Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

January 2009

Subunit Compositions of Arabidopsis DNA-Dependent RNA Polymerases and the Roles of the Plant-Specific RNA Polymerases IV and V in Gene Silencing

Thomas Ream
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

Recommended Citation

Ream, Thomas, "Subunit Compositions of Arabidopsis DNA-Dependent RNA Polymerases and the Roles of the Plant-Specific RNA Polymerases IV and V in Gene Silencing" (2009). *All Theses and Dissertations (ETDs)*. 288.
https://openscholarship.wustl.edu/etd/288

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY

Division of Biology and Biomedical Sciences

Program in Plant Biology

Dissertation Examination Committee:
Craig S. Pikaard, Ph.D., Chairman
Douglas L. Chalker, Ph.D.
Sarah R. Elgin, Ph.D.
Roberto Galletto, Ph.D.
Joseph Jez, Ph.D.
John Majors, Ph.D.

SUBUNIT COMPOSITIONS OF *ARABIDOPSIS* DNA-DEPENDENT RNA POLYMERASES AND THE ROLES OF THE PLANT-SPECIFIC RNA POLYMERASES IV AND V IN GENE SILENCING

by

Thomas Scott Ream

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2009

Saint Louis, Missouri

Copyright by

Thomas Scott Ream

2009

ABSTRACT OF THE DISSERTATION

Subunit compositions of *Arabidopsis* DNA-dependent RNA polymerases and the roles of the plant-specific RNA Polymerases IV and V in gene silencing

by

Thomas Scott Ream

Doctor of Philosophy in Plant Biology

Washington University in St. Louis, 2009

Professor Craig S. Pikaard, Ph.D., Chairman

In addition to RNA polymerases I, II and III, the essential RNA polymerases present in all eukaryotes, plants have two additional nuclear RNA polymerases, abbreviated as Pol IV and Pol V, that play non-redundant roles in siRNA-directed DNA methylation and gene silencing. Using a combination of affinity purification and protein identification by mass spectrometry, my studies show that *Arabidopsis* Pol IV and Pol V are composed of subunits that are paralogous or identical to the twelve subunits of Pol II. Four subunits of Pol IV are distinct from their Pol II paralogs, six subunits of Pol V are distinct from their Pol II paralogs, and four subunits differ between Pol IV and Pol V. Importantly, the subunit differences occur in key positions relative to the template entry and RNA exit paths. My findings support the hypothesis that Pol IV and Pol V are Pol II-like enzymes that evolved specialized roles in the production of noncoding transcripts for RNA silencing and genome defense. In addition, I have determined the subunit compositions of *Arabidopsis* RNA polymerases I, II and III, providing a novel framework

for comparative and functional analyses of subunits from all five DNA-dependent RNA polymerases in plants.

ACKNOWLEDGEMENTS

First and foremost, I thank my wife, Lauren, for being so supportive and loving over the last 8 years. Without her love and encouragement I would not be where I am today.

I thank my family and friends for their love and support in and out of graduate school. In particular, I thank my parents, Thomas and Mary, and my sister, Jennifer.

I thank all the members of the Pikaard lab during my tenure in the lab for their technical help and valuable insight. In particular, I thank Jeremy Haag and Keith Earley, who are role models for the research community and have helped me to make great strides as a researcher.

I thank the members of my committee for their help and insight over the course of my thesis.

I thank Craig Pikaard for the opportunity to study in his laboratory and for his excellent mentoring and willful encouragement to pursue projects of my choice.

I thank the following institutions for funding: National Institutes of Health, Department of Energy and the Department of Agriculture, Monsanto Company and the Washington University Division of Biology and Biomedical Sciences.

Lastly, I would like to thank my undergraduate research mentors, Dr. James Birchler and Dr. Donald Auger, for introducing me into the world of plant research and for their enthusiastic encouragement to attend graduate school.

TABLE OF CONTENTS

Acknowledgements iv

Chapter 1: Introduction 1-112
i. Prologue 2
ii. RNA-mediated gene silencing and heterochromatin formation 3
a. The chromatin landscape 4
i. Heterochromatin and euchromatin 4
ii. DNA methylation in plants 6
iii. Interplay between DNA methylation and histone modifications 8
iv. Reversibility of chromatin modifications 9
b. RNA interference 12
i. Discovery of RNAi 12
ii. Argonaute and Dicer: central players in RNAi 14
iii. RNAi in yeast 16
iv. RNAi in *Drosophila* 19
v. RNAi in mammals 23
vi. RNAi in plants 24
vii. The roles of plant-specific RNA polymerases IV and V 36
viii. Ancient origin of RNAi and Pol IV/V 45
iii. DNA-dependent RNA polymerases: subunit compositions and roles of their subunits in transcription 48
a. Overview of prokaryotic, archaeal and eukaryotic RNA polymerases 48
i. Nomenclature of yeast and plant DNA-dependent RNA polymerase subunits 55
ii. Identification of *Arabidopsis* RNA polymerase subunits 58
iii. Conservation and evolution of RNA polymerase subunits 60
iv. Structural analysis of RNA polymerases 62
b. Roles of the different RNA polymerase subunits in transcription 65
i. The catalytic subunits 66
ii. Subunits involved in assembly 67
iii. Subunits shared by all three polymerases 68
iv. Unique subunits with functional equivalents among the three polymerases 70
v. Pol I-specific subunits 76
vi. Pol III-specific subunits 77
c. Early studies of *Arabidopsis* RNA polymerase subunits 78
i. Rpb3-related subunits 79
ii. Rpb4 and Rpb7-related subunits 80
iii. Rpb5-related subunits 81
iv. Scope of this thesis 82
v. References cited 85

Chapter 2: Plant nuclear RNA polymerase IV mediates siRNA and DNA-methylation dependent heterochromatin formation 113-146

Chapter 3: Subunit structure of the RNA silencing enzymes, Pol IV and Pol V, reveal their origins as specialized forms of RNA polymerase II 147-206

Chapter 4: Subunit characterization of RNA polymerases I, II, III, IV and V from *Arabidopsis thaliana* reveals redundancy and functional diversification among subunit families 207-293

Chapter 5: Conclusions and future directions 294-351

i. Prologue 295

ii. Elucidating the role of non-catalytic subunits in multi-subunit DNA-dependent RNA polymerase function 295

a. Introduction 295

b. Testing for functional redundancy between subunits 297

i. NRPB6 and NRPB8 families 297

ii. The NRPB9 family 300

iii. The NRPB3 family 302

iv. Isoforms of polymerases? 303

v. Roles of the NRPB5 family 307

vi. NRPD4/NRPE4, NRPD7 and NRPE7 functions 310

vii. The roles of NRPB4 and NRPB7 319

viii. Subunits with unknown functions 322

ix. Characterization of RNA polymerase subunit antibodies 324

iii. Toward mechanisms of Pol IV and Pol V gene silencing 325

a. Introduction 325

b. Activities of Pol IV and Pol V 327

c. Understanding where Pol IV and Pol V act in the nucleus 330

d. Role of post-translational modification in RNA polymerase-mediated silencing 332

e. Pol IV and Pol V in the context of other gene silencing components 334

f. A role for Pol V in heterochromatin stability 340

iv. Evolution of plant RNA polymerase subunits 341

Appendix A: Roles of RNA polymerase V in gene silencing 352-361

Appendix B: The *Arabidopsis* Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center 362-381

Appendix C: Sex-biased lethality or transmission of defective transcription machinery in *Arabidopsis* 382-401

Appendix D: RNA Polymerase V transcription guides AGO4 to chromatin 402-408

Appendix E: *Curriculum vitae* 409-414

LIST OF TABLES AND FIGURES

Chapter 1: Introduction

Figure 1: Subunit composition of a prokaryotic RNA polymerase. 49

Table 1: Relationships of DNA-dependent RNA polymerase subunits from bacteria, archaea, yeast and *Arabidopsis*. 50

Figure 2: Subunit compositions of prokaryotic vs. eukaryotic RNA polymerases. 52

Figure 3: Subunit compositions of yeast RNA polymerases I, II and III. 53

Figure 4: Subunit structure of yeast Pol II. 54

Figure 5: Prominent structural features of yeast RNA polymerase II. 63

Chapter 2

Figure 1: Evidence for RNA Pol IV in Plants 116

Figure 2: NRPD2 Does Not Cofractionate with Pol II or with DNA-Dependent RNA Polymerase Activity 118

Figure 3: Heterochromatin Is Disrupted in *nrpd2* Mutants 119

Figure 4: NRPD1 and NRPD2 Are Required for 5S Gene and AtSN1 Cytosine Methylation and siRNA Accumulation 121

Tables S1 and S2: GenBank Accessions for the DNA-Dependent RNA Polymerase Largest Subunits Analyzed in Figure 1. 127

Figure S1: Multiple Alignment of RPD1 with DNA-Dependent RNA Polymerase Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec). 130

Figure S2: Multiple Alignment of RPD2 with DNA-Dependent RNA Polymerase Second-Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec). 137

Figure S3: Comparison of Conserved Domains A–H in RPD1a and DNA-Dependent RNA Polymerase Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec). 141

Figure S4: Comparison of Conserved Domains A–I in RPD2 and DNA-Dependent RNA Polymerase Second- Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec). 142

Figure S5: Determination of the Full-Length mRNA Sequence for RPD2a by RT-PCR, 5' RACE, and Primer Extension. 143

Table S3: Cytological Changes in *rpd2* Mutants. 145

Chapter 3

Figure 1: Relationships of *Arabidopsis* Pol II, IV and V subunits to *E. coli*, archaeal, and yeast RNA Pol II subunits 151

Figure 2: Verification of Pol V subunit associations 152

Figure 3: Pol V utilizes a distinct RPB3 variant, NRPE3b, as well as an NRPE3a variant corresponding to the Pol II NRPB3 subunit 153

Figure 4: Co-IP tests of Pol V, IV and II subunit associations 154

Figure 5: *nrpe5* mutants are defective in RNA-directed DNA methylation and retrotransposon silencing 155

Figure 6: Comparison of RNA polymerase subunits in Pol II, IV and V 157

Table S1: Genes whose known or predicted sequences were used in peptide coverage maps and/or protein alignments. 165

Table S2: Subunits of Arabidopsis Pol V identified by LC-MS/MS analysis of immunoprecipitated FLAG-NRPE5. 166

Table S3: List of primer sequences. 167

Figure S1: Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V NRPE1-FLAG). 168

Figure S2: Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol IV (NRPD1-FLAG). 174

Figure S3: Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol II (NRPB2-FLAG). 180

Figure S4: RPB4 family alignment. 186

Figure S5: RPB5 family alignment. 187

Figure S6: RPB6 family alignment. 188

Figure S7: RPB7 family alignment. 189

Figure S8: RPB8 family alignment. 190

Figure S9: RPB9 family alignment. 190

Figure S10: RPB10 family alignment. 190

Figure S11: RPB11 family alignment. 190

Figure S12: RPB12 family alignment. 191

Figure S13: Expression patterns of the RPB5 family. 191

Figure S14: Analysis of *nrpd5-1* and *nrpe11-1* T-DNA insertion mutants. 192

Figure S15: Flowering time of individual plants from wild-type (ecotype Col-0) and *nrpe5-1* populations. 193

Figure S16: Alignment of RPB5 family variants in diverse plants with non-plant RPB5s. 194

Figure S17: The N-terminal extension of NRPE5 is required for the protein's stability and function. 198

Figure S18: Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS analysis in affinity purified FLAG-NRPE5 samples. 200

Chapter 4

Table 1: Number of Pol I and Pol III-specific subunits in plants and yeast. 244

Table 2: Peptide coverage corresponding to RNA polymerase subunits identified in affinity purified Pol I and Pol III complexes. 245

Figure 1: Immunoblots testing the specificity of antibodies directed against Pol I or Pol III specific subunits, NRPA3 and NRPC7, respectively. 246

Figure 2: Verification of Pol I and Pol III subunit associations using co- 247

immunoprecipitation tests.

Figure 3: Alternative RNA polymerase subunits are expressed but do not associate with Pol I, II, III, IV or V. 248

Table 3: Genotyping analysis of F2 progeny segregating T-DNAs in *NRPA11*, *NRPA3* or *NRPC3* genes. 249

Table 4: Partitioning of DNA-dependent RNA polymerase subunits in the five *Arabidopsis* RNA polymerases compared to the equivalent subunits of yeast, archaea and bacteria. 250

Table S1: List of primers used in this study. 254

Table S2: List of Genbank accession numbers for proteins used in alignments. 255

Figure S1: ClustalW-aligned Rpa49-like sequences from plants, fungi and animals. 256

Figure S2: ClustalW-aligned Rpc17-like sequences from plants, fungi and animals. 258

Figure S3: ClustalW-aligned Rpc31-like sequences from plants, fungi and animals. 259

Figure S4: ClustalW-aligned Rpc34-like sequences from plants, fungi and animals. 262

Figure S5: ClustalW-aligned Rpc37-like sequences from plants, fungi and animals. 264

Figure S6: ClustalW-aligned Rpc53-like sequences from plants, fungi and animals. 267

Figure S7: ClustalW-aligned Rpc82-like sequences from plants, fungi and animals. 271

Figure S8: ClustalW-aligned Rpac40-like sequences from plants, fungi and animals. 273

Figure S9: Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V (NRPA2-FLAG). 276

Figure S10: Peptide coverage maps of DNA-directed RNA polymerase 283

subunits detected by LC-MS/MS in affinity purified Pol V (NRPC2-FLAG).

Figure S11: ClustalW-aligned Rpa34/PAF49 sequences from plants, fungi and animals. 291

Conclusions and future directions

Table 1: T-DNA insertion lines available for RNA polymerase subunits. 298

Figure 1: DNA methylation and siRNA accumulation in *nrpe9* mutants is identical to wild-type plants. 301

Figure 2: RT-PCR of *NRPE6* and *NRPE8* genes shows both variants are ubiquitously expressed in the tissues tested. 303

Table 2: Transgenic lines and antibodies available for studying RNA polymerase subunits. 305

Figure 3: *nrpe7* mutants affect DNA methylation at Pol V-dependent loci. 312

Figure 4: FLAG-NRPB4 rescues an *nrpb4* mutant and is active for promoter-independent transcription. 320

Table 3: Total number of proteins identified in affinity purified Pol IV and Pol V lines. 335

Table 4: Peptides matching Reptin and Pontin homologs in Pol IV or Pol V affinity purifications. 336

Table 5: Proteins associated with Pol IV and Pol V after filters. 337

Figure 5: RNA polymerase subunits arose by duplication of the *Arabidopsis* genome. 342

Appendix A:

Glossary 354

Figure 1: Catalytic subunits of DNA-dependent RNA polymerases. 355

Box 1: Pol IV subunit nomenclature 356

Figure 2: A variety of proteins participate in Pol IVa-dependent silencing pathways. 358

Figure 3: Possible modes of Pol IVa function. 359

Appendix B:

Figure 1: Loss of siRNAs and Cytosine Methylation at Repeated DNA Sequences in Mutants of the Nuclear siRNA Pathway 365

Figure 2: Nuclear Localization of siRNAs 367

Figure 3: Immunolocalization of Nuclear siRNA Pathway Proteins 369

Figure 4: siRNAs Colocalize with NRPD1b, RDR2, DCL3, and AGO4 370

Figure 5: Pol IV Colocalizes with Endogenous Repeat Loci 371

Figure 6: Effects of Mutations on the Localization of Proteins Involved in Nuclear siRNA Biogenesis 372

Figure 7: A Spatial and Temporal Model for Nuclear siRNA Biogenesis 373

Figure S1: Antibody Specificity Controls 378

Figure S2: NRPD1a and NRPD1b Immunolocalization Signals Are Not Lost in DNase I-Treated Nuclei 379

Figure S3: Immunolocalization of DRD1, NRPD1b, and NRPD1a in *drd1-6* Mutant Nuclei 379

Table S1: Supporting Data for Figure 2A: siRNA Probe Hybridization Patterns and Frequencies 380

Table S2: Supporting Data for Figure 3A: Protein Localization and Effects of RNase 380

Table S3: Supporting Data for Figure 3C: Pairwise Detection of Nuclear siRNA Pathway Proteins 380

Table S4: Supporting Data for Figure 4: Protein-siRNA Colocalization 381

Table S5: Supporting Data for Figure 5: Localization of Proteins Relative to NORs and 5S Gene Loci 381

Table S6: Supporting Data for Figure 6: Protein Localization in Various Nuclear siRNA Pathway Mutants. 381

Appendix C:

Figure 1: Sex-biased transmission of disrupted alleles for second-largest subunits of RNA polymerases I, II, and III (NRPA2, NRPB2, and NRPC2, respectively). 387

Table 1: Genotypes of progeny of Pol I, II, and III hemizygotes. 388

Table 2: Male-specific transmission of Pol I, II, and III mutant alleles. 388

Figure 2: Failed seed development in siliques of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* hemizygotes. 389

Figure 3: Developmental arrest of mutant female gametophytes in flowers just prior to anthesis was visualized by confocal fluorescence microscopy. 390

Table 3: Female gametophyte development in polymerase mutants. 391

Figure 4: Development and early tube elongation of pollen are unaffected by defects in RNA polymerases. 392

Figure 5: Reduced paternal transmission of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles relative to wild-type alleles in self-fertilized hemizygotes is due to decreased, competitive fertilization of ovules farthest from the stigma. 393

Table S1: Male-specific transmission of *RPB12a* mutant alleles. 398

Table S2: Transgene rescue allows maternal transmission of mutant alleles. 399

Figure S1: Developmentally arrested mutant female gametophytes within pistils just prior to anthesis, visualized by confocal fluorescence microscopy. 400

Appendix D:

Figure 1: Pol V, AGO4 and DMS3 work non-redundantly in heterochromatin formation. 405

Figure 2: AGO4 is not required for Pol V transcription. 405

Figure 3: Pol V transcription is necessary for AGO4-chromatin interactions. 406

Figure 4: AGO4 physically interacts with Pol V transcripts. 406

Figure 5: The SMC hinge-domain protein DMS3 is required for Pol V transcription and detectable Pol V-chromatin interactions. 407

Figure 6: A model for Pol V and siRNA-dependent heterochromatin formation. 407

CHAPTER ONE

INTRODUCTION

Prologue

The central dogma of molecular biology is that DNA is transcribed into RNA by DNA-dependent RNA polymerases, followed by the subsequent translation of RNA into protein. This essential process is highly regulated at each step in all forms of life, from bacteria to fungi to plants to animals. Over the last twenty years, mounting evidence has suggested a pivotal role for RNA in negatively regulating gene expression, a phenomenon known as RNA-mediated gene silencing, or RNA interference (RNAi). In addition to gene regulation, RNA is a major influence in the formation of distinct nuclear domains, such as euchromatin and heterochromatin, and in mediating defense responses to viruses, transposable elements and pathogens. RNA plays a crucial role in other basic cellular processes such as chromosome segregation, centromere formation, developmental transitions and maintaining or promoting cell identity. The following introduction will summarize RNA-mediated gene silencing and the structure and function of DNA-dependent RNA polymerases. Recently, it has become clear that these two processes are mechanistically linked. I begin with a general overview of chromatin and RNAi in different kingdoms and then shift towards RNAi mechanisms in plant systems, focusing on the model plant *Arabidopsis thaliana*. Along the way I will discuss relevant features of the silencing process, such as the proteins involved and their activities, the role of heterochromatin in silencing and RNA-directed DNA methylation. I close the first part of the introduction with a review of Pol IV and Pol V-mediated silencing pathways in *Arabidopsis*, many of which feed back into the processes listed above. In the second part of the introduction, I review the structure and function of DNA-dependent RNA polymerases, focusing on what has been demonstrated in yeast and plants.

i.

RNA-MEDIATED GENE SILENCING AND HETEROCHROMATIN FORMATION

RNA-mediated gene silencing, or RNA interference (RNAi), is a widespread phenomenon that occurs in many different eukaryotic organisms within the plant, fungi and animal kingdoms (Agrawal et al. 2003). RNA silencing typically involves production of double-stranded RNA trigger molecules by viral replication, bi-directional transcription or amplification of single-stranded RNA by RNA-dependent RNA polymerases (RDRs). These dsRNA triggers are used as substrates by Dicer enzymes, which cleave the dsRNA into small RNAs (sRNAs) of 20-40 nucleotides in length. These sRNAs then incorporate into Argonaute (AGO)-containing RNA induced silencing complexes (RISC) to act as guide molecules that direct either transcriptional or post-transcriptional gene silencing (TGS or PTGS, respectively). In TGS, genes are turned off by using siRNAs to guide repressive chromatin modifications to the target gene sequence or associated histones; in PTGS, siRNAs are used to seek and destroy complementary mRNAs or to mediate translational inhibition, resulting in high mRNA turnover and decreased expression of the target gene (Sijen et al. 2001; Pal-Bhadra et al. 2002; Vazquez 2006).

In plants and animals, RNAi is required for maintaining genome integrity, defense against viruses and transposases, gene regulation and heterochromatin formation, allof which could potentially compromise the viability of the organism if left unregulated (Birchler et al. 2000; Henikoff 2000; Waterhouse et al. 2001; Agrawal et al. 2003;

Lippman et al. 2004; Lippman and Martienssen 2004; Almeida and Allshire 2005; Blevins et al. 2006; Fritz et al. 2006; Vagin et al. 2006; Wang et al. 2006; Riddle and Elgin 2008; Huisinga and Elgin 2009; Matzke et al. 2009). Several mechanisms are used in RNA-based silencing to control gene expression in a variety of organisms, including yeast, plants, flies and humans. In one mechanism, RNAi facilitates heterochromatin formation by targeting the activities of chromatin-modifying proteins in a locus-specific manner to regulate gene expression (Volpe et al. 2002; Zilberman et al. 2003; Chan et al. 2004; Morris et al. 2004; Pal-Bhadra et al. 2004). Other RNAi based mechanisms control gene regulation using specialized small RNAs, termed microRNAs (miRNAs), to post-transcriptionally regulate genes that are important for developmental timing, stress responses or innate immunity (reviewed in (Banerjee and Slack 2002; Dugas and Bartel 2004; Millar and Waterhouse 2005)). In addition, plants utilize two classes of small RNAs, termed trans-acting siRNAs and cis-acting natural anti-sense siRNAs, regulate transcript levels of key regulatory genes similar to miRNAs but by using different mechanisms (Peragine et al. 2004; Vazquez et al. 2004; Borsani et al. 2005). Before elaborating on these RNAi mechanisms, it is important to understand the broader context of chromatin and chromatin-modifying proteins that are intricately linked to the RNAi mechanism.

<u>The chromatin landscape</u>

Heterochromatin and euchromatin

RNA-based silencing can occur post-transcriptionally by the destruction of mRNA or transcriptionally by inducing repressive chromatin states that preclude efficient

transcription. Chromatin is defined as DNA and its associated proteins, most notably histones. Histones are comprised of an octamer of four proteins, H2A, H2B, H3 and H4, with each protein contributing two copies (Luger et al. 1997). The nucleosome core particle is wrapped by 147 base pairs of DNA (Luger et al. 1997; Richmond and Davey 2003). Histone-fold motifs are conserved among each subunit and mediate the interaction with the looping DNA (Luger et al. 1997). The N-terminal tails are variable among the histone subunits and are subject to a wide and dynamic range of reversible modifications, including methylation, acetylation, ubiquitination and phosphorylation (reviewed in (Taverna et al. 2007; Marmorstein and Trievel 2009)). These modifications are hypothesized to encode information that dictates the local chromatin state, known as the "histone code hypothesis" (Jenuwein and Allis 2001). For example, histone hypo-acetylation and methylation of H3 lysine (K) 9 or H3K27 are conserved marks of heterochromatin across higher eukaryotes, whereas methylation of H3K4 and hyper-acetylated histones are conserved euchromatic marks (Fuchs et al. 2006). These marks have the potential to act alone or in combination with other histone modifications to regulate chromatin (reviewed in (Fuchs et al. 2006; Suganuma and Workman 2008)).

Chromatin is typically grouped into one of two functional classes—euchromatin and heterochromatin. Euchromatin is generally more accessible to RNA polymerases and is enriched for active gene marks such as acetylation of histone tails and methylation of H3K4 (Grewal and Elgin 2007). Heterochromatin is defined as densely staining portions of the genome throughout the cell cycle, including interphase, which correlates with dense chromatin packaging, transcriptionally repressive chromatin marks such as H3K9 methylation, ordered nucleosome structures, and low gene density (Richards and Elgin

2002; Grewal and Moazed 2003). Heterochromatin replicates late in S phase and is generally found near the centromeres and telomeres, although it may be located along the chromosome arms (Elgin and Grewal 2003; Zhimulev and Belyaeva 2003; Wallace and Orr-Weaver 2005). There is essentially no meiotic recombination within heterochromatin. Heterochromatin is typically enriched in repeated sequences and transposable elements, suggesting that heterochromatin formation helps control the expression of these potentially deleterious sequences (Elgin and Grewal 2003; Wallace and Orr-Weaver 2005). Heterochromatin formation facilitates the assembly of the functional centromere, although it may not be sufficient for centromere formation (Ekwall 2004b). At the yeast centromere, heterochromatin recruits cohesin to hold sister chromatids together until anaphase. To this end, mutants that compromise heterochromatin formation in yeast, such as mutants in the RNAi machinery, display chromosome segregation defects during anaphase (Volpe et al. 2002; Volpe et al. 2003).

DNA methylation in plants

In addition to histone modifications, DNA is also epigenetically modified by the addition of a methyl group to the cytosine base. In plants and animals, cytosines in a symmetric CpG context are maintained by maintenance methyltransferases during replication. In *Arabidopsis*, maintenance CG methylation is accomplished by MET1 (METHYLTRANSFERASE1) (Vongs et al. 1993; Finnegan et al. 1996; Ronemus et al. 1996; Kankel et al. 2003). In plants, methylation also occurs outside the context of CpG, in contrast to animals. A second type of symmetric methylation can occur in a CpNpG context, where N is any nucleotide but C. A majority of this maintenance CNG

methylation is performed by CMT3 (CHROMOMETHYLASE3), but at some loci DRM2 (DOMAINS REARRANGED METHYLTRANSFERASE) is required for CNG maintenance methyltransferase activity, either alone or in combination with CMT3 (Cao and Jacobsen 2002; Cao et al. 2003; Kankel et al. 2003). In addition, DNA methylation can also occur in an asymmetric context in plants. Maintenance of asymmetric methylation is more complicated because there is no template for the maintenance of DNA methylation during replication. Therefore, this type of DNA methylation is thought to require the presence of an RNA silencing trigger in order to be established by the *de novo* methyltransferase, DRM2, a process known as RNA-directed DNA methylation (see below) (Cao and Jacobsen 2002). However, evidence suggests that MET1 or CMT3 may also play direct or indirect roles in *de novo* asymmetric methylation because this type of methylation is lost genome wide in *met1* and *cmt3* mutants (Cokus et al. 2008; Lister et al. 2008).

RNA-directed DNA methylation was first described in plants and was discovered when viral cDNA sequences integrated into the genome became methylated in a manner that was strictly dependent on replication of the virus, which involves an RNA intermediate (Wassenegger et al. 1994). A later study revealed that dsRNA specifically could initiate RNA-directed DNA methylation to homologous sequences in plants (Mette et al. 2000). As little as 30 base pairs (bp) of RNA can trigger methylation of the homologous sequence, suggesting that this targeted modification may provide higher resolution gene regulation than modification of histones, which occur in nucleosome core particles spaced every ~200 bp of DNA (Pelissier and Wassenegger 2000). Although the precise mechanism by which RNA is used to direct DNA methylation is unknown,

siRNAs are leading candidates because of their small size (~21-24 bp on average in *Arabidopsis thaliana*). In addition, a siRNA-mediated mechanism is supported by the requirement for Argonaute and Dicer proteins in RNA-directed DNA methylation (Zilberman et al. 2003; Chan et al. 2004; Xie et al. 2004). In many cases, RNA-directed DNA methylation is usually confined to the regions of siRNA complementarity, but there are notable exceptions when silencing spreads out from the initial target (see section below on "Spreading of silencing in plants"). RNA-directed *de novo* DNA methylation can occur on cytosines in any context (CG, CNG, or asymmetric C). Methylation that occurs in CG or CNG contexts can be subsequently maintained during DNA replication by cytosine DNA maintenance methyltransferases MET1 and CMT3 in the absence of the RNA trigger.

As much as 30% of the methylated genome results from targeting by RNA in *Arabidopsis*. As expected, much of the DNA methylation is clustered near peri-centromeres and other heterochromatic domains, although there are examples of significant DNA methylation in euchromatin, often near transposons or repeats (Cokus et al. 2008; Lister et al. 2008). Some of these euchromatic repeats or remnant transposable elements are known to regulate endogenous genes (Liu et al. 2004b; Chan et al. 2006; Saze and Kakutani 2007; Henderson and Jacobsen 2008).

Interplay between DNA methylation and histone modifications

Drosophila and fission yeast lack significant DNA methylation. Therefore, heterochromatin formation relies primarily on the interplay between histone modifications and heterochromatin-associated proteins. However, plants utilize both

DNA methylation and histone modifications to mediate heterochromatin formation. For example, H3K9 methylation is significantly reduced in plants that are deficient for the CG maintenance methyltransferase, MET1, suggesting that DNA methylation occurs upstream of H3K9 methylation (Soppe et al. 2002; Tariq et al. 2003). In these cases, loss of H3K9 correlates with activation of transcription (Johnson et al. 2002). However, mutations in the *Arabidopsis* H3K9 methyltransferase, KRYPTONITE (KYP), lead to loss of CNG methylation, suggesting that H3K9 methylation acts upstream of DNA methylation (Jackson et al. 2002). In other organisms with both DNA methylation and histone modifications, such as *Neurospora*, histone methylation directs DNA methylation (Tamaru and Selker 2001). It appears that plants have evolved a complex interplay between histone modifications and DNA methylation that likely depends on the context of a given set of modifications.

Reversibility of chromatin modifications

It was previously thought that chromatin modifications such as cytosine methylation and histone lysine methylation were permanent, covalent marks. However, in the last few years multiple reports have demonstrated that enzymes exist to mediate the removal of these groups (reviewed in (Agger et al. 2008; Ikeda and Kinoshita 2009; Ng et al. 2009)). These reports support a dynamic model of the heterochromatin state in contrast to a static model. Targeted, reversible chromatin modifications may allow for precise spatial and temporal changes in chromatin environments, which could result in changes in gene expression to meet the demands of the cell under certain conditions such as changes in the environment, developmental cues, stressors or the cell cycle.

Recently, DNA glycosylases have been shown to mediate DNA demethylation by a base pair excision mechanism in plants (Agius et al. 2006; Gehring et al. 2006; Morales-Ruiz et al. 2006). These demethylases provide evidence for the dynamic nature of DNA methylation, particularly in euchromatic regions where genes may need to be switched rapidly from the "off" to "on" state based on environmental cues (reviewed in (Huettel et al. 2007)). To this end, euchromatic genes tend to be enriched in RNA-directed, asymmetric DNA methylation compared to the more "permanent", repressive chromatin modifications such as CG methylation and H3K9 methylation (reviewed in (Huettel et al. 2007)). Of the four known DNA demethylases in plants, ROS1 (Repressor of Silencing 1) is the best characterized. In *ros1* mutants, DNA methylation increases throughout the genome, suggesting that ROS1 actively removes methylation catalyzed by the DNA methylation machinery (Lister et al. 2008). These opposing activities may promote a balance of DNA methylation that ultimately influences gene expression (Gong et al. 2002).

As stated above, DNA demethylation in plants is proposed to occur via a base-excision repair pathway (Agius et al. 2006; Gehring et al. 2006; Morales-Ruiz et al. 2006). In mammals, experiments with proteins involved in active demethylation or the mechanisms involved are either hotly debated or not repeatable (reviewed in (Ooi and Bestor 2008)). A recent hypothesis proposes that the mammalian DNA methylation machinery, DNMT3A and DNMT3B, is involved in both methylation and demethylation. In the absence of the available methyl group donor, DNMT3 can deaminate the cytosine to thymine, which recruits the mismatch repair machinery and leads to replacement of methyl-cytosine with cytosine, resulting in demethylation (Kangaspeska et al. 2008;

Metivier et al. 2008). Plants differ from mammals in that they do not erase epigenetic marks from the parent generation in early development. To achieve gene imprinting in plants, DNA demethylases actively demethylate the parent-specific allele to be expressed (Choi et al. 2002; Kinoshita et al. 2004; Jullien et al. 2006). In mammalian imprinting, both genes initially have the capacity to be expressed but one is selectively methylated (Ooi and Bestor 2008). This fundamental difference explains why plants need to actively demethylate their genomes, at least in the context of early development, and would argue against a reason for active demethylation early in mammalian development (Ooi and Bestor 2008).

Enzymes mediating the deacetylation, demethylation or deubiquitination of histones reinforce the idea that covalent chromatin modifications are dynamically regulated. SUP32 is a *bona fide* histone deubiquitinase that is required for RNA-directed DNA methylation and H3K9 di-methylation in *Arabidopsis* (Sridhar et al. 2007). Consistent with this idea, H2B levels of ubiquitination increase in *sup32* mutants, coincident with an increase in the euchromatic mark, H3K4 tri-methylation. Mutations in *IBM1* (*Increase in BONSAI methylation 1*) result in increased levels of H3K9 di-methylation at the *BONSAI* gene that spill over from a neighboring transposable element. IBM1 is a putative histone lysine demethylase that is hypothesized to prevent the spread of H3K9 methylation into the *BONSAI* gene under normal circumstances in *Arabidopsis* (Saze et al. 2008).

Lysines of H3 and H4 can be acetylated or deacetylated by a number of different histone acetylases and deacetylases. In plants, these modifications have many consequences on transcription and development (Chen and Tian 2007; Hollender and Liu

2008). HDA6, one of eighteen histone deacetylases in *Arabidopsis*, is required for silencing transgenes, transposons and rDNA and affects DNA methylation patterns at some of these loci (Murfett et al. 2001; Aufsatz et al. 2002; Probst et al. 2004; Earley et al. 2006). In addition, HDA6 is important for regulating genes involved in senescence, flowering time and pathogen or wound-responses (Wu et al. 2008). In contrast, histone acetylases add acetyl groups to the lysines of H3 and H4 histone tails to promote transcription. This may occur by altering the charge of the histone tails and reducing their affinity for DNA or by recruiting specific chromatin remodeling proteins or transcriptional activators (Kuo and Allis 1998; Jenuwein and Allis 2001). In *Arabidopsis*, histone acetyltransferases regulate a wide array of developmental programs including cell fate, cold acclimation and light responses (Stockinger et al. 2001; Vlachonasios et al. 2003; Bertrand et al. 2005; Long et al. 2006).

RNA interference

Discovery of RNAi

RNA interference was initially observed in plants, although the mechanism was unclear at the time and the phenomenon was referred to as co-suppression. In the first case, plants engineered to carry kanamycin and hygromycin selectable markers with regions of homology in their respective transgenes often resulted in plants with a silenced kanamycin gene (Matzke et al. 1989). Methylation of the promoter correlated with silencing and was released upon segregation of the two selectable markers in later generations (Matzke et al. 1989). In a second case, attempts to increase pigment production in petunias by the addition of exogenous copies of the *chalcone synthase* gene

resulted in lower pigmentation levels than expected in some plants (Napoli et al. 1990). Similar results were observed in later studies using *Neurospora* and *Drosophila* (Romano and Macino 1992; Pal-Bhadra et al. 1997; Fire et al. 1998). However, these studies fell short of identifying the specific cause of the co-suppression, or gene silencing, effects. In 1998, Fire and Mello demonstrated that injecting *C. elegans* with as little as a few molecules of dsRNA per cell caused substantial gene silencing effects when compared to injection of either single strand alone (Fire et al. 1998). They further hypothesized that the sub-stoichiometric levels of dsRNA to endogenous mRNA could implicate an amplification mechanism for the dsRNA. That same year, a systematic study using a plant model showed that transgene expression of dsRNA could induce PTGS and viral resistance more effectively than either RNA strand alone (Waterhouse et al. 1998). Indeed, the first screens designed to identify gene silencing factors uncovered RNA-dependent RNA polymerases in *C. elegans* and plants (Tabara et al. 1999; Dalmay et al. 2000; Smardon et al. 2000). The requirement for an RNA-dependent RNA polymerase in gene silencing provided substantial support for the role of dsRNA in silencing and also provided a mechanism for amplification of the silencing trigger.

In 1999, Hamilton and Baulcombe were the first to show that small RNAs complementary to silenced RNAs correlate with post-transcriptional silencing in plants (Hamilton and Baulcombe 1999). This study provided the platform for a series of papers making the connection between dsRNA cleavage into small RNAs by an RNase III-like enzyme, Dicer, and incorporation of these small RNAs into an RNA-induced silencing complex (RISC). From these studies, hypotheses emerged suggesting that sRNAs

incorporated into RISC provided the vehicle for achieving silencing specificity (Bernstein et al. 2001; Hammond et al. 2001).

The discovery of the Argonaute family proteins provided a critical link between small RNAs and silencing. Tabera et al. identified the *C. elegans rde-1* mutant as an *Argonaute* gene defect in a genetic screen for loss of RNA silencing (Tabara et al. 1999). Indeed, purification of the first RISC and RNA-induced initiation of transcriptional gene silencing (RITS) complexes revealed that Argonautes were central components that bind small RNAs and use them as guides to direct post-transcriptional or transcriptional gene silencing (Hammond et al. 2001; Verdel et al. 2004). The importance of Argonaute family proteins for development was revealed by isolation of *ago1* mutants that affect leaf development and shoot apical meristem integrity in *Arabidopsis* (Bohmert et al. 1998; Moussian et al. 1998). In *Drosophila*, *piwi* mutations (mutants representing a sub-class of the *Argonaute* gene family) compromise germ-line stem cell development (Lin and Spradling 1997; Cox et al. 1998).

Argonaute and Dicer: central players in RNAi

Although the specific mechanisms of RNAi differ among organisms and even among different RNAi pathways within a given organism, there are two components that are required for nearly every RNAi pathway, a Dicer and an Argonaute protein. Dicers are RNase III-like enzymes that bind and cleave dsRNA into small RNAs ranging from 20-40 nucleotides (nt) in size, although 21-24 nt are typically the most common size classes of small RNAs, especially in *Arabidopsis*. Dicers consist of a PAZ domain that can bind RNA and tandem RNase III domains to perform the dsRNA cleavage.

Crystallization of a Dicer from *Giardia* has shown that Dicers themselves are the measuring sticks for determining how dsRNA is cleaved into small RNAs of a particular size (Macrae et al. 2006a; Macrae et al. 2006b).

After dicing, small RNAs are loaded into Argonaute proteins that possess PAZ, MID and PIWI domains. The strand of the small RNA duplex that is selectively incorporated into an Argonaute protein is determined by several factors. The 5' strand of the siRNA duplex at the least thermodynamically stable end of the duplex is the strand that is typically bound by the Argonaute protein. However, the direction of processing, or processing polarity, also appears to have a role in determining the fate of the duplex (Elbashir et al. 2001b; Rose et al. 2005). Argonautes bind the 3' end of small RNAs using their PAZ domain and bind the 5' phosphate of the siRNA using their MID domain, which can also serve as a protein-protein interaction domain (Lingel et al. 2003; Song et al. 2003; Lingel et al. 2004; Ma et al. 2004; Ma et al. 2005; Parker et al. 2005). Target RNA cleavage is carried out by their PIWI domains (Liu et al. 2004a; Song et al. 2004). Argonaute-containing ribonucleoprotein complexes are often referred to as "effector complexes" because they mediate the mRNA cleavage, translational inhibition or chromatin modifications at the target, guided by the small RNA.

Small RNAs are sorted into Argonaute proteins by different mechanisms in plants and animals. In *Drosophila*, miRNAs are sorted into Ago1 complexes and siRNAs are sorted into Ago2 complexes (Forstemann et al. 2007; Tomari et al. 2007). This sorting is dependent on the inherent structure of the small RNA duplex. miRNAs typically contain mismatches that allow them to be selectively incorporated by Ago1 and selectively rejected by Ago2. Conversely, siRNA duplexes are typically perfect matches and this

degree of complementarity is favored by Ago2 but not Ago1 (Forstemann et al. 2007; Tomari et al. 2007). In *Arabidopsis*, 10 Argonautes can potentially form effector complexes. In contrast to animals, it appears that the 5' nucleotide of the small RNA dictates assortment into a particular AGO (Mi et al. 2008). AGO1 binds miRNAs that have a 5' uracil. AGO5 binds siRNAs that start with cytosine. AGO2 and AGO4 both prefer siRNAs that begin with adenosine (Mi et al. 2008).

The discovery that production of dsRNA, Dicer cleavage and incorporation of small RNAs into Argonaute complexes (or RISC-like complexes) is common to a variety of RNA silencing phenomenon provides a basic platform for modeling RNA-mediated gene silencing in diverse organisms.

RNAi in yeast

S. pombe is the one of the best-studied models for RNAi because of the simplicity in the number of its silencing components and the ease with which they can be manipulated and purified. *S. pombe* has one Dicer, one Ago and one RNA-dependent RNA polymerase (RdRP) protein (Volpe et al. 2002). Several loci serve as models for studying the mechanism of RNAi-induced heterochromatin formation, including the telomeres, the mating type loci and the outer centromere repeats (White and Allshire 2008). These loci share a pair of similar repeats that are thought to be critical for initiating silencing, known as the *dg* and *dh* repeats (Takahashi et al. 1992; Steiner et al. 1993). Of these loci, the outer centromere repeats have served as a model for RNAi-mediated gene silencing because there is an absolute requirement for Dicer, Ago and RdRP activity in establishing and maintaining silencing at this region (Volpe et al. 2002).

Silenced endogenous loci and reporter genes targeted to silenced regions, such as the outermost centromere repeats, have served as invaluable markers for loss of silencing (Allshire et al. 1995).

Silencing is carried out by the RNAi machinery in conjunction with several heterochromatin-associated proteins. Histone deacetylases such as Clr3, Clr6 and Sir2 deacetylate histone tails, which allows for H3K9 methylation by the H3K9 methylase, Clr4 (Grewal et al. 1998; Grewal 2000; Rea et al. 2000; Shankaranarayana et al. 2003). H3K9 methylation by Clr4 provides a binding platform for the Heterochromatin Protein 1 (HP1) homolog, Swi6 (Bannister et al. 2001; Nakayama et al. 2001), which helps nucleate the heterochromatin state. Interestingly, Clr4 can also be recruited by Swi6, resulting in a self-reinforcing mechanism for the spreading of silencing within heterochromatin domains. Mutations in Clr4, Clr3, Clr6 and Sir2 lead to loss of silencing and/or mislocalization of Swi6 (Ekwall and Ruusala 1994; Thon et al. 1994; Ekwall et al. 1996; Grewal et al. 1998).

Current models propose that many of the above-mentioned players are part of a self-propagating silencing cycle, illustrated best at the centromere repeats (reviewed in (Moazed et al. 2006; Zofall and Grewal 2006; White and Allshire 2008)). Low-level transcription of one strand of the *dg* and *dh* repeats by Pol II is hypothesized to produce RNA that recruits the RITS complex via siRNA-transcript base pairing (Motamedi et al. 2004; Buhler et al. 2006). The RITS complex is comprised of outer centromere repeat siRNAs, Ago, Chp1 (a chromodomain protein), and Tas3, a protein of unknown function (Ekwall 2004a; Verdel et al. 2004). The RDRC complex, comprised of an RNA-dependent RNA polymerase (Rdp1), a poly-A polymerase (Cid12) and a helicase (Hrr1),

is required for producing dsRNA and is recruited by RITS via a physical interaction (Motamedi et al. 2004). Dicer uses the dsRNA as a substrate to produce siRNAs, which are incorporated into the RITS complex (Colmenares et al. 2007). In addition to targeting RDRC to chromatin, RITS uses the siRNAs generated by Dicer as a guide to target heterochromatin modifications in *cis* through Clr4 and Swi6, but the precise mechanism is unknown (Noma et al. 2004). Knocking out any of these genes or abolishing their activities prevents the nucleation of heterochromatin at a reporter gene, suggesting that these proteins are dependent on each other to establish and maintain silencing (Volpe et al. 2002; Motamedi et al. 2004; Verdel et al. 2004; Buhler et al. 2006; Irvine et al. 2006). Tethering of the RITS complex to an unsilenced reporter gene is sufficient to induce the silencing of that reporter gene *in cis*, further suggesting that all of these processes likely occur in *cis* at the target locus (Buhler et al. 2006).

Transcription by Pol II is required to initiate silencing because several mutants in Pol II subunits are viable but lead to loss of silencing of a reporter gene or at the centromere repeats. A mutation in Rpb2, the second largest subunit of Pol II, results in decreased H3K9 methylation, loss of siRNAs and increase in the stability of certain centromere transcripts that are normally degraded very quickly in wild-type cells (Kato et al. 2005). Likewise, an *rpb7* mutant is also defective in siRNA production at the centromere repeats, but this is achieved by blocking Pol II transcription from the start, therefore precluding any substrate that could be funneled into the RNAi pathway (Djupedal et al. 2005).

RNAi in *Drosophila*

Compared to yeast, *Drosophila melanogaster* has an expanded repertoire of the RNAi machinery core, including five Ago proteins (Ago1, Ago2, Ago3, Piwi and Aubergine) and two Dicer proteins (Dcr-1 and Dcr-2) (reviewed in (Huisinga and Elgin 2009)). However, *Drosophila* lacks an identifiable RdRP, suggesting that dsRNA substrates are produced in the absence of this enzyme. In addition to these core factors, *Drosophila* employs a third Dicer-like enzyme, Drosha, which participates in the initial step of miRNA precursor processing (Denli et al. 2004). Several RNA helicases, namely Homeless, Armitage and Lip, alos participate in silencing to varying degrees (Csink et al. 1994; Cook et al. 2004; Pal-Bhadra et al. 2004; Tomari et al. 2004). R2D2 and Loquacious are dsRNA binding proteins that facilitate the processing of siRNAs or miRNAs, respectively (reviewed in (Riddle and Elgin 2008)).

Three major small RNA pathways have been characterized in *Drosophila*. These include a canonical microRNA pathway (miRNA), a PIWI-associated siRNA pathway (piRNA), and a canonical siRNA pathway that acts in PTGS. In the miRNA pathway, RNA polymerase II transcribes loci that adopt hairpin structures that are recognized by the RNaseIII-like enzyme, Drosha (Denli et al. 2004). Drosha, in cooperation with the dsRNA binding protein Pasha, cleaves the primary miRNA transcripts in the nucleus into a pre-miRNA (Denli et al. 2004). After nuclear export, Loquacious helps mediate Dcr-1 cleavage of the pre-miRNA into 22-nt mature miRNAs in the cytoplasm (Denli et al. 2004; Lee et al. 2004; Saito et al. 2005). Then, one strand of the miRNA duplex is typically incorporated into an Ago1 effector complex to mediate either transcript

cleavage if the miRNA has high complementarity to the target, or translation repression in cases of mismatches between the miRNA and its target (Okamura et al. 2004).

Recently, a germ-line specific class of small RNAs, or piRNAs (Piwi-interacting small RNAs), has been identified in *Drosophila*. piRNAs correspond to repeats and transposable element sequences and associate with Piwi, Aub and Ago3 in a Dicer-independent manner (Brennecke et al. 2007). Their biogenesis, therefore, has been hypothesized to be dependent on the Piwi proteins themselves, which are capable of RNA cleavage (Brennecke et al. 2007; Gunawardane et al. 2007). In this way, piRNAs self-regenerate and post-transcriptionally degrade their targets simultaneously. piRNAs are important because they serve as an epigenetic mechanism to repress promiscuous transposition and maintain genome integrity (Aravin et al. 2007; Brennecke et al. 2008). Another potential role for piRNAs might be in mediating the timely establishment of heterochromatin early in development, but evidence is limiting at this point (Brennecke et al. 2007; Riddle and Elgin 2008; Huisinga and Elgin 2009).

The third small RNA pathway in *Drosophila* is the siRNA pathway directed against endogenous transposons, sense/anti-sense transcript pairs and viral RNA sequences. In this pathway, Dcr-2 partners with Loquacious or a second dsRNA binding protein, R2D2, to produce 21-nt siRNAs from dsRNA (Liu et al. 2003; Lee et al. 2004). This contrasts with piRNA biogenesis, which uses a piRNA-Ago complex to cleave ssRNA precursors. Ago2 is required to bind siRNAs and mediate target RNA destruction (Okamura et al. 2004; van Rij et al. 2006; Czech et al. 2008; Kawamura et al. 2008).

RNAi and heterochromatin formation in Drosophila

Drosophila is an excellent model for studying heterochromatin formation and maintenance. To this end, position effect variegation (PEV) is a very useful tool for such studies (reviewed in (Schotta et al. 2003; Girton and Johansen 2008)). PEV occurs when a normally euchromatic gene is placed in or near a heterochromatic environment such that the expression of the euchromatic gene now is under the control of the heterochromatin domain, which can result in silencing (Girton and Johansen 2008). J.H. Muller documented the first case of PEV with the white gene in 1930 (Muller 1930). X-ray mutations caused a rearrangement of the X chromosome, which resulted in the white gene residing in proximity to heterochromatin (Muller 1930). Affected flies had stochastic but heritable silencing of the white gene, resulting in a mottled red- and white-eye phenotype (Muller 1930).

PEV of a gene can result from a chromosomal rearrangement or by insertion of a gene into heterochromatin using a P-element (Girton and Johansen 2008). Genes subjected to PEV are silenced in a stochastic manner, leading to variegated phenotypes where some cells retain expression of the gene and others have silenced the gene. PEV of the *white* gene in the fly eye is a widely used marker because the eye phenotypes can be scored and quantified. A gradient of white to orange to red can be observed in different eye cells depending on the degree of silencing.

PEV has been associated with altered chromatin structure, providing early evidence for chromatin-based regulation (Wallrath and Elgin 1995). Screens have identified many factors that either suppress or enhance PEV (reviewed in (Girton and Johansen 2008)). Suppressors of variegation, or Su(var)s, encode components required

for silencing, whereas Enhancers of variegation, or E(var)s, are components that normally repress silencing. HP1 (the Swi6 homolog) and Su(var)3-9 (the H3K9 methyltransferase) both suppress PEV and are required for heterochromatin formation (James and Elgin 1986)(Reuter et al. 1986). HP1 and Su(var)3-9 act to nucleate and spread silencing marks in a self-reinforcing manner similar to what is observed in yeast with Swi6 and Clr4 (Bannister et al. 2001). Among the RNAi players listed above, Ago2, Piwi, Aubergine, Lip and Homeless mutants all suppress variegation (Csink et al. 1994; Pal-Bhadra et al. 2004; Deshpande et al. 2005). Mutants in these genes also impact heterochromatin structure, linking RNAi and heterochromatin formation in higher eukaryotes (Pal-Bhadra et al. 2004; Deshpande et al. 2005; Haynes et al. 2006).

The importance of maintaining a functional RNAi system and heterochromatin integrity in *Drosophila* is illustrated by mutations in many components comprising different pathways. Whereas *dcr-2* mutants are viable, *dcr-1* mutants are lethal (Lee et al. 2004; Meyer et al. 2006). Ago1 and Ago2 act redundant to control embryo patterning and cell polarity (Lee et al. 2004; Meyer et al. 2006). *ago2* mutants affect early embryo development by causing aberrations in chromosome condensation and spindle assembly (Deshpande et al. 2005). RNAi components that localize to the germline, such as Piwi, Aubergine, Homeless and Armitage, result in female sterility (reviewed in (Huisinga and Elgin 2009)). In addition, HP1, Piwi, Aubergine, Homeless, Dcr-2 and Su(var)3-9 mutants lead to extrachromosomal DNA accumulation and variation in the number of nucleoli (Peng and Karpen 2007).

RNAi in mammals

Both post-transcriptional and transcriptional gene silencing (PTGS and TGS) mechanisms appear to be conserved in mammals and plants. The first evidence for PTGS in humans came in 2001 with the transfection of HeLa and embryonic kidney cells with small dsRNAs complementary to reporter genes or endogenous targets, which decreased both the activity of the reporter genes and protein levels of the endogenous genes (Elbashir et al. 2001a). A role for RNAi in directing heterochromatin modifications and gene silencing has been observed in humans, although much less is known about the mechanisms in humans compared to other organisms. siRNAs delivered to the nucleus of human cells can trigger TGS that is associated with DNA methylation and/or histone modifications (Morris et al. 2004; Castanotto et al. 2005; Weinberg et al. 2006). Histone methylation appears to require both siRNAs and Ago1 (Morris et al. 2004; Janowski et al. 2006). In addition, Dicer-defective chicken cells exhibit increased transcription of satellite repeats and mislocalization of heterochromatin-associated proteins (Fukagawa et al. 2004). In humans, both Ago1 and Ago2 are required for transcriptional and post-transcriptional gene silencing (Janowski et al. 2006), linking these two silencing mechanisms in human cells and suggesting there is conservation of the RNAi mechanism in humans, yeast, *Drosophila* and plants. Ago2 appears to be the predominant Ago associated with PTGS, given its catalytic ability to cleave targeted mRNAs or participate in miRNA induced translation inhibition (Meister et al. 2004). Interestingly, Ago1 interacts with RNA polymerase II in a transcription-dependent manner (Kim et al. 2006; Weinberg et al. 2006). Models have proposed that transcription by Pol II could recruit Ago1-siRNA complexes as well as other factors such as the DNA methyltransferase

DNMT3a, the histone deacetylase HDAC-1 and Suv39H1, a histone methyltransferase (Morris 2008). Recruitment of these factors would then result in histone and DNA methylation of the target locus. This model is very similar to one proposed for silencing in *S. pombe*, where transcription by Pol II recruits the RDRC and RITS complexes to initiate silencing.

RNAi in plants

In contrast to yeast, which has a single Dicer, RNA-dependent RNA polymerase and Argonaute, *Arabidopsis* has extensive functional diversification of the core proteins involved in silencing. *Arabidopsis* has four Dicer proteins (DCL1-DCL4), ten Argonautes (AGO1-AGO10) and six RNA-dependent RNA polymerases (RDR1-RDR6).

Studies have attributed functions to all four Dicers, three of the six RDRs, and at least five of the ten AGOs to date. DCL2, DCL3 and DCL4 are redundant in some contexts and act primarily to produce siRNAs corresponding to endogenous transcripts, transgenes and viruses to different extents (Xie et al. 2004; Dunoyer et al. 2005; Gasciolli et al. 2005; Xie et al. 2005; Blevins et al. 2006). DCL1 is mainly involved in microRNA production (Park et al. 2002; Kurihara and Watanabe 2004). AGO1 is a slicing endonuclease that cleaves miRNA and trans-acting siRNA target genes as well as transgenes (Baumberger and Baulcombe 2005). There is partial redundancy between two AGO proteins, AGO4 and AGO6, in performing RNA-directed chromatin modifications (Zheng et al. 2007). AGO4 acts independently as both a slicer and as a chromatin-modifying effector complex (Qi et al. 2006). AGO7 is involved in the targeting of trans-acting siRNAs to their targets and for targeting viral RNAs (Hunter et al. 2003; Adenot et

al. 2006; Qu et al. 2008). AGO10 is required for maintaining the pool of meristem cells (undifferentiated cells similar to stem cells) in the shoot apical meristem and for regulating leaf polarity (Moussian et al. 1998; Lynn et al. 1999; Liu et al. 2008).

Of the six RDRs, three are tightly linked and have no reported function, namely RDR3, RDR4 and RDR5. RDR1 plays a role in the accumulation of siRNAs directed against some viruses (Xie et al. 2004; Diaz-Pendon et al. 2007; Donaire et al. 2008). RDR6 is involved in production of dsRNAs in the trans-acting siRNA pathway, the natural anti-sense siRNA pathway and in silencing of transgenes and viruses (Dalmay et al. 2000; Peragine et al. 2004; Allen et al. 2005; Borsani et al. 2005; Katiyar-Agarwal et al. 2006; Diaz-Pendon et al. 2007). RDR6 also functions in the spreading of silencing over long distances (Brosnan et al. 2007). RDR2 is the main RDR involved in the amplification of RNAs that are diced to mediate chromatin modifications (Xie et al. 2004). RDR2 is also required for short-range spreading of silencing and long-distance reception of silencing between cells (Brosnan et al. 2007; Dunoyer et al. 2007; Smith et al. 2007). RDR6 is the only *Arabidopsis* RDR for which activity has been demonstrated *in vitro* (Curaba and Chen 2007).

In addition to these core components, there is a plethora of other silencing factors, including chromatin remodeling proteins, double-stranded RNA binding proteins (DRBs), RNA helicases, exonucleases, RNA methylases, DNA and histone methyltransferases, histone deacetylases, DNA demethylases and proteins of unknown function, among others. Specific mention of these latter components will occur in their relative context throughout the Introduction. As the Introduction will later elaborate,

many of these factors are not confined to a role in a single pathway, but may be found in multiple pathways and have multiple genetic partners.

Two of the most interesting components of RNAi in *Arabidopsis* are the plant-specific RNA polymerases IV and V. These polymerases play non-redundant roles in mediating siRNA production, DNA methylation and gene silencing in multiple silencing pathways directed against endogenous genes and transgenes (Chapter 2)(Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). Because work on these two protein complexes serves as the focus of my thesis, the RNAi silencing pathways of *Arabidopsis* will be discussed below. Later in the Introduction, I will follow up with a review of the specific roles of Pol IV and V in some of these pathways.

MicroRNA pathway

microRNAs (miRNAs) are small RNAs that are produced from transcribed, non-coding loci in the genome. miRNAs function to regulate gene expression of developmentally important genes at the post-transcriptional level through target mRNA cleavage or translational inhibition. Plants primarily cleave their miRNA targets because the miRNAs generally display a high degree of complementarity to their targets, although there are cases where translation inhibition occurs, which is the primary mechanism of miRNA action used by animals. miRNAs were first discovered as “small temporal RNAs” in *C. elegans* (Lee et al. 1993). Later work in plants revealed that miRNAs are a conserved mechanism of gene regulation across kingdoms (Lee et al. 1993; Aukerman and Sakai 2003; Palatnik et al. 2003). In *Arabidopsis*, 187 miRNA genes have been identified using a combination of forward genetics, small RNA cloning and

bioinformatics prediction (Chen 2008)(miRBASE: http://microrna.sanger.ac.uk/). Conservation between plant and animal miRNAs has been detected but not yet confirmed, suggesting that miRNA function has a deep evolutionary history (Chen 2008). Interestingly, however, there are also families of miRNAs in *Arabidopsis* that have no apparent homologs in other flowering plants, such as rice and Poplar, suggesting that some miRNA families are evolving rapidly (Chen 2008).

miRNA precursor transcripts are typically generated by Pol II and form an imperfect hairpin structure called the primary miRNA, or pri-miRNA. The pri-miRNA is processed into a shorter pre-miRNA and finally to a 21-nt imperfect double-stranded miRNA. Processing in animals typically occurs through the action of the RNase-III enzyme, Drosha, to generate the pre-miRNA (Denli et al. 2004). Dicer subsequently cleaves the pre-miRNA into the mature miRNA duplex. In plants, however, both steps appear to be performed by DCL1 in concert with a dsRNA binding protein, HYL1 (HYPONASTIC LEAVES 1), which interacts with DCL1 and is thought to stabilize the cleavage of the pri-miRNA by DCL1 (Kurihara and Watanabe 2004; Kurihara et al. 2006). HYL1 probably acts analogously to PASHA, a related dsRNA binding protein, in the animal Microprocessor complex (Denli et al. 2004). A zinc finger protein, SE (SERRATE), interacts with HYL1 and also assists with the initial miRNA cleavage events (Lobbes et al. 2006; Yang et al. 2006a). The final miRNA duplex becomes methylated on the 2' OH of the 3' terminal nucleotide of the mature miRNA through the RNA methylase activity of HEN1 (HUA ENHANCER 1) (Li et al. 2005; Yu et al. 2005; Yang et al. 2006b). Methylation of the miRNA is thought to stabilize the miRNA and prevent poly-uridiylation of the 3' miRNA, which is observed in *hen1* mutants (Li et al.

2005). Indeed, almost all siRNAs and miRNAs appear to be methylated by HEN1, suggesting that this mechanism is critical for the action of small RNA function (Li et al. 2005).

After methylation by HEN1, one strand of the miRNA duplex is incorporated into a RISC-effector complex. In *Arabidopsis*, AGO1 is the primary miRNA-RISC that binds miRNA and results in target mRNA cleavage (Baumberger and Baulcombe 2005; Qi et al. 2005). However, it is possible that other AGOs, particularly AGO4, have a small role in binding miRNAs (Qi et al. 2006). miRNA function occurs post-transcriptionally, leading primarily to target mRNA cleavage or translational inhibition, which occur in the cytoplasm. Consistent with this, miRNA accumulation and function requires export from the nucleus to the cytoplasm by HST (HASTY), a homolog of the mammalian exportin 5 (Park et al. 2005).

Trans-acting siRNA pathway

Trans-acting siRNAs constitute a novel siRNA pathway in plants (Peragine et al. 2004; Vazquez et al. 2004). This pathway is initiated with the production of a single miRNA species that is dependent upon AGO1, DCL1, HEN1 and HYL1 (Allen et al. 2005). The miRNA then targets a trans-acting siRNA gene locus, or *TAS* gene. The initial miRNA facilitates cleavage of the TAS RNA followed by subsequent conversion of the cleaved *TAS* RNA into dsRNA by RDR6, with the assistance of SGS3 (SUPPRESSOR OF GENE SILENCING 3), a plant-specific coiled-coil protein of unknown function (Allen et al. 2005). One of the dsRNA ends serves as the initiation point for phased dicing by DCL4 (Gasciolli et al. 2005; Xie et al. 2005). DCL4 is

stabilized by the dsRNA binding protein DRB4, analogous to R2D2 or Loquacious in *Drosophila* (Adenot et al. 2006; Nakazawa et al. 2007). The resulting cascade of trans-acting siRNAs produced by DCL4 then incorporate into AGO7 or AGO1 effector complexes and guide the cleavage of their target mRNAs (Baumberger and Baulcombe 2005; Adenot et al. 2006). The name “trans-acting siRNA” is derived from the observation that the tasiRNAs often target genes that share little or no homology with the *TAS* genes, from which the tasiRNAs are generated (Vazquez et al. 2004). To date, only a handful of *TAS* genes have been discovered, but each gene is capable of producing multiple, distinct tasiRNAs. Downstream targets of the tasiRNAs include auxin-related transcription factors and members of a large family of pentatricopeptide repeat (PPR) proteins (reviewed in (Poethig et al. 2006)). The importance of the trans-acting siRNA pathway in plant development is exemplified by the phenotypes resulting from mutations in the *TAS3* gene, which confer an accelerated transition from juvenile to adult leaf development (Adenot et al. 2006). Likewise, *dcl4*, *rdr6*, *sgs3* and *ago7* mutations all display similar accelerations in development toward adult leaf features (Hunter et al. 2003; Yoshikawa et al. 2005; Poethig et al. 2006).

Given the small number of tasiRNAs identified to date, it is unclear whether tasiRNAs are a widespread mechanism for gene regulation, or if bioinformatics and deep sequencing approaches simply have not identified all the existing tasiRNAs in *Arabidopsis*. Furthermore, no such pathways have been reported in other metazoans, suggesting that tasiRNA-mediated gene regulation might be specific to the plant kingdom.

Abiotic and biotic stress-inducible siRNA pathways

miRNAs and tasiRNAs in *Arabidopsis* provide a critical means for regulating plant development. However, it has become evident that other types of siRNAs also regulate plant responses to the environment, complementing the actions of the miRNAs and tasiRNAs. Such diversity in small RNA function suggests that plants have evolved a series of related but distinct RNAi mechanisms for regulating growth and development.

Two emerging classes of siRNAs are the abiotic and biotic stress-induced siRNAs, or natural anti-sense siRNAs. These siRNAs regulate the expression of the overlapping gene pairs from which they are derived. The biogenesis of abiotic and biotic stress-inducible siRNAs is regulated by components of the siRNA and miRNA pathways. Pol IV, RDR6, DCL1, HEN1 and HYL1 are the core constituents of this pathway (Borsani et al. 2005; Katiyar-Agarwal et al. 2006; Katiyar-Agarwal et al. 2007). DCL4, DCL2, SGS3, SDE3 (an RNA helicase), AGO7, Pol V and HST1 are also involved but act in a locus-dependent manner (Borsani et al. 2005; Katiyar-Agarwal et al. 2006; Katiyar-Agarwal et al. 2007). Notably, these pathways use Pol IV and Pol V but not DCL3 or RDR2, all of which play prominent roles in the 24-nt siRNA directed heterochromatin pathway (reviewed below).

The first abiotic-stress inducible siRNA identified is derived from two genes, *SRO5* and *P5CDH,* whose 3' ends overlap on opposite strands. Under normal conditions, only *P5CDH* is expressed. Under salt stress conditions, *SRO5* expression is induced, resulting in transcription from both loci. The resulting transcripts are thought to anneal to produce dsRNA from their complementary regions. Through an unknown mechanism, a single

Pol IV-dependent 24-nt siRNA is produced to guide the phased cleavage of *P5CDH* transcripts (Borsani et al. 2005). Down-regulation of *P5CDH* results in increased proline synthesis, a key response to salt tolerance. However, lower P5CDH levels in salt stressed plants also triggers an increase in the number of reactive oxygen species. To prevent oxidative stress, SRO5 quenches the effects of the higher reactive oxygen species during salt stress.

Pathogen-inducible siRNAs constitute a novel class of biotic stress-induced siRNAs that positively regulate plant pathogen resistance. In one case, a 39-41-nt long siRNA, or lsiRNA, is induced upon infection by the *Pseudomonas syringae* DC3000 strain carrying the avrRpt2 effector (Katiyar-Agarwal et al. 2007). This lsiRNA is complementary to the overlapping region of two opposing genes (*SRRLK* and *AtRAP*), but specifically down-regulates *AtRAP*, a negative regulator of plant basal defense response, resulting in increased resistance to the pathogen.

Interestingly, these stress response strategies may be widespread among kingdoms because pairs of overlapping, anti-sense-oriented genes appear relatively frequently in genomes (Werner and Berdal 2005; Numata et al. 2007). In *Arabidopsis*, more than a thousand gene pairs overlap in an anti-sense orientation (Wang et al. 2004; Jen et al. 2005; Wang et al. 2005; Henz et al. 2007; Jin et al. 2008). Most of these pairs have a small RNA that maps to at least one of the overlapping loci, suggesting a significant role for small RNA regulation in regulating expression of natural anti-sense gene pairs (Jin et al. 2008).

siRNA-directed heterochromatin formation

The pathways outlined above regulate gene expression using small RNAs at the post-transcriptional level. In addition, plants have also evolved elaborate mechanisms using siRNAs to modify chromatin environments, similar to yeast, flies and humans, which are important for maintaining genome integrity and for regulating gene expression at the transcriptional level. Similar to PTGS, small RNAs are produced by Dicers and are incorporated into AGO-effector complexes, but in contrast to PTGS the end readout is a modification at the DNA or histone level. In plants, siRNAs direct DNA methylation as well as histone modifications, such as H3K9 methylation.

Similar to yeast, the biogenesis of heterochromatic siRNAs in *Arabidopsis* involves the concerted action of a specific set of core silencing components, namely Pol IV, DCL3, RDR2 and AGO4. In *S. pombe*, Pol II-generated non-coding RNA transcripts are the initial triggers for making dsRNA and subsequent heterochromatic siRNAs (Kato et al. 2005; Buhler et al. 2006). However, the role of Pol II in plant silencing is not clear. Surprisingly, genes encoding catalytic subunits for two additional RNA polymerases have been found (*Arabidopsis* genome initiative (2000)). Genetic and biochemical approaches have determined that these catalytic subunits form two non-redundant, multi-subunit plant-specific polymerases involved in gene silencing, named RNA Polymerase IV and RNA Polymerase V (Pol IV and Pol V respectively), which in plants may be analogous to the role of RNA Polymerase II (Pol II) in gene silencing in yeast (Chapters 2-3)(Appendices A-B, D)(Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005; He et al. 2009; Huang et al. 2009; Ream et al. 2009).

Silencing of heterochromatin is proposed to begin by the generation of an RNA trigger that is produced by Pol IV from a DNA template, perhaps with the cooperation of a putative chromatin remodeling protein, CLSY1 (CLASSY1), which was identified in a screen for factors defective in the spreading of silencing (Pontes et al. 2006; Smith et al. 2007). RDR2 amplifies this transcript into dsRNA that is then processed by DCL3 into 24-nt siRNAs in a Cajal body-like siRNA processing center in the nucleolus (Appendix B) (Xie et al. 2004; Qi et al. 2005; Li et al. 2006; Pontes et al. 2006). This process is consistent with complete loss of 24-nt siRNAs and DNA methylation corresponding to target loci in *nrpd1*, *clsy1*, *nrpd2* and *rdr2* mutants (Xie et al. 2004; Herr et al. 2005; Kanno et al. 2005; Pontier et al. 2005; Pontes et al. 2006). These 24-nt siRNAs are incorporated into AGO4 effector complexes, or in some cases AGO6 complexes, which use the siRNA as a guide to mediate chromatin modifications to homologous regions of DNA or associated histones in cooperation with Pol V (Zilberman et al. 2003; Qi et al. 2006; Zheng et al. 2007). siRNA-directed chromatin modifications at the target loci require the action of the *de novo* DNA methyltransferase, DRM2, a second chromatin remodeling protein, DRD1, and the H3K9 methyltransferase, KRYPTONITE (KYP). AGO4, Pol V, DRM2 and DRD1 act downstream of siRNA biogenesis such that mutations in these genes that lead to changes in DNA methylation and histone modifications but little or no change in siRNA accumulation (Cao and Jacobsen 2002; Cao et al. 2003; Zilberman et al. 2003; Chan et al. 2004; Kanno et al. 2004; Kanno et al. 2005; Li et al. 2006; Pontes et al. 2006; Mosher et al. 2008).

The order of the pathway presented above is well supported by extensive cytogenetic studies in which mutants of the pathway were used to order the activity of

proteins relative to one another (Appendix B) (Li et al. 2006; Pontes et al. 2006). Pol IV and RDR2 act in the nucleoplasm and co-localize with sites such as the 5S rDNA loci that are targeted for RNA-directed DNA methylation (Pontes et al. 2006). Consistent with acting early in the pathway, NRPD1, NRPD2 or RDR2 localization is not disrupted in mutants of *dcl3*, *ago4* and *nrpe1* (Pontes et al. 2006). DCL3, RDR2 and AGO4 localization occurs within a Cajal body-like structure in the nucleolus, in addition to some nucleoplasmic localization. Fittingly, siRNAs co-localize with these proteins in the nucleolus, suggesting that siRNA processing and even RISC loading may occur in the nucleolus. Localization of DCL3, RDR2 and AGO4 is severely disrupted by mutations in *nrpd1* or *nrpd2*, the largest and second-largest catalytic subunits of Pol IV (Li et al. 2006; Pontes et al. 2006). Likewise, AGO4 localization is also disrupted by *dcl3* and *rdr2* mutations, consistent with the idea that AGO4 activity requires DCL3 and RDR2 activity for either localization or stabilization of the siRNA-AGO4 effector complex (Li et al. 2006). AGO4 protein levels, but not transcripts, are dependent on NRPD1 and RDR2, suggesting that AGO4 is unstable when siRNAs are not present (Li et al. 2006) (Wierzbicki et al. 2009). Acting downstream to coordinate chromatin modifications, Pol V and DRD1 localize to 5S rDNA genes where chromatin modifications are targeted (Pontes et al. 2006).

Interestingly, the siRNA machinery (DCL3, AGO4, RDR2 and NRPE1) and the miRNA machinery (DCL1, HYL1 and SE) localize to Cajal body-like structures in the nucleolus (Fang and Spector 2007; Fujioka et al. 2007; Song et al. 2007). miRNA precursors also co-localize with the miRNA processing machinery, such as DCL1, suggesting that miRNA processing may also occur in the nucleolus (Fang and Spector

2007; Fujioka et al. 2007). It has been proposed that these plant nucleolar bodies may share some functions with animal P-bodies. P-bodies are where siRNA-mediated RNA degradation and miRNA-mediated translational inhibition occur in the cytoplasm (Pontes and Pikaard 2008).

Physical interaction between AGO4 and Pol V is an important step of the 24-nt heterochromatic siRNA pathway. This interaction is mediated by a conserved WG/GW motif present in the C-terminal domain (CTD) of NRPE1, the largest subunit of Pol V (Li et al. 2006; El-Shami et al. 2007). WG/GW motifs are conserved in proteins that interact with AGO proteins in yeast (TAS3), *Drosophila* (GAWKY) and humans (GW182, TNRC6B) (Jakymiw et al. 2005; Behm-Ansmant et al. 2006; Partridge et al. 2007), suggesting that this motif, known as an "AGO-hook", is an AGO docking platform. The robustness of the interaction between AGO4 and Pol V is debatable because the interaction *in vivo* has only been demonstrated in one direction (Li et al. 2006). In addition, AGO4 was not identified in affinity purified Pol V complexes by two independent labs (Huang et al. 2009; Ream et al. 2009; Wierzbicki et al. 2009).

Recent analyses have identified a role for an SPT5-like transcription factor, also known as KTF1, in the 24-nt heterochromatic siRNA pathway (Bies-Etheve et al. 2009; Huang et al. 2009). KTF1/SPT5-like peptides were identified in affinity purified Pol V, although secondary confirmation of this interaction is lacking (Huang et al. 2009). In addition, AGO-hook motifs (WG/GW repeats) are present in KTF1/SPT5-like, which interacts with AGO4 (Bies-Etheve et al. 2009). *ktf1/spt5-like* mutants display a reduction in DNA methylation and siRNAs at some Pol V-dependent loci, similar to mutations in the *nrpe5* subunit of Pol V (Bies-Etheve et al. 2009; Huang et al. 2009).

One proposed role of KTF1/SPT5-like is that it acts as a transcription factor that modulates the activity or specificity of Pol V (Bies-Etheve et al. 2009). This is reminiscent of the role of Spt5 proteins in other organisms, which are required for the elongation activity of Pol II (Hartzog et al. 1998; Wada et al. 1998; Sims et al. 2004).

The roles of plant-specific RNA polymerases IV and V

Identification of the Pol IV and Pol V using forward genetics

Alleles corresponding to the largest and second largest catalytic subunits of Pol IV (NRPD1, NRPD2/NRPE2) and Pol V (NRPE1, NRPE2/NRPD2), as well as the homolog of yeast Rpb4 shared by Pol IV and Pol V (NRPD4/NRPE4), have been isolated from genetic screens that assayed for factors required in maintaining silencing of a reporter gene. Herr et al. screened for reactivation of a GFP marker that was initially silenced by a *Potato virus X-GFP* transgene. In their screen, they identified mutant alleles of *NRPD1* and used reverse genetics to reveal the requirement for the *NRPD2* gene in silencing (Herr et al. 2005). Both siRNAs and DNA methylation corresponding to the reporter gene were lost in the *nrpd1* and *nrpd2* mutants, providing some of the first evidence that Pol IV subunits are required for RNA-directed DNA methylation and siRNA-mediated gene silencing (Herr et al. 2005). In addition, transcripts corresponding to a SINE retroelement, *AtSN1*, were up-regulated in *nrpd1* and *nrpd2* mutants, suggesting that Pol IV subunits are necessary for maintaining silencing of potentially harmful genetic elements.

In a similar screen, Kanno et al. used a seed-specific promoter driving a GFP reporter gene that was silenced by an inverted repeat corresponding to the seed-specific

promoter. In their screen, they identified the largest and second largest subunits of Pol V, NRPE1 and NRPE2. Notably, NRPE2 and NRPD2 correspond to the same protein, which is used by Pol IV and V. The *nrpe1* and *nrpe2* mutants displayed loss of RNA-directed DNA methylation at the seed specific promoter, but siRNAs still accumulated (Kanno et al. 2005). This evidence suggested that Pol V was not required for siRNA production but was essential for RNA-directed DNA methylation. In this screen, no alleles for *RDR2* or *NRPD1* were identified, likely because these components are not required for silencing at this locus because the inverted repeat is sufficient for making enough dsRNA to feed into the silencing pathway. This idea fits with the proposed pathway where Pol IV and RDR2 act upstream of siRNA biogenesis and Pol V acts downstream of or independent of siRNA biogenesis, but upstream of chromatin modifications such as DNA methylation (Appendix B)(Pontes et al. 2006). Transcriptional up-regulation of transposable elements was also observed in *nrpe1* and *nrpe2* mutants.

A screen using a heavily methylated luciferase transgene, whose silencing is dependent on RNA-directed DNA methylation, identified a homolog of yeast *rpb4*, named *NRPD4/NRPE4* (He et al. 2009). NRPD4/NRPE4 interacts with Pol IV and Pol V and is required for siRNA production and RNA-directed DNA methylation of a variety of endogenous repeats and transposons, including the 5S rRNA genes and *AtSN1*-class retroelements, similar to mutations in other Pol IV and Pol V subunits (Chapter 3)(He et al. 2009; Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). However, the reduction of methylation and siRNAs is less severe than in mutants disrupting the

catalytic subunits of Pol IV and Pol V (Chapter 2)(Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005).

Reverse genetics analyses by two other labs trying to understand the role of these atypical polymerases in plant development also demonstrated the requirements for Pol IV and Pol V catalytic subunits in RNA-directed DNA methylation and siRNA production (Onodera et al. 2005; Pontier et al. 2005) (Chapter 1). Pontier et al. demonstrated that Pol IV and Pol V are non-redundant polymerases required for regulating flowering time. Molecular analysis attributed this phenotype to loss of silencing of a flowering repressor gene, *FWA*. In addition, they confirmed that Pol IV is required to produce siRNAs whereas Pol V is only partially required, whereas both polymerases are needed for RNA-directed DNA methylation (Pontier et al. 2005). They further demonstrated that the stability of the NRPD2/NRPE2 subunit was dependent on the NRPE1 subunit, but not the NRPD1 subunit, providing an alternative explanation for why siRNAs decrease in *nrpe1* mutants. Pol IV and Pol V eluted as fairly large molecular weight complexes from a gel filtration column, arguing that they exist as multi-subunit RNA polymerases, similar to the canonical eukaryotic RNA polymerases I, II and III (Pontier et al. 2005).

Roles of Pol IV and Pol V in siRNA production

Pol IV is required for producing nearly 95% of the siRNAs in *Arabidopsis*. Of these, Pol IV is essential for the 24-nt siRNA size class that directs heterochromatin modifications and silencing of endogenous repeats and transposons, especially within peri-centromeric regions where these elements are enriched (Mosher et al. 2008). Consistent with earlier reports, Pol V is less important—at some loci Pol V is required for

siRNA production, but there are many examples where siRNAs are produced independently of Pol V. In no cases are there siRNAs that are independent of Pol IV but are Pol V-dependent. Therefore, Pol IV is intimately linked to siRNA production and the corresponding DNA methylation readout. The role of Pol V in siRNA production is locus-dependent, but Pol V is still capable of directing DNA methylation in the absence of siRNAs in some cases (Mosher et al. 2008).

Pol IV and Pol V in the spreading of silencing

Spreading of silencing can be divided into two major classes—spreading along the chromosome and spreading between cells. In the first case, spreading of silencing can occur from a site of nucleation along the chromosome to silence adjacent genes. This type of spreading is often cell autonomous and is best highlighted by position effect variegation in yeast and flies. This type of spreading of silencing in plants has been observed using transgenes and, in recent cases, in silencing of endogenous genes (Saze and Kakutani 2007; Daxinger et al. 2009; Matzke et al. 2009). One form of this silencing spread is termed "transitivity" because secondary siRNAs are generated to regions outside of the initial silenced region (Vaistij et al. 2002). A feature of silencing in a GFP reporter screen for RNA-directed DNA methylation mutants by the Matzke lab is transitivity that spreads downstream of the initial target region to induce DNA methylation (Kanno et al. 2008; Daxinger et al. 2009). The secondary siRNAs and corresponding DNA methylation resulting from this transitivity are Pol IV, Pol V and RDR2 dependent. In another case, an endogenous gene appears to undergo spreading of silencing via DNA methylation and siRNAs, and the siRNAs in this case are Pol IV and

Pol V dependent (Henderson and Jacobsen 2008). The criteria for why some loci experience transitivity and why transitivity spreads 3' to 5' vs. bi-directionally is unclear, but some models have suggested the number of primary small RNAs targeting a region has an influence in these mechanisms (Axtell et al. 2006; Moissiard et al. 2007).

The second major type of spreading involves the mobilization of the silencing signal between cells and across organs of an individual in a non-cell autonomous manner. This type of spreading is further divided into two forms—short-range spreading and long-range (systemic) spreading. These cases are best documented in plants and worms (Voinnet 2005). In plants, short-range silencing between cells is postulated to occur via the plasmodesmata, the cytoplasmic channels that connect individual cells (reviewed in (Voinnet 2005; Kalantidis et al. 2008)). Systemic silencing is initiated in one part of the organism and subsequently spreads throughout the entire organism. Systemic silencing is mediated through the phloem, or vascular system of the plant, where the signal moves in the direction of the phloem from source-to-sink tissues of the plant (i.e. from mature photosynthesizing leafs to emerging tissues that require extra nutrition) (Tournier et al. 2006).

Several elegant genetic screens have sought to identify factors involved in short-range cell-to-cell spreading or systemic silencing through the phloem. Screens for factors required for short-range spreading of silencing have employed a phloem-specific promoter driving a hairpin targeted to an exogenous reporter gene or an endogenous marker gene. These screens have identified NRPD1, RDR2 and the chromatin remodeling protein, CLSY1, all part of the classical 24-nt heterochromatic siRNA pathway (Dunoyer et al. 2007; Smith et al. 2007). Interestingly, NRPD1 and RDR2

proteins are not required for the initial production of siRNAs, but for the transmission or reception of the silencing signal between cells via an unknown mechanism (Dunoyer et al. 2007). A screen to detect factors involved in long-range silencing utilized a GFP reporter-silencer system. In this screen, silenced plant rootstocks were grafted with un-silenced scions (aerial plant tissue) from different RNAi mutants. Grafting is a technique where two different plant halves can be fused together to make one plant. This experiment revealed that Pol IV, RDR2, DCL3 and to some degree AGO4 are all required for the perception and amplification of the silencing signal (Brosnan et al. 2007).

The nature of the long-distance silencing signal has remained elusive in plants, although RNA of some form is a leading candidate. Evidence arguing against siRNAs as the RNA signal comes from grafting experiments that show *dcl1* or *dcl2 dcl3 dcl4* mutants are not compromised in generating the initial silencing signal in the silenced rootstock (Brosnan et al. 2007). In addition, other work has shown that viral suppressors of silencing can bind and sequester multiple size classes of siRNAs without inhibiting the movement of the long-range silencer (Mallory et al. 2003). The short-range signal, however, may be siRNAs because *dcl4* mutants compromise the spread of silencing from cell to cell (Dunoyer et al. 2005). In conclusion, transitivity and cell-to-cell movement of silencing requires Pol IV for reception (or potentially transmission), but less is known about the generation of the long distance signal.

Role of Pol IV in paramutation

Paramutation is a classic epigenetic phenomenon first described by Brink in 1956 at the maize *R1* locus (Brink 1956). Paramutation is the conversion of an active allele to

an inactive allele that results from the exposure to an inactive allele. The allelic transition does not involve a change in DNA sequence and is heritable over many generations in the absence of the original inactive allele. If the converted allele is crossed with a naïve, or active allele, the converted allele is able to mediate the transformation of the active allele to an inactive allele. For decades, the molecular mechanism underlying paramutation was elusive. In the last few years, several groups have isolated mutations in genes required for paramutation. Alleles of *RDR2*, a *CLSY1* or *DRD1*-like protein, and *NRPD1* have been recovered, indicating that Pol IV-mediated silencing is involved in paramutation (Alleman et al. 2006; Hale et al. 2007; Erhard et al. 2009). Paramutation has been observed in several plants and in animals, indicating that this phenomenon may be widespread, although presently there are a limited number of examples to support this claim (reviewed in (Chandler et al. 2000; Chandler 2007)).

Pol V transcribes non-coding RNA

While *in vitro* biochemical activity of Pol V has not yet been demonstrated, *in vivo* efforts to identify Pol V-dependent transcripts have yielded insight into Pol V function. Hypothesizing that Pol V transcription may occur at very low levels compared to Pol I, II or III, Wierzbicki et al. examined loci in the *Arabidopsis* genome that had no detectable transcription using published gene expression data, but were enriched for DNA methylation and/or small RNAs that might be dependent on Pol V. This approach led to the identification of several loci in both euchromatic and heterochromatic regions that produce low abundance transcripts in wild-type plants that are abolished in *nrpe1* or *nrpe2* mutants, but persist in *nrpd1* mutants. These transcripts vary in their start sites,

lack poly-A tails and have 5' caps or triphosphates characteristic of capped or newly initiated RNAs, but not cleaved RNAs. These transcripts are dependent on the conserved Metal A active site of NRPE1 and can be cross-linked to NRPE1, suggesting that Pol V catalyzes their synthesis (Wierzbicki et al. 2008). NRPE1 transcription is also required for recruiting repressive chromatin modifications at the loci it transcribes and regulates, as indicated by loss of DNA methylation, H3K27 mono-methylation and H3K9 di-methylation at these loci in Pol V active site mutants (Wierzbicki et al. 2008; Haag et al. 2009). Interestingly, Pol V transcription does not require small RNA production, suggesting that Pol V transcription occurs independently of Pol IV, RDR2 and DCL3 action (Wierzbicki et al. 2008).

The chromatin remodeling protein, DRD1, and a structural maintenance of chromosomes (SMC) protein, DMS3, are required for Pol V transcription (Wierzbicki et al. 2008) (Wierzbicki et al. 2009). DMS3 was identified in a screen for mutants defective in RNA-directed DNA methylation and has limited homology to a domain of cohesin (Kanno et al. 2008). In *drd1* and *dms3* mutants, Pol V is not efficiently recruited to the loci it transcribes, suggesting that DRD1 and DMS3 play a role in facilitating Pol V recruitment, either by opening the chromatin template or mediating interaction with other proteins that recruit Pol V (Wierzbicki et al. 2008) (Wierzbicki et al. 2009).

Interestingly, AGO4 appears to antagonize Pol V transcription because Pol V transcripts accumulate at higher levels in *ago4* mutants (Wierzbicki et al. 2009). These data show that Pol V does not require AGO4 for transcription. Interestingly, AGO4 localizes to Pol V loci in a manner that is dependent on Pol V transcription and siRNAs

(Wierzbicki et al. 2009). RNA-ChIP results further suggest that the AGO4 recruitment to chromatin may occur via siRNA-AGO4 base pairing to nascent Pol V transcripts.

Pol V regulates higher-order heterochromatin structure

Interphase heterochromatin is organized into higher-order structures called chromocenters, which consist of transcriptionally repressed rRNA genes, retrotransposons and various repetitive elements (Fransz et al. 2002; Soppe et al. 2002). Mutations in the CG maintenance methyltransferase, MET1, or an ATP dependent SWI/SNF chromatin remodeler, DDM1, result in disruption of normally condensed chromocenters, suggesting that DNA methylation is important for nucleating higher-order heterochromatin in *Arabidopsis* (Soppe et al. 2002). Mutations in a H3K9 methyltransferase, KRYPTONITE, do not affect chromocenter organization, suggesting that H3K9 methylation is downstream of chromocenter formation or that redundancy exists among other H3K9 methyltransferases (Jasencakova et al. 2003).

Surprisingly, mutations in *nrpd2* strongly impact chromocenter organization (Chapter 2) (Onodera et al. 2005). Analysis of chromocenters in *nrpd1* vs. *nrpe1* mutants has shown that Pol V is primarily responsible for nucleating chromocenters, whereas Pol IV is largely dispensable (Pontes et al., 2009). In fact, 24-nt siRNA production, assayed by using *dcl3* and *rdr2* mutants, is not required for heterochromatin formation (Pontes et al., 2009). Rather, Pol V association with DRD1 is critical because *drd1* mutants also impact chromocenter integrity, similar to *met1* and *ddm1* mutants (Pontes et al., 2009). The association of Pol V with the chromatin remodeler, DRD1, is reminiscent of a Pol V-dependent transcription pathway (Wierzbicki et al. 2008). However, this pathway

mediating chromocenter integrity is novel because DDM1, another chromatin remodeler, and the CG methyltransferase, MET1, are also involved with Pol V and DRD1 (Pontes et al., 2009). The mechanism by which these proteins nucleate heterochromatin into higher-order chromocenters is unclear, but intriguing models involve Pol V-dependent transcripts acting as scaffolds for the recruitment of other chromatin modifications necessary to maintain higher order structures (Pontes et al., 2009). This is consistent with dispersal of chromocenters after RNaseA treatment of interphase nuclei (Pontes et al., 2009).

Ancient origin of RNAi and Pol IV/V

Phylogenetic analysis from 25 genomes representing members of all five supergroups of eukaryotes revealed that the core RNAi machinery (Ago, Dicer and RdRP) is quite common throughout each of the eukaryotic lineages examined. In addition, chromatin modifications, such as histone methylation, are also found within many of these lineages, and organisms from each of these supergroups are susceptible to dsRNA-induced degradation of target mRNAs. Based on this evidence, RNAi is hypothesized to be an ancient process that likely functioned at the transcriptional and post-transcriptional level. Loss of RNAi has occurred independently in organisms that lack identifiable homologs to the core RNAi machinery, such as the budding yeast, *S. cerevisiae* (Cerutti and Casas-Mollano 2006).

Of the core RNAi components, the Ago family is the most conserved, and all organisms utilizing the RNAi process contain at least one Ago protein. The Dicer family is also highly conserved, with identifiable representatives in nearly all RNAi utilizing

organisms, in particular higher eukaryotes. RdRPs are the least conserved and appear to have been lost independently in several lineages, such as *Drosophila*, humans and some green algae (Cerutti and Casas-Mollano 2006). However, in these cases, mechanisms or triggers exist to maintain functional RNAi pathways.

Pol IV and Pol V are believed to have evolved in the last common ancestor of land plants and the Charales, a subgroup of complex algae often referred to as stoneworts (a type of freshwater plant) (Luo and Hall 2007). Although this dates to at least 500-700 million years ago, this is well after the evolution of the core silencing components, Dicer, Ago and RdRP, in the last common ancestor of all eukaryotes (Cerutti and Casas-Mollano 2006). Therefore, it is likely that silencing mechanisms were already well established in the plant kingdom before Pol IV and Pol V evolved. *NRPD1*, the gene encoding the largest subunit of Pol IV, is hypothesized to have evolved first by duplication of *NRPB1*, the largest subunit of Pol II. This is supported by the presence of *NRPD1*-like genes in all land plants and in at least some species of green algae (Luo and Hall 2007). *NRPD2* and *NRPE1* are hypothesized to have evolved later, because neither gene has been identified in green algae to date, although it is possible that previous studies have missed these homologs due to insufficient data mining or organism diversity coverage (Luo and Hall 2007). *NRPE1* is thought to have evolved by duplication of *NRPD1* sometime after land plants evolved but before the emergence of angiosperms (flowering plants). Interestingly, the relative substitution rates in the conserved domains of *NRPD1* and *NRPE1* show that these proteins are evolving at a significantly faster rate than the same regions of the largest subunits of Pol I, II and III (Luo and Hall 2007).

Notably, the evolution of sister silencing components Ago, Dicer and RdRP, is also occurring at a relatively fast rate (Cerutti and Casas-Mollano 2006).

Because only the largest subunit of Pol IV was identified in green algae, it is not known whether a fully functional Pol IV emerged in green algae. Given the similarity of NRPD2 to NRPB2, it is possible that Pol IV initially used the NRPB2 subunit of Pol II until a functional NRPD2 subunit evolved (Luo and Hall 2007).

An intriguing model is that Pol IV and Pol V evolved to replace the silencing functions carried out by Pol II. As discussed previously, mutations in Pol II subunits have been recovered that link Pol II function to silencing in fission yeast (Djupedal et al. 2005; Kato et al. 2005). Because these mutants are viable, they separate the housekeeping function of Pol II transcription with transcription devoted to triggering silencing.

ii.

DNA-DEPENDENT RNA POLYMERASES: SUBUNIT COMPOSITIONS AND ROLES OF THEIR SUBUNITS IN TRANSCRIPTION

Overview of prokaryotic, archaeal and eukaryotic RNA polymerases

DNA-dependent RNA polymerases are multi-subunit enzymes that serve as the catalytic engines to synthesize RNA from DNA. Whereas bacteria and archaea contain a single essential RNA polymerase, eukaryotes contain three functionally diversified and essential RNA polymerases, namely Pol I, II and III. Pol I transcribes the precursors of the ribosomal RNA (rRNA) genes (Grummt 2003). Pol II transcribes messenger RNAs (mRNAs) and non-coding RNAs, such as some small nuclear RNAs (snRNAs) and microRNAs (miRNAs) (Woychik and Hampsey 2002). Pol III transcribes 5S rRNA gene precursors, transfer RNAs (tRNAs) and some snRNAs such as U6 and 7SL RNAs (Schramm and Hernandez 2002).

E. coli RNA polymerase is composed of five core subunits- ß', ß, ω, and two copies of α (Figure 1)(Zhang et al. 1999). ß' and ß are the catalytic subunits that form the bulk of the enzyme surface. The α subunits homodimerize and are involved in the assembly of the catalytic subunits, along with the ω subunit (Minakhin et al. 2001). Whereas loss of function mutations in ß', ß and α are lethal, mutations in ω are viable but lead to slow growing cells and instability of the purified complex (Gentry et al. 1991; Mukherjee and Chatterji 1997).

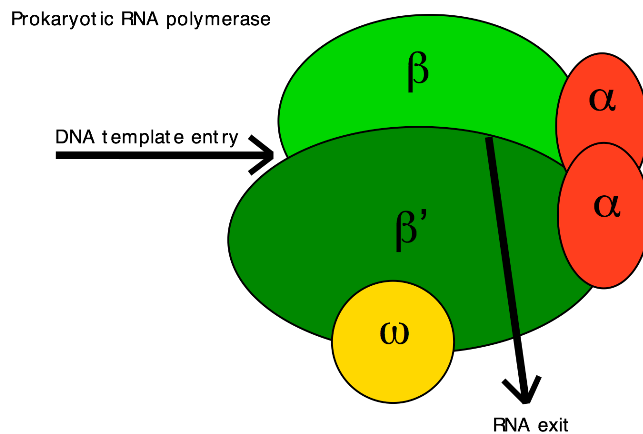


Figure 1. Subunit composition of a prokaryotic RNA polymerase. Colors distinguish subunits encoded by different genes. Note that both α subunits are encoded by the same gene. Arrows indicate the relative template entry and nascent RNA exit channels. The sizes and placement of the subunits are designed to reflect their approximate positions in the *T. aquaticus* crystal structure (Zhang et al. 1999).

The single archaeal RNA polymerase is more complex than *E. coli* RNA polymerase, with a core structure of eleven subunits, but is less complex than eukaryotic RNA polymerases. The two largest subunits, RpoA and RpoB, are equivalent to ß' and ß, and in some species are divided into two proteins, RpoA' and RpoA", and RpoB' and RpoB" (Werner 2007; Hirata et al. 2008b). (Rpo stands for "RNA polymerase".) RpoD and RpoL are homologous to the α subunit of *E. coli* and coordinate with RpoN and RpoP to mediate assembly of the two largest subunits. RpoK is most similar to the *E. coli* ω subunit, whereas RpoG, RpoH, RpoF and RpoE are subunits that evolved in archaea and have homologs in eukaryotes but not bacteria (Werner 2007). TFS/X is an RNA cleavage factor with homology to TFIIS in eukaryotes (Hausner et al. 2000; Lange

and Hausner 2004). Table 1 provides a subunit equivalency chart to illustrate the orthologous subunits among bacteria, archaea and eukaryotes.

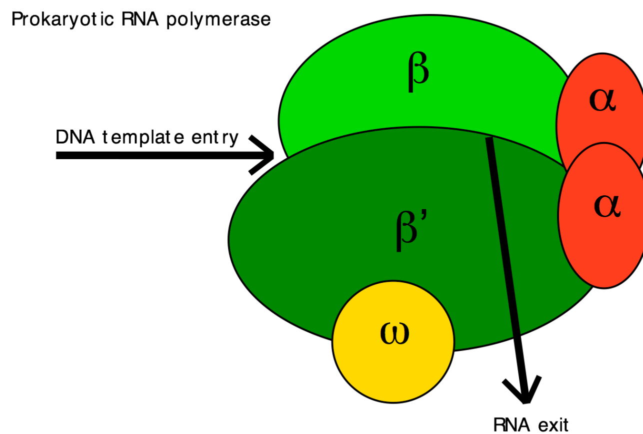
			yeast				*Arabidopsis thaliana*		
Bacteria	Archaea	Pol I	Pol II	Pol III	Pol I	Pol II	Pol III	Pol IV	Pol V
ß'	RpoA'	Rpa1	Rpb1	Rpc1	NRPA1	NRPB1	NRPC1	NRPD1	NRPE1
	RpoA"								
ß	RpoB'	Rpa2	Rpb2	Rpc2	NRPA2	NRPB2	NRPC2	NRPD2	NRPE2
	RpoB'								
α	RpoD	Rpac40	Rpb3	Rpac40	NRPA3	NRPB3	NRPC3	NRPD3	NRPE3a
									NRPE3b
α	RpoL	Rpac19	Rpb11	Rpac19	NRPA11	NRPB11	NRPA11	NRPD11	NRPE11
	RpoN	Rpb10	Rpb10	Rpb10	NRPA10	NRPB10	NRPC10	NRPD10	NRPE10
	RpoP	Rpb12	Rpb12	Rpb12	NRPA12?	NRPB12	NRPC12	NRPD12	NRPE12
ω	RpoK	Rpb6	Rpb6	Rpb6	NRPA6a	NRPB6a	NRPC6a	NRPD6a	NRPE6a
					NRPA6b	NRPB6b	NRPC6b	NRPD6b?	NRPE6b
	RpoG	Rpb8	Rpb8	Rpb8	NRPA8a	NRPB8a	NRPC8a	NRPD8a?	NRPE8a
					NRPA8b	NRPB8b	NRPC8b	NRPD8b	NRPE8b
	RpoH	Rpb5	Rpb5	Rpb5	NRPA5	NRPB5	NRPC5	NRPD5	NRPE5
	RpoF	Rpa14	Rpb4	Rpc17	?*	NRPB4	NRPC4	NRPD4	NRPE4
	RpoE	Rpa43	Rpb7	Rpc25	NRPA7?	NRPB7	NRPC7	NRPD7	NRPE7
	TFS/X	Rpa12	Rpb9	Rpc11	NRPA9	NRPB9a	NRPC9a?	NRPD9a	NRPE9a
						NRPB9b	NRPC9b?	NRPD9b	NRPE9b
		Rpa49			NRPA13				
		Rpa34			NRPA14				
				Rpc82			NRPC13		
				Rpc53			NRPC14a		
							NRPC14b		
				Rpc37			NRPC15		
				Rpc34			NRPC16		
				Rpc31			NRPC17		

Table 1. Relationships of DNA-dependent RNA polymerase subunits from bacteria, archaea, yeast and *Arabidopsis*. Same-shaded subunits are encoded by the same gene. Subunits with homologous functions are placed in the same row. "?" indicates that this subunit has not yet been shown to associate with the given RNA polymerase either by mass spectrometry or co-IP in *Arabidopsis*. "*" indicates that a homolog to this subunit has not been identified in *Arabidopsis* using BLASTp searches.

Eukaryotic RNA polymerase II is comprised of twelve core subunits, Pol I has fourteen core subunits and Pol III contains seventeen core subunits (Figures 2-3). Each polymerase contains subunits that are homologs of the five subunits of *E. coli* RNA polymerase (Figure 2). The largest and second-largest subunits of Pol I, II and III are encoded by unique genes and are the eukaryotic homologs of the ß' and ß prokaryotic subunits, respectively. The two prokaryotic α subunits share homology with two different proteins in all three eukaryotic RNA polymerases. In yeast Pol II, these subunits are Rpb3 and Rpb11. In yeast Pol I and Pol III, these subunits are Rpac40 and Rpac19. Note that these latter subunits are shared between Pol I and Pol III in yeast. Rpb6, a common subunit of Pol I, II and III, is homologous to ω (Minakhin et al. 2001). Four additional subunits are common to all three canonical polymerases in yeast, namely Rpb5, Rpb8, Rpb10 and Rpb12 (Woychik et al. 1990; Carles et al. 1991). Rpa12, Rpb9 and Rpc11 are functional equivalents specific to Pol I, II or III, respectively (Woychik et al. 1991; Nogi et al. 1993; Chedin et al. 1998). Rpb4 and Rpb7 complete the twelve subunit Pol II core in yeast (reviewed in (Choder 2004; Sampath and Sadhale 2005)); Rpc17-Rpc25 and Rpa14-Rpa43 form the Rpb4-Rpb7 subunit equivalents in yeast Pol I and III, respectively (Peyroche et al. 2002; Siaut et al. 2003). In addition to these twelve subunits, there are two Pol I-specific subunits (Rpa34 and Rpa49) and five Pol III-specific subunits (Rpc82, Rpc53, Rpc37, Rpc34 and Rpc31). Mutations in nearly all of these subunits result in lethal phenotypes. The notable exceptions are mutations in *rpb9, rpa12, rpb4, rpa49, rpa34* and *rpa14*, which are conditionally lethal at temperature extremes (Woychik and Young 1989; Woychik et al. 1991; Liljelund et al. 1992; Nogi et al. 1993; Smid et al. 1995; Gadal et al. 1997). Recent evidence from plants indicates that

mutations in the second-largest subunits *NRPA2*, *NRPB2* and *NRPC2* are lethal, as are mutations in the non-catalytic subunits *NRPB5*, *NRPB11* and *NRPB12* (Onodera et al. 2008)(Chapter 2), similar to reports from yeast.

The structures of Pol I, Pol II and Pol III have been determined in yeast (Cramer et al. 2001; Jasiak et al. 2006; Kuhn et al. 2007; Lorenzen et al. 2007). Overall, the three polymerases share conserved regions that participate in RNA catalysis and nucleotide triphosphate entry (Cramer et al. 2001; Fernandez-Tornero et al. 2007; Lorenzen et al. 2007). The major differences occur in areas near the periphery of the enzymes where the unique subunits of Pol I and Pol III bind (Figure 3).



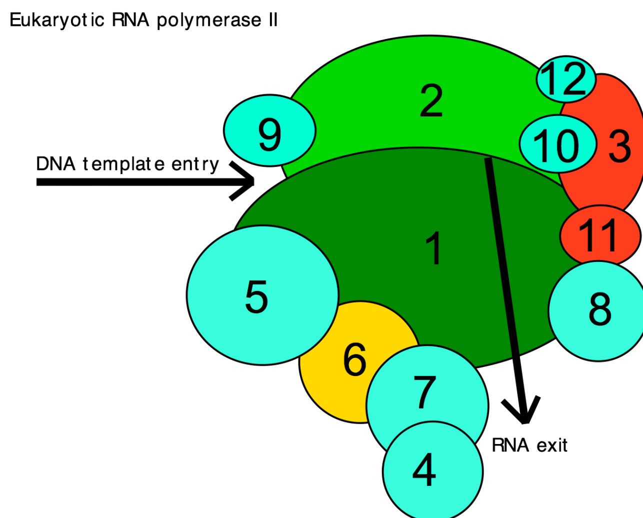


Figure 2. Subunit compositions of prokaryotic vs. eukaryotic RNA polymerases. Colors indicate homology between subunits from prokaryotes to eukaryotes. The cyan-colored subunits represent subunits that evolved in eukaryotes or archaeabacteria. Arrows indicate the relative template entry and nascent RNA exit channels. The sizes and placement of the eukaryotic subunits are designed to reflect their approximate positions in the yeast Pol II crystal structure (Armache et al. 2005).

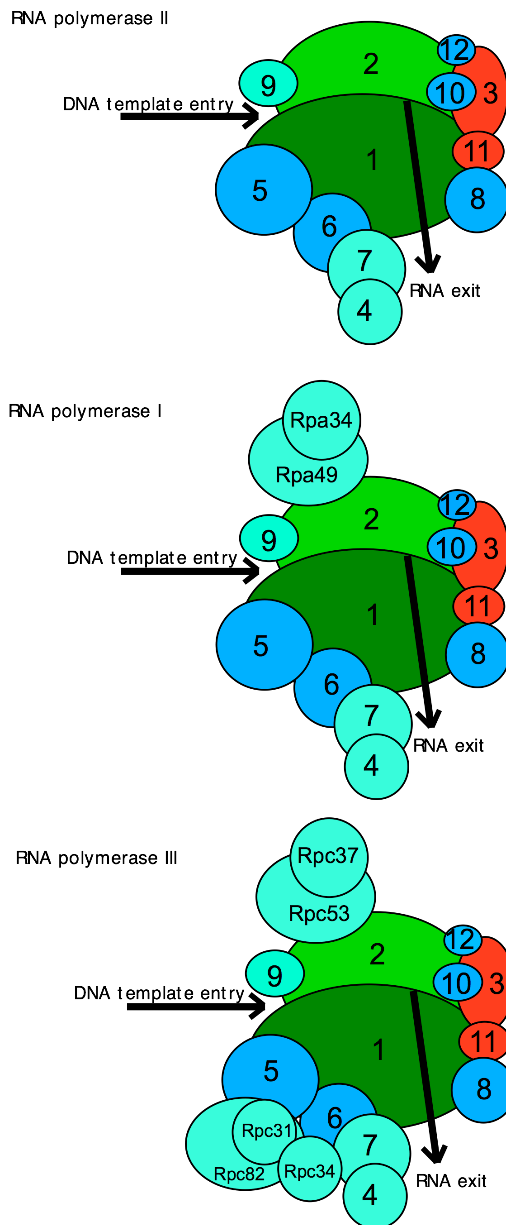


Figure 3. Subunit compositions of yeast RNA polymerases I, II and III. The green and red subunits reflect their homologies to prokaryotic subunits, as in Figures 1-2. Blue subunits are shared among Pol I, II and III. Cyan subunits represent either subunits specific to Pol I or III (denoted with "Rpa" or "Rpc" prefixes, respectively), or equivalent subunits among the three RNA polymerases that are encoded by different genes.

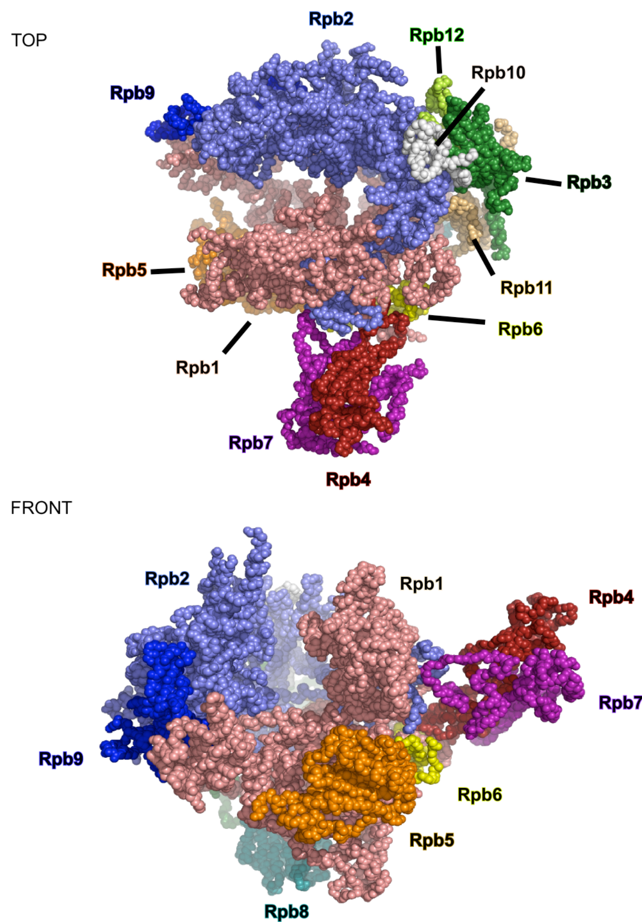


Figure 4. Subunit structure of yeast Pol II. Different colors highlight the twelve subunits in the crystal structure of Pol II. Top and front views are provided to illustrate all subunits. This figure was created in MacPymol and adapted from the yeast Pol II crystal structure (PDB code 1wcm or 1n9t) (Armache et al. 2005).

Nomenclature of yeast and plant DNA-dependent RNA polymerase subunits

Nomenclature of RNA polymerase subunits is confusing at best. Multiple systems exist in yeast, where RNA polymerases and the functions of their subunits have been studied most extensively. As a result, most of this introduction focuses on what is known in yeast. In yeast, three nomenclature systems exist (reviewed in (Archambault and Friesen 1993)). The first system uses the prefix Rpa, Rpb or Rpc, representing subunits that associate with Pol I, Pol II or Pol III, respectively (Young and Davis 1983). In system 1, Rpb is followed by a number that designates a subunit's rank according to its size in the polymerase, with Rpb1 representing the largest subunit of Pol II, Rpb2 the second-largest Pol II subunit and so on. The second system uses the Rpo1, Rpo2 and Rpo3 prefixes to indicate association with Pol I, II or III, respectively (Ingles et al. 1984). This system also uses a number that designates a subunit's rank according to its size in the polymerase. For example, Rpo21 represents the largest subunit of Pol II, Rpo22 the second-largest Pol II subunit and so on. The third system uses the prefixes Rpa, Rpb and Rpc, as in system one, but instead of a 1-12 numbering system, subunits are defined by their molecular weight in kilodaltons. For example, Rpb1=Rpb220 and Rpb6=Rpb23. To complicate the matter more, subunits that are shared by all three polymerases have in some cases been named for the polymerase in which they were first discovered. For example, Rpb6 is a shared subunit of Pol I, II and III in yeast and humans, yet its Pol I and Pol III roles are not reflected in its name. Rpb6/Rpo26/Rpb23 could easily be referred to as Rpa6/Rpo16/Rpa23 or Rpc6/Rpo36/Rpc23 in the contexts of Pol I and III, respectively, but such references are not consistent in the literature. In other cases, shared subunits have been referred to in the context of all three RNA polymerases using a

single name. For example, RPB6 has also been referred to as ABC23—ABC for Pol I, II and III and 23 standing for its size of 23 kDa. Unfortunately, nomenclature based on molecular weight is species-specific such that equivalent subunits in different species would have different names under this system.

The lack of a unified nomenclature system has muddled the literature and has made it very difficult to understand which polymerase subunit is being referred to in a manuscript. It is even more difficult to remember a given subunit's synonyms among papers that do not use the same nomenclature. Multiple nomenclature systems have complicated the designation of mutant alleles in yeast, where multiple different alleles corresponding to unique mutations have been given the same name. For example, *rpb1-5* and *rpo21-5* are referring to mutations of the same subunit, the largest subunit of Pol II, but these mutations are substitutions in *rpb1-5* (R335C) and insertions in *rpo21-5*, (Q124(ARAR)A) (Kolodziej and Young 1991; Archambault and Friesen 1993).

Each lab has adopted its favorite nomenclature system and this is sometimes clarified in the introduction of reports. However, the challenge still remains for becoming acquainted with the nomenclature, and for non-polymerase researchers it is not an easy task. In plants, far less is known about the subunit structures of the DNA-dependent RNA polymerases, and relatively few subunits have been given formal names in the literature (Ulmasov and Guilfoyle 1992; Larkin and Guilfoyle 1993; Ulmasov et al. 1995; Larkin and Guilfoyle 1996; Larkin and Guilfoyle 1997; Larkin and Guilfoyle 1998; Larkin et al. 1999). Therefore, this presents an opportunity to establish a single nomenclature system for plant DNA-dependent RNA polymerases that can be adopted by plant researchers studying multiple plant model systems. This system should be designed

such that equivalent subunits can easily be distinguished between plant species and between polymerases of a given plant species. Ideally, this system should be informative not just within plants but also translatable to yeast or mammalian nomenclature systems.

Establishing a unified nomenclature system for RNA polymerase subunits in plants

Several challenges exist to establishing a unified nomenclature system for DNA-dependent RNA polymerase subunits in plants. First, The *Arabidopsis* Information Resource, which oversees assignment of names and functions to genes, limits the character number to five letters in a name. Second, the use of RPA and RPD as prefixes to Pol I and Pol IV, respectively, is complicated because these names are already used to signify other genes. To circumvent this issue, we adopted a nomenclature system that includes the prefix "N" representing "Nuclear", such that NRPA stands for "Nuclear RNA Polymerase I", NRPB for "Nuclear RNA Polymerase II", NRPC for "Nuclear RNA Polymerase III", NRPD for "Nuclear RNA Polymerase IV" and NRPE for "Nuclear RNA Polymerase V" (Table 1).

Another challenge in establishing a plant nomenclature is determining the specific subunit designation numbers. In general, the sizes of subunits (in kDa) roughly correlate between yeast and plants, such that the largest subunits in yeast are the subunit equivalents of the largest subunits in plants. However, notable exceptions occur. One example is the RPB4 protein in plants, which is actually the eighth largest subunit rather than the fourth largest subunit, as in yeast Pol II. Therefore, we have proposed a nomenclature whereby plant RNA polymerase subunits are assigned a number according to their subunit equivalent in yeast Pol II, whose crystal structure is known and whose

subunits are familiar to RNA polymerase researchers (Armache et al. 2005). In the special cases of the two Pol I specific subunits and five Pol III specific subunits, these subunits are given the number designations 13-14 and 13-17 based on their homology to the corresponding yeast subunits, and are ranked in order of decreasing size according to their molecular weights in yeast. Again, naming Pol I or Pol III subunits strictly by their size in plants is not translatable to yeast because the Pol III-specific subunit Rpc37, which has an apparent molecular weight of 37 kDa and a predicted size of 32 kDa in yeast, is predicted to be 78 kDa in *Arabidopsis*.

There are several subunits that are present in multiple copies in the *Arabidopsis* genome. Therefore, to distinguish between these paralogs, we add a lowercase "a", "b", "c" and so on to differentiate these family members. In the case of the NRPB6 family, there are two members, referred to as NRPB6a and NRPB6b (Table 1).

Our nomenclature system developed for *Arabidopsis* should allow for naming subunits as their functions are determined in other plant species. Our system also allows for multiple synonyms of a subunit--which synonym is used largely depends on context. For example, if NRPB6a is being discussed as a subunit of Pol I, it is more appropriate to address this subunit as NRPA6a, even though it is the same protein as NRPB6a, NRPC6b, NRPD6b and NRPE6b (Table 1)(Chapters 3-4).

Identification of *Arabidopsis* RNA polymerase subunits

In *Arabidopsis thaliana*, database searches and cloning approaches have revealed homologs of nearly all Pol I, Pol II and Pol III subunits, with the exception of yeast Rpa14 (see below). Interestingly, many of the subunits are encoded by gene families, in

contrast to yeast and humans where single genes encode nearly all of the RNA polymerase subunits. A notable exception is the RPB11-like subunit in humans, which has diverged into at least three different isoforms (hRPB11a, hRPB11bα and hRPB11bβ), of which only one, hRPB11a, assembles into Pol II. However, hRPB11a does not complement a yeast *rpb11* mutant, whereas hRPB11bα does complement the yeast *rpb11* mutant, despite being excluded from human Pol II (Grandemange et al. 2001). One explanation for the subunit diversity in *Arabidopsis* stems from the evolution of Pol IV and Pol V, which utilize distinct largest and second-largest subunits in addition to unique NRPB4-like, NRPB7-like and NRPB5-like subunits (Chapter 3)(Ream et al. 2009). However, there has also been expansion of subunit gene families that are not associated with Pol IV or Pol V. Two genes related to the yeast Rpac40 subunit, namely NRPAC42 and NRPAC43, were shown in previous studies to be expressed in *Arabidopsis* suspension cell cultures; however, only NRPAC42 was shown to interact with Pol III, such that the function of NRPAC43 has been unknown (Ulmasov et al. 1995). Therefore, NRPAC42 and NRPAC43 may be functionally redundant or they could be variants of the yeast Rpac40 subunit that associate with distinct polymerases in *Arabidopsis*. In Trypanosomes, different Rpb5 and Rpb6 isoforms associate with Pol I and Pol II, which presumably reflects their functional diversification (Devaux et al. 2007). Likewise, the Pol II Rpb7 subunit has been shown to associate with Pol I in Trypanosomes, for which no Rpa43 homolog has been identified (Penate et al. 2009). (Rpa43 is the Rpb7 equivalent in yeast Pol I.)

In addition to multiple genes encoding each subunit shared by Pol I, II or III, there are multiple genes encoding homologs of several Pol III-specific subunits in *Arabidopsis*,

which correspond to the yeast Rpc11, Rpc17, Rpc31, and Rpc53 subunits (Table 1). Whether some or all of these subunits are incorporated into *Arabidopsis* Pol III is not known.

Bioinformatics searches have failed to identify the Rpa14 homolog of yeast Pol I in other eukaryotes. This subunit has been identified in fission yeast and budding yeast, but not in higher eukaryotes, making it unclear if this subunit has been conserved. One reason for the unsuccessful bioinformatics searches may be the rapid divergence of polymerase-specific subunits among species, as noted in previous studies (Imazawa et al. 2005; Proshkina et al. 2006). Therefore, purification of Pol I or performing a yeast-two-hybrid screen with the Rpa43 homolog might succeed in identifying this homolog in plants.

Conservation and evolution of RNA polymerase subunits

Studies of the single archaeal RNA polymerase have shown that its structure and subunit composition is very similar to that of eukaryotic RNA polymerases, with two exceptions (Langer et al. 1995; Hirata et al. 2008b; Kusser et al. 2008). Rpb9 is replaced with the TFS/X homolog, which has homology to the TFIIS subunit involved in polymerase pausing and RNA cleavage of stalled complexes. Rpb8 is only found in a subset of archaeabacteria, suggesting that these species might be the link in evolution to eukaryotic RNA polymerases (Langer et al. 1995; Koonin et al. 2007; Kwapisz et al. 2008).

Bioinformatics analyses have identified many of the twelve core subunits in Pol I, II and III throughout eukaryotes, suggesting that the core subunit structures are conserved

among eukaryotes (Proshkina et al. 2006). In support of this, experimental evidence has shown that the core subunit compositions of yeast and human RNA polymerases are nearly identical (Huang and Maraia 2001; Hu et al. 2002). Likewise, purification of RNA polymerases in other eukaryotes has also identified a majority of the common subunits (Das et al. 2006; Devaux et al. 2006; Nguyen et al. 2006; Martinez-Calvillo et al. 2007).

Hetero-complementation occurs with four out of the five subunits common to Pol I, II and III between yeast and humans, namely the functional equivalents of yeast Rpb6, Rpb8, Rpb10 and Rpb12 (Shpakovski et al. 1995). This is consistent with conservation of function of these core subunits during evolution. Yeast and human Rpb5, however, are not able to cross-complement, but fission yeast and budding yeast Rpb5s are able to cross-complement (Shpakovski et al. 1995). Cross-complementation of the Pol III subunits Rpc11, Rpc17, Rpc25, Rpac19 and Rpac40 has also been documented between budding and fission yeast (Shpakovski 1994; Chedin et al. 1998; Imai et al. 1999; Imazawa et al. 1999; Shpakovski and Shematorova 1999; Voutsina et al. 1999; Shpakovski et al. 2000; Proshkina et al. 2006). Recent work in plants has shown that the yeast Rpb4 homolog in *Arabidopsis*, NRPB4, is able to complement the yeast subunit *in vivo* (He et al. 2009). Interestingly, none of the five Pol III-specific subunits Rpc31, Rpc34, Rpc37, Rpc53 and Rpc82, nor the two catalytic subunits Rpc1 and Rpc2 from fission yeast are able to cross-complement mutations in the corresponding genes in budding yeast (Proshkina et al. 2006). This suggests that Pol I and Pol III-specific subunits may be evolving at a faster rate than shared subunits (i.e. Rpb5, Rpb6, Rpb8, Rpb10, Rpb12), or subunits with functional equivalents in other polymerases (i.e. Rpc11,

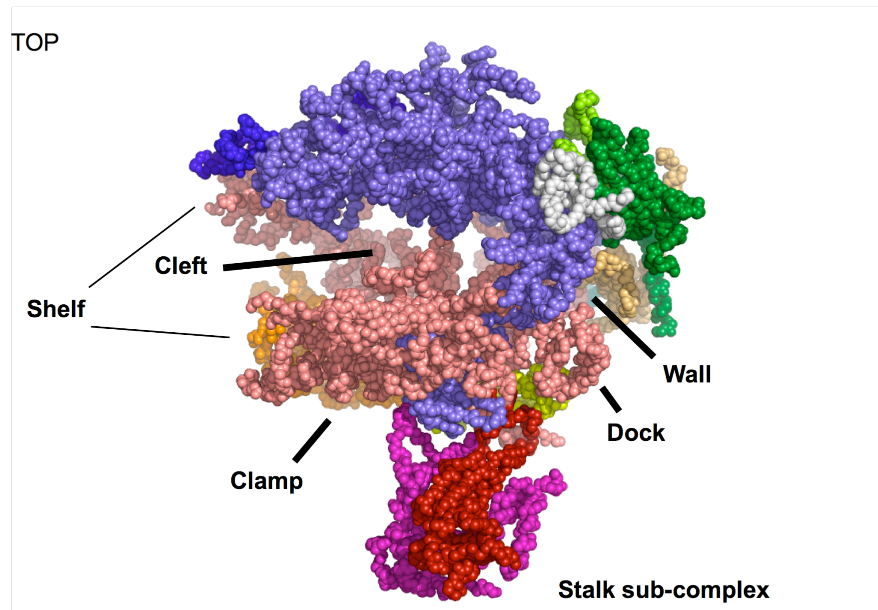
Rpac19, Rpac40, Rpb4, Rpc17, Rpc25). In contrast, subunits found to interact with all three polymerases tend to be more conserved across species compared to polymerase-specific genes.

Structural analysis of RNA polymerases

As stated earlier, subunit structures have been determined for bacterial, archaeal and eukaryotic RNA polymerases (Zhang et al. 1999; Cramer et al. 2001; Hirata et al. 2008b). Bacteria have the least complex RNA polymerase, composed of five subunits with an overall structure resembling a "crab claw" conformation. The two pincers represent the largest and second-largest subunits (β' and β), which interact to form unique structural features such as the active site, a positively charged cleft that accommodates the template DNA and a pore in the floor of the cleft that allows entry for incoming nucleotides (Zhang et al. 1999). The α dimer resides on the upstream side of the crab claw, at the base of the pincers, to stabilize the two largest subunits. The ω subunit resides on the β' side of the polymerase and does not interact with the α or β subunits. This overall structure has been retained in both archaea and eukaryotic polymerases, with the addition of new subunits that mediate a higher degree of regulatory complexity and functional diversification, especially in eukaryotes (Figure 4).

In Pol II, the "cleft", or space between the pincers, is formed on one side by the second-largest subunit, Rpb2, and by the "clamp" domain of Rpb1 on the other side (Figure 5). The active site is enclosed by several features: the opposing walls of the cleft mentioned above, a "bridge helix" that spans the cleft on the downstream side and a back "wall" on the upstream side of the polymerase formed by Rpb2. A stalk complex

comprised of an Rpb4-Rpb7 heterodimer is situated on the clamp side of the polymerase and protrudes upward and outward from the polymerase core (Figures 4 and 5). RNA leaving the active site emerges between the wall and the clamp and juts outward, in close proximity to the stalk complex. A "dock" domain spans a region in between the wall and the stalk complex and is proposed to bind initiation factors (Figure 5)(Cramer et al. 2000; Cramer et al. 2001; Gnatt et al. 2001).



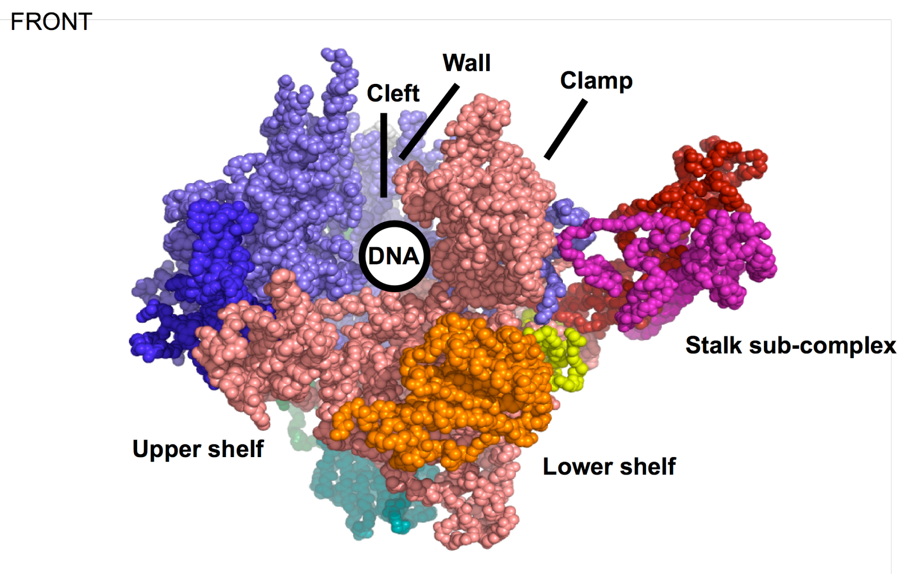


Figure 5. Prominent structural features of yeast RNA polymerase II. Colors represent the different subunits in top and front views, as in Figure 4. This figure was created in MacPymol and adapted from the yeast Pol II crystal structure (PDB code 1wcm or 1n9t) (Armache et al. 2005).

The non-catalytic subunits bind around the periphery of the catalytic subunits and participate in either stabilization of the polymerase or in initiation, elongation or termination in various contexts (Figures 4 and 5). The yeast Rpb3-Rpb11 or Rpac40-Rpac19 heterodimers promote the assembly and stabilization of the largest subunits on the upstream side of Rpb1 and Rpb2, at the base of the two crab claw pincers. Rpb10 and Rpb12 also participate in this process. Rpb5 and Rpb9 reside on the downstream face of the polymerase near the template DNA entry site and form the lower and upper "jaw" domains through their interactions with Rpb1 (in the case of Rpb5) or Rpb1 and Rpb2 (in the case of Rpb9). Rpb6 binds solely to Rpb1 on the lower clamp side of the polymerase and mediates interactions with the stalk complex (Cramer et al. 2000; Cramer et al. 2001; Gnatt et al. 2001).

As mentioned earlier, structural studies have revealed that yeast Pol I, II and III are similar in organization and have subunits that are shared or have functional equivalents (Figure 3). The Pol I-specific subunits Rpa34 and Rpa49 form a heterodimer that attaches on the Rpb2-side of the crab claw pincer. In Pol III, Rpc37 and Rpc53 heterodimerize and reside in a similar position. The remaining Pol III-specific subunits Rpc31, Rpc34 and Rpc82 interact to form a trimer that nests on the Rpb1 side of the

polymerase close to the stalk sub-complex comprised of Rpc17-Rpc25 (Cramer et al. 2001; Fernandez-Tornero et al. 2007; Kuhn et al. 2007).

Roles of the different RNA polymerase subunits in transcription

Non-specific (promoter-independent) transcription is carried out by RNA polymerases *in vitro* in the absence of initiation factors or activator proteins. This is in marked contrast to transcription *in vivo*, which is a highly regulated process requiring many subunit interactions. In the case of Pol II, the twelve-subunit core enzyme must be able to recognize specific genes in order to transcribe mRNA. Core Pol II is incapable of doing this efficiently and therefore requires a host of basal transcription factors to assist with promoter recognition, DNA melting and transcription initiation. In addition to basal transcription factors, this elaborate process is facilitated by a large protein complex, called “Mediator”, which as its name implies is able to translate information from activator and repressor proteins to the basal transcription machinery to influence transcription.

Studies on the function of each RNA polymerase subunit in relation to the whole enzyme have been most extensively studied in yeast, although progress has been achieved in other models such as flies, humans and mice. The following summary reviews the roles played by individual subunits or sub-complexes, focusing primarily on yeast.

The catalytic subunits

The largest and second-largest subunits of DNA-dependent RNA polymerases interact to form the active site that catalyzes nucleotide incorporation into the nascent mRNA. The active site consists of a pocket containing two magnesium ions bound at the so-called Metal A and Metal B sites. The magnesium ions stabilize the transition state of the reaction, which involves a nucleophilic attack by the 3' OH group of ribose in the RNA chain on the α-phosphorus atom of the incoming nucleotide triphosphate. These magnesium ions are also important for coordinating the active site in Rpb9/TFIIS-mediated RNA cleavage to overcome transcriptional pausing or arrest states (Sosunov et al. 2005). The Metal A site consists of a highly conserved motif in the largest subunit with three absolutely conserved aspartate residues. Metal B binding is coordinated by essential aspartate and glutamate residues in a highly conserved motif in the second-largest subunit (Cramer et al. 2001). Mutations within either of these motifs are dominant lethal and result in a catalytically inactive polymerase that is nonetheless capable of binding DNA (Dieci et al. 1995; Zaychikov et al. 1996; Werner and Weinzierl 2002).

A plethora of null mutations and conditional mutations have been identified within the largest and second-largest subunits in yeast and *E. coli* RNA polymerases. Whereas null mutations are lethal, conditional mutations take many different forms and have broad effects on the polymerase, including loss of interaction with smaller subunits, defects in assembly, failure to recruit transcription initiation factors, promiscuous start site selection, impaired nucleotide binding and defective promoter clearance, elongation or termination (reviewed in (Archambault and Friesen 1993)). These molecular

phenotypes indicate that the two largest subunits play many critical roles in the transcription cycle.

A unique feature of Pol II is the presence of a large C-terminal domain (CTD) within Rpb1. The CTD is thought to comprise a long tail separated from the rest of Rpb1 by a linker region. The CTD contains a series of short repeats that are dynamically and post-translationally modified. A minimum number of repeats are required for viability in yeast and other organisms (Nonet et al. 1987). The CTD is highly disordered in the crystal structure, but fully extended has the capacity to interact with almost any region of the core polymerase (Cramer 2004). The CTD mediates interaction with mRNA processing factors during transcription and also facilitates the transition state from initiation to elongation, processes that are tightly regulated by phosphorylation events (reviewed in (Meinhart et al. 2005)).

Interestingly, Pol V contains an elaborated CTD that is critical for its role in RNA-directed DNA methylation, siRNA production and silencing of endogenous genes (Haag et al., unpublished)(El-Shami et al. 2007). The CTD is a putative interaction target for AGO4 via an AGO-hook domain (a series of reiterated WG/GW motifs) that are universally important for Argonaute protein binding (see above)(Li et al. 2006; El-Shami et al. 2007). The Pol V CTD also contains at least three other domains in addition to the WG/GW motifs.

Subunits involved in assembly

The Rpb3-Rpb11 and Rpac40-Rpac19 heterodimers are the eukaryotic RNA polymerase homologs of the α homodimer in bacteria. In bacteria, the α subunits

dimerize before recruiting the second-largest subunit and finally the largest subunit (Yura and Ishihama 1979). Likewise, in eukaryotes, the Rpb3-Rpb11 dimer binds Rpb2 prior to assembling with Rpb1, and this Rpb3-Rpb11-Rpb2 trimer is quite stable in up to 6M urea (Kolodziej and Young 1991; Kimura et al. 1997). Structural and mutational studies have revealed that the Rpb3-Rpb11 dimer is critical for interaction with the Mediator complex and for activation-dependent transcription, in agreement with the roles of the α dimer in *E. coli* (Ebright and Busby 1995; Tan et al. 2000b; Davis et al. 2002). Consistent with a role in assembly and activation, Rpb3, Rpb11, Rpac40 and Rpac19 are essential for viability (Mann et al. 1987; Kolodziej and Young 1989; Dequard-Chablat et al. 1991).

Subunits shared by all three RNA polymerases

Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12 are shared subunits of Pol I, II and III and are essential for viability in yeast (Woychik et al. 1990; Woychik and Young 1990; Treich et al. 1991). A notable exception, however, is the protozoan *Trypanosoma brucei*, which uses distinct paralogs of Rpb5 and Rpb6 in Pol I vs. Pol II (Devaux et al. 2007). With the exception of Rpb6, the homolog to the bacteria ω subunit (Minakhin et al. 2001), these common subunits are found only in eukaryotes and archaeabacteria.

Both ω and Rpb6 are hypothesized to promote assembly or stability of the largest subunit, based on several lines of evidence. First, Rpb6 and ω physically interact with their corresponding largest subunits and some conditional mutations in the largest subunits are suppressed by over-expressing the Rpb6 or ω subunits (Cramer et al. 2001; Minakhin et al. 2001). Second, conditional *rpb6* mutant alleles destabilize the accumulation of Pol II (Nouraini et al. 1996). Polymerase fractions lacking Rpb6 lack

activity and activity can be restored upon the addition of recombinant Rpb6, suggesting that Rpb6 has a role in promoting transcription (Lanzendorfer et al. 1997).

As revealed by 2.8 Å and 3.1 Å crystal structures of the Pol II enzyme, yeast Rpb5 is located near the front of the enzyme in the lower jaw domain where it interacts with the largest subunit and with Rpb6 to form the "shelf" module (Figure 5)(Cramer et al. 2001; Gnatt et al. 2001). The shelf module is implicated in rotational orientation or stabilization of the template DNA as it enters the polymerase (Cramer et al. 2001). Rpb5 has also been implicated in structural stabilization of the bridge helix spanning the cleft and of the "switch 1 loop", which aids in clamp movement (Zaros et al. 2007). The potential role of Rpb5 as a target for transcriptional activation and inactivation is supported by several studies, although it is unclear whether this role is direct or indirect. TFIIB and TIP120, which facilitate recruitment of Pol II to the pre-initiation complex, have been shown to interact with Rpb5, as has Hepatitis B transcriptional activator protein X (HBx) (Cheong et al. 1995; Lin et al. 1997; Makino et al. 1999). Rpb5 is composed of two domains separated by a short linker, the N-terminal jaw domain and the C-terminal assembly domain. The assembly domain is involved primarily in interactions with the core polymerase, whereas the jaw domain is relatively free to interact with either DNA or regulatory proteins (Cramer et al. 2000; Gnatt et al. 2001). Deletion studies have demonstrated that regions of the Rpb5 jaw domain mediate the interactions between TFIIB and HBx (Lin et al. 1997). Rpb5 is also a link between Pol I, II and III transcription and chromatin remodeling because it interacts with the chromatin remodeling complex RSC in yeast (Soutourina et al. 2006).

Very little is known about the precise roles of Rpb10 and Rpb12. Based on their placement in the crystal structure of Pol II, these subunits reside near each other, the second-largest subunits, and the two α homologs, Rpb3 and Rpb11. Based on their locations, they are predicted to be involved in polymerase complex assembly or stability. Rpb10 contains an invariant HVDLIEK motif at its C-terminus motif that is essential for viability and for the stability of the largest subunit of Pol I within the core complex (Gadal et al. 1999). Rpb12 is an interesting subunit because in addition to being lethal, yeast with only one copy are haplo-insufficient and therefore grow more slowly than wild-type cells. Based on these data, it has been argued that this subunit is involved in a rate-limiting step in the assembly of polymerases (Rubbi et al. 1999). Rpb12 contains zinc-binding motifs, which are hypothesized to coordinate interaction between subunits and are involved in complex stability (Carles et al. 1991; Archambault and Friesen 1993; Rubbi et al. 1999). Zinc binding is a property of other subunits, too, including the catalytic subunits, and is utilized in a catalytic-independent manner to promote assembly, stability or interaction between subunits (Carles et al. 1991; Werner et al. 1992; Archambault and Friesen 1993; Rubbi et al. 1999). In addition, it is likely that zinc binding is critical for Pol IV or Pol V function, because a mutation in a zinc-binding motif of NRPD1 in maize abolishes paramutation (Erhard et al. 2009).

Unique subunits that have functional equivalents in the other polymerases

The RPB4-RPB7 stalk sub-complex of Pol II

Yeast Rpb7 is an essential polymerase subunit, whereas *rpb4* knockout mutants have been recovered and display conditional growth phenotypes, becoming lethal at very

high or low temperatures (Woychik and Young 1989; McKune et al. 1993). Likewise, in archaeabacteria, the Rpb7 equivalent, RpoE, is essential for growth but the Rpb4 equivalent, RpoF, is able to grow under normal conditions but is essential at extreme growth temperatures (Hirata et al. 2008a).

In purified yeast Pol II from *rpb4* knockout strains, both Rpb4 and Rpb7 are missing from the complex (Edwards et al. 1991; Sheffer et al. 1999). Over-expression of Rpb7 is able to partially suppress the effects of the *rpb4* mutation, (Sheffer et al. 1999). These data support the observations that Rpb4 and Rpb7 interact with each other (Armache et al. 2005). Point mutations in yeast RNA polymerase subunits that disrupt Rpb4-Rpb7 association have demonstrated that the Rpb4-Rpb7 complex anchors to Rpb1, Rpb2 and Rpb6 of the Pol II core, in close agreement with structural studies (Figures 4-5)(Tan et al. 2003; Armache et al. 2005; Sampath et al. 2008).

In *rpb4* mutants in yeast, promoter-dependent transcription is impaired (Edwards et al. 1991; Orlicky et al. 2001). In archaea, however, there are no apparent defects in transcription initiation, elongation or termination, suggesting that the eukaryotic Rpb4-Rpb7 sub-complex has evolved a more complex role in the function of Pol II (Hirata et al. 2008a). A role for Rpb4-Rpb7 in transcription initiation is apparent by structural studies that have shown that the sub-complex interacts with TFIIF during the assembly of the pre-initiation complex (Chung et al. 2003).

Several lines of evidence support the hypothesis that Rpb7 and Rpb4 associate with the polymerase throughout the transcription cycle. The nascent mRNA exiting the active site is able to interact with Rpb7 (Ujvari and Luse 2006; Chen et al. 2009). Rpb4 interacts with Fcp1, a phosphatase that acts on the CTD of Rpb1, and Rpb7 interacts with

a CTD-binding termination factor (Kimura et al. 2002; Mitsuzawa et al. 2003). Global mapping of Rpb4-Rpb7 binding sites relative to the core Pol II subunit Rpb3 has shown that the Rpb4-Rpb7 complex is present at nearly every step of transcription along the DNA and is bound at nearly all Pol II loci (Jasiak et al. 2008; Runner et al. 2008; Verma-Gaur et al. 2008). The presence of Rpb4 is required for the efficient recruitment of mRNA 3' end processing factors, which are decreased in *rpb4* mutants (Runner et al. 2008). This evidence challenges previous assumptions that the Rpb4-Rpb7 complex dissociates from Pol II under certain conditions, or once Pol II has been recruited to the initiation complex (Edwards et al. 1991; Jasiak et al. 2008).

As stated above, *rpb7* knockout mutants are lethal, which begs the question: how are *rpb4* mutants viable if Rpb7 dissociates from the polymerase in this background? One explanation is that Rpb7 plays a critical role in the cell outside the context of its role in the Pol II core enzyme. Mounting evidence demonstrates that the roles of Rpb4 and Rpb7 in transcription and transcription-coupled processes are complex (reviewed in (Choder 2004; Sampath and Sadhale 2005)). In addition, both subunits play roles outside the context of the polymerase core, including functions in the cytoplasm, suggesting that these subunits bridge the cross-talk between transcription and post-transcriptional processing.

A major role of Rpb4 in the context of Pol II may be to stabilize the association of Rpb7. However, evidence demonstrates that Rpb4 plays an important role that is independent of stabilizing Rpb7 outside the context of Pol II. Rpb4 is required for mediating selective mRNA export under stress conditions and shuttles from the nucleus to the cytoplasm to accomplish this task (Farago et al. 2003). This role is consistent with

a requirement for Rpb4 in Pol II transcription under stress conditions and with rescue of the *rpb4* mutant phenotype by overexpression of a nuclear pore complex protein (Rosenheck and Choder 1998; Tan et al. 2000a).

Rpb4 and Rpb7 are required for degradation of mRNAs in cytoplasmic P-bodies, sites where mRNA decay occurs (Sheth and Parker 2003; Lotan et al. 2005; Lotan et al. 2007). Rpb4 and Rpb7 localize to P-bodies and stimulate the de-adenylation of mRNAs as an initial step towards their destruction (Tharun and Parker 2001; Lotan et al. 2005; Lotan et al. 2007). Rpb4 and Rpb7 accomplish this by direct interaction with subunits of the mRNA decay complex, such as Pat1p, the first protein that binds the RNA targeted for decapping and degradation (Tharun and Parker 2001; Lotan et al. 2005; Lotan et al. 2007). Rpb7 appears to have a general role in mRNA decay, whereas Rpb4 is involved in the degradation of a subset of mRNAs representing protein biosynthesis genes (Lotan et al. 2005; Lotan et al. 2007), suggesting that Rpb4 and Rpb7 may have non-redundant roles in mRNA decay.

Interestingly, the roles of Rpb4 and Rpb7 in mRNA decay require their initial association with Pol II transcription, providing strong evidence for Rpb4-Rpb7-mediated cross-talk between transcription in the nucleus, mRNA export and mRNA decay kinetics in the cytoplasm (Goler-Baron et al. 2008). This may represent a form of “channeling”, where Rpb4 and Rpb7 never dissociate from the mRNA from the point of transcription initiation until the point of mRNA translation or degradation. However, the roles of Rpb4 and Rpb7 within P-bodies is probably performed outside the context of Pol II, since the Rpb3 or Rpb2 subunits do not interact with the Pat1p protein of the decapping complex and are not exported from the nucleus under stress conditions (Farago et al.

2003; Lotan et al. 2005). Nonetheless, Pol II transcription is required for the shuttling of Rpb4 and Rpb7 from the nucleus to the cytoplasm under normal growth conditions (Selitrennik et al. 2006). In summary, Rpb4 and Rpb7 are key proteins that play integral roles in transcription. They also interact with factors that mediate co-transcriptional and post-transcriptional processing of mRNAs.

The Rpc17-Rpc25 stalk sub-complex of Pol III

Rpc17 and Rpc25 are essential genes in yeast (Sadhale and Woychik 1994; Ferri et al. 2000). Similar to Rpb4-Rpb7, Rpc17 and Rpc25 subunits interact with each other in yeast and humans and form a protruding stalk module that interacts with the largest subunit and Rpb6 in core Pol III, in a manner analogous to Rpb4-Rpb7 in Pol II (Hu et al. 2002; Jasiak et al. 2006; Fernandez-Tornero et al. 2007; Lorenzen et al. 2007). Rpc25 is not essential in promoter-independent transcription assays, but is required for promoter-dependent initiation of transcription (Zaros and Thuriaux 2005). Elongation is not affected in *rpc25* conditional mutants (Zaros and Thuriaux 2005). Consistent with a role for the sub-complex in transcription initiation, Rpc17 interacts with a component of TFIIIB, a Pol III initiation factor that mediates recruitment of core Pol III to the promoter via TFIIIC, and Rpc31, a Pol III-specific core subunit involved in initiation (Ferri et al. 2000; Jasiak et al. 2006). Rpc17-Rpc25 may also play roles downstream of initiation because the sub-complex is able to bind tRNAs *in vitro*, but these roles are uncharacterized (Jasiak et al. 2006).

The Rpa14-Rpa43 stalk sub-complex of Pol I

Similar to Rpc25 and Rpb7, the Rpa43 counterpart in Pol I is essential for viability (Thuriaux et al. 1995). However, in contrast to Rpc17, but similar to Rpb4, Rpa14 is non-essential for growth at normal temperatures but is essential for viability at temperature extremes (Smid et al. 1995). The stalk module formed by the direct interaction of Rpa14-Rpa43 in Pol I is critical for interaction with the Pol I-specific initiation factor, Rrn3 (Peyroche et al. 2000; Peyroche et al. 2002; Armache et al. 2003; Meka et al. 2003). *Rrn3* is an essential gene necessary for Pol I promoter-dependent transcription *in vitro* and *in vivo* (Yamamoto et al. 1996; Milkereit and Tschochner 1998). To this end, *rpa43* point mutants abolish contact with Rrn3 and subsequently show no promoter-dependent Pol I activity *in vitro* (Peyroche et al. 2000). This result helps to explain why non-specific activity of Pol I *in vitro* can be obtained in the absence of Rpa43, but not *in vivo*, where Rrn3 is required (Hager et al. 1977). Interestingly, Pol I purified from *rpa14* mutants lacks Rpa43 and Rpb6 and is not active *in vitro*, suggesting that either Rpb6 or Rpa14 are critical for *in vitro* activity (Smid et al. 1995). Rrn3 association typically marks the active fraction of Pol I in the cell, which represents a relatively small portion of the total Pol I fraction (Milkereit and Tschochner 1998). The loss of association of Rpa14 and Rrn3 with *S. pombe* Pol I in post-log phase cells suggests a mechanism for regulating rRNA transcription in non-dividing cells, but the precise details of this mechanism are unclear (Imazawa et al. 2005).

Rpb9, Rpa12 and Rpc11

Rpa12, Rpb9, and Rpc11 represent equivalent subunits in yeast Pol I, II and III based on sequence homology, function and location within the polymerase structure. Structural studies place these subunits on the Rpb2 side of the polymerase near the upper front jaw of the enzyme (Cramer et al. 2000; Fernandez-Tornero et al. 2007; Kuhn et al. 2007). Only Rpc11 is absolutely essential; knockouts of yeast Rpa12 or Rpb9 lead to viable cells that exhibit slow growth at temperature extremes (Woychik et al. 1991; Nogi et al. 1993; Chedin et al. 1998). All three proteins have homology to TFIIS, a cleavage factor involved in overcoming transcriptional arrest at pause sites during elongation (reviewed in (Wind and Reines 2000)). Rpa12 and Rpc11 play important roles in termination of transcription, either through their intrinsic abilities to cleave the nascent RNA (Rpc11), or by recruiting factors involved in termination or RNA processing (Rpa12) (Awrey et al. 1997; Chedin et al. 1998; Prescott et al. 2004). Rpb9 also plays a role in TFIIS-mediated read-through of pause sites during elongation and is critical for accurate start site selection during initiation (Hull et al. 1995; Awrey et al. 1997; Hemming and Edwards 2000; Hemming et al. 2000).

Pol I-specific subunits

Rpa34 and Rpa49 are two subunits that interact with each other on the periphery of the Pol I complex, located on the Rpa2 side of the polymerase approximately opposite of the Rpa14-Rpa43 stalk (Kuhn et al. 2007; Beckouet et al. 2008). These subunits are not essential for viability and do not effect promoter-independent *in vitro* transcription of Pol I (Liljelund et al. 1992; Gadal et al. 1997). In support of their direct interaction, Pol I

purified from *rpa34* mutants in yeast lacks Rpa49, and Rpa34 is lost from Pol I in *rpa49* mutants (Gadal et al. 1997; Beckouet et al. 2008). Loss of Rpa34 does not affect cell growth, but *rpa49* mutations lead to slow growing cells (Liljelund et al. 1992; Gadal et al. 1997). Rpa34 may play an important role in recruiting DNA topoisomerase to relieve barriers to Pol I transcription. This conclusion is based on two lines of evidence. First, double *rpa34 top1* yeast mutants are synthetically lethal, unlike the nearly wild-type phenotypes of either single mutant, and second, these two proteins directly interact (Gadal et al. 1997; Beckouet et al. 2008). Rpa49 is involved in Rrn3 binding by an unknown mechanism that is dependent on Rpa43, suggesting that Rpa49 is important for Pol I transcription initiation (Beckouet et al. 2008).

Pol III-specific subunits

There are five Pol III-specific subunits in yeast, namely Rpc31, Rpc34, Rpc37, Rpc53 and Rpc82, each of which have homologs in other eukaryotes. All five of these subunits are essential for viability in yeast (Mosrin et al. 1990; Chiannilkulchai et al. 1992; Mann et al. 1992; Stettler et al. 1992; Landrieux et al. 2006). As observed for the Pol I-specific subunits, the Pol III-specific subunits associate near the periphery of the core enzyme (Fernandez-Tornero et al. 2007; Lorenzen et al. 2007). These subunits partition into two distinct complexes with non-overlapping functions.

Rpc31-Rpc34-Rpc82 interact to form a trimer that is important for transcriptional initiation and binding to the TFIIIB complex that is involved in promoter-specific recognition, such that mutations in these subunits inhibit promoter-dependent initiation of transcription (Werner et al. 1992; Werner et al. 1993; Thuillier et al. 1995; Brun et al.

1997; Wang and Roeder 1997). Rpc34 mediates the interaction of the trimer with TFIIIB, similar to Rpc17 in the Rpc17-Rpc25 dimer (Werner et al. 1993; Brun et al. 1997). The trimer is situated on the Rpc1 side of the cleft near the stalk sub-complex, where it would be in a position to interact with initiation factors. Rpc31 is hypothesized to act as a bridge protein between Rpc34-Rpc82, Rpc17 and Rpc1 (Flores et al. 1999; Ferri et al. 2000; Geiduschek and Kassavetis 2001; Fernandez-Tornero et al. 2007; Lorenzen et al. 2007).

Rpc37 and Rpc53 interact to form a complex implicated in termination and facilitated re-initiation of transcription, in combination with Rpc11 in the latter case (Flores et al. 1999; Hu et al. 2002; Landrieux et al. 2006; Fernandez-Tornero et al. 2007). The Rpc37-Rpc53 dimer is positioned opposite to the Rpc17-Rpc25 stalk on the Rpc2 side of the cleft, similar to, but not precisely the same as, the position of Rpa34-Rpa49 in Pol I (Fernandez-Tornero et al. 2007).

In conclusion, it is apparent from a plethora of genetic and biochemical studies that each of the RNA polymerase subunits play a vital role in transcription, either by recruiting initiation, elongation or termination factors, or by being integral to those processes themselves. Many of the subunits also promote the stability of the complex. In addition, some subunits may function outside the context of the polymerase. Future studies will no doubt uncover new functions for all of these subunits.

Early studies of *Arabidopsis* DNA-dependent RNA polymerase subunits

In contrast to yeast, little is known about the roles of each RNA polymerase subunit in plant RNA polymerases. Like all eukaryotes, plants also contain three

essential, functionally diversified DNA-dependent RNA polymerases (Appendix C) (Guilfoyle et al. 1984; Onodera et al. 2008). The three DNA-dependent RNA polymerases were initially discriminated by their different sensitivities to α-amanitin, a fungal toxin that is a potent inhibitor of Pol II transcription and a mild inhibitor of Pol III transcription (Lindell et al. 1970). In *Arabidopsis*, Pol I, II and III have been purified according to their binding affinities on a variety of chromatography columns (Ulmasov et al. 1995; Ulmasov et al. 1996; Saez-Vasquez and Pikaard 1997). In these cases, polyacrylamide gel electrophoresis determined that these enzymes consisted of multiple subunits. However, information about subunit composition was limited to whether polypeptides in a preparation cross-reacted with a limited number of available antibodies to subunits from different plant species. These results left open the questions of which subunits comprise each of these enzymes and whether multiple subunits from a given family are utilized by only one or multiple polymerases.

Rpb3-related subunits

Despite these limitations, several RNA polymerase subunit families have been characterized. Ulmasov et al. determined that two Rpb3-like subunits are expressed in *Arabidopsis*, named AtRPB36a (now NRPB3a) and AtRPB36b (now NRPE3b), but only one NRPB11 subunit was identified (Ulmasov et al. 1996). These three subunits are homologous to the *E. coli* α-subunits involved in assembly of the core enzyme. Interestingly, both NRPB3 subunits are able to interact with NRPB11 based on yeast-two-hybrid and *in vitro* pull-down assays, similar to the Rpb3-Rpb11 heterodimer interaction in yeast Pol II. However, in highly purified fractions of Pol II, only NRPB3a

was detected, leaving unanswered the function of AtRPB36b (now NRPE3b) until this thesis (Chapter 3)(Ulmasov et al. 1996). To complement the NRPB3-NRPB11 studies, Larkin et al. purified recombinant Rpac40-like subunits, AtRPAC42 (now NRPA3) and AtNRPAC43 (now NRPC3), and demonstrated that both interact *in vitro* with AtRPAC19 (now NRPA11/NRPC11), the Pol I and Pol III functional equivalent of the *E. coli* α-dimer (Larkin and Guilfoyle 1997). These results are consistent with the interaction between Rpac40 and Rpac19 in yeast Pol I and III. NRPAC19 (now NRPA11/NRPC11) co-purifies with Pol I and III from a variety of plant species (Larkin and Guilfoyle 1996). However, only AtRPAC42 (now NRPA3) was found to associate with purified Pol III, making the role of AtRPAC43 (now NRPC3) unclear (Ulmasov et al. 1995). Whereas the *Arabidopsis* NRPB3 and NRPB11 subunits can form heterodimers with their yeast counterparts, no interaction was observed between *Arabidopsis* NRPB3 and NRPA11/NRPC11 subunits nor NRPA3 or NRPC3 subunits with NRPB11 subunits. These results are consistent with studies from yeast that suggested Pol I and III share a conserved α-related dimer that is different from the α-related dimer found in Pol II (Larkin and Guilfoyle 1997).

Rpb4- and Rpb7-related subunits

Other insights into *Arabidopsis* RNA polymerase subunits came from studies on the NRPB4-NRPB7 sub-complex. Similar to fission yeast and humans, the *Arabidopsis* NRPB4-NRPB7 subunits stay associated with the polymerase throughout multiple chromatographic steps, suggesting that NRPB4-NRPB7 is stably bound to the core enzyme, in contrast to budding yeast where these subunits dissociate from Pol II

(Ulmasov and Guilfoyle 1992; Khazak et al. 1998; Larkin and Guilfoyle 1998; Sakurai et al. 1999). *Arabidopsis* NRPB4 and NRPB7 interact *in vitro*, in agreement with studies of yeast Pol II (Larkin and Guilfoyle 1998). However, little is known about the function of the NRPB4-NRPB7 sub-complex in plant Pol II transcription, or if there is a role of this complex or the individual subunits outside the context of Pol II transcription.

Rpb5-related subunits

Arabidopsis has five genes encoding full-length NRPB5-like subunits, whereas yeast and humans have only one gene (Chapter 2) (Larkin et al. 1999; Ream et al. 2009). Initial cloning studies revealed that two of these *Arabidopsis* NRPB5-like family subunits, NRPB5/NRPD5 and NRPE5 are expressed in suspension cell cultures, whereas the expression and function of the other three NRPB5-like family members was not tested (Larkin et al. 1999). NRPB5, but not NRPE5, was detected in highly purified fractions of Pol II and Pol III in *Arabidopsis*, and later in Pol I (Saez-Vasquez and Pikaard 1997; Larkin et al. 1999). Virtually nothing is known about the functional role(s) of the NRPB5 subunit in plants, except what can be inferred from studies in other organisms, especially yeast (reviewed above). I have shown that NRPE5 is required for RNA-directed DNA methylation, siRNA production and silencing of endogenous transposable elements and repeats and is a subunit of Pol V, as reported in Chapter 2 (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009).

Scope of this thesis

The establishment of transcriptional silencing is triggered by noncoding RNAs, resulting in dicing and targeting of repressive chromatin modifications to the target locus. Therefore, in order to maintain silencing, it is increasingly clear that some level of transcription is required for providing the silencing triggers (Grewal and Elgin 2007). Insight into this apparent paradox was resolved by the isolation of viable mutations in yeast Pol II subunits that were defective in silencing, providing evidence that low-level transcription by Pol II is necessary for maintaining silencing (Djupedal et al. 2005; Kato et al. 2005). However, transcription of silent heterochromatin presents another anomaly in itself, because heterochromatin has long been considered inaccessible to Pol II transcription (reviewed in (Studitsky et al. 2004)). A solution to this paradox is that Pol II preferentially transcribes heterochromatin during the S phase of the cell cycle, when chromatin is less condensed, therefore accommodating both models (Chen et al. 2008; Kloc and Martienssen 2008; Kloc et al. 2008).

The discovery of two plant-specific RNA polymerases, Pol IV and Pol V, has provided insight into the transcription of plant heterochromatin, the subsequent generation of silencing triggers and the maintenance of silencing. Prior to work undertaken in this thesis, the functions of Pol IV and Pol V were unknown. In addition, it was unclear whether Pol IV and Pol V function as simple, prokaryotic-like RNA polymerases or if they utilize a larger complement of subunits to direct silencing by transcription. Given the homology of their catalytic subunits to those of Pol II, it seemed likely that they would be multi-subunit enzymes, but the degree to which they shared any subunits was undefined. My thesis work shows that Pol IV and Pol V are multi-subunit

RNA polymerases that are structurally related to Pol II, and that non-catalytic subunits of Pol V are required for endogenous gene silencing.

The purification of all five *Arabidopsis* DNA-dependent RNA polymerases, revealed in this thesis, is unparalleled in any given organism. Before my thesis, little information was known about the precise subunit compositions of plant RNA polymerases. My thesis work has systematically determined the polymerase association of nearly every predicted RNA polymerase subunit in *Arabidopsis*, of which there are over fifty. My thesis work has also determined the functional role of some of these subunits and has created tools to address the functions of other subunits in the future. My work has revealed that many subunits are shared among some or all five of the RNA polymerases, whereas some subunits are highly specific and presumably mediate polymerase-specific functions. This information is essential for future studies that will examine the in-depth roles of each of these subunits.

Chapter 2 begins with the first published genetic and biochemical analyses of the catalytic subunits of RNA polymerase IV from *Arabidopsis thaliana*. The findings in Chapter 2 demonstrate that Pol IV subunits are required for RNA-directed DNA methylation, siRNA production and heterochromatin formation at endogenous loci. Chapter 3 reveals the multi-subunit structures of Pol IV and Pol V. Based on these results, Pol IV and Pol V likely evolved from Pol II because they share many subunits, some of which are excluded from Pol I and Pol III. I show that Pol IV and Pol V have evolved subunits that are specific to each polymerase, which likely mediate their specific functions *in vivo*. Moreover, a non-catalytic subunit, NRPE5, is essential for gene silencing. Chapter 4 reports the first comprehensive, parallel analysis of the subunit

compositions of all five DNA-directed RNA polymerases from an organism. These results provide a valuable framework for studying RNA polymerase subunit families in *Arabidopsis* and other model plants in the future. I complete my thesis with an outlook on the next ten years of RNA polymerase research in *Arabidopsis* and other plants, with a focus on Pol IV and Pol V and the roles of their non-catalytic subunits in transcription and gene silencing. Additional insight is provided towards the functions of RNA polymerase I, II, and III subunits.

Lastly, I include appendices that showcase the depth and breadth of my contributions to RNAi and RNA polymerase research, both intellectually and physically, over the past six years.

REFERENCES CITED

2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**(6814): 796-815.

Adenot, X., Elmayan, T., Lauressergues, D., Boutet, S., Bouche, N., Gasciolli, V., and Vaucheret, H. 2006. DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. *Curr Biol* **16**(9): 927-932.

Agger, K., Christensen, J., Cloos, P.A., and Helin, K. 2008. The emerging functions of histone demethylases. *Curr Opin Genet Dev* **18**(2): 159-168.

Agius, F., Kapoor, A., and Zhu, J.K. 2006. Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc Natl Acad Sci U S A* **103**(31): 11796-11801.

Agrawal, N., Dasaradhi, P.V., Mohmmed, A., Malhotra, P., Bhatnagar, R.K., and Mukherjee, S.K. 2003. RNA interference: biology, mechanism, and applications. *Microbiol Mol Biol Rev* **67**(4): 657-685.

Alleman, M., Sidorenko, L., McGinnis, K., Seshadri, V., Dorweiler, J.E., White, J., Sikkink, K., and Chandler, V.L. 2006. An RNA-dependent RNA polymerase is required for paramutation in maize. *Nature* **442**(7100): 295-298.

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. 2005. microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**(2): 207-221.

Allshire, R.C., Nimmo, E.R., Ekwall, K., Javerzat, J.P., and Cranston, G. 1995. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev* **9**(2): 218-233.

Almeida, R. and Allshire, R.C. 2005. RNA silencing and genome regulation. *Trends Cell Biol* **15**(5): 251-258.

Aravin, A.A., Hannon, G.J., and Brennecke, J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**(5851): 761-764.

Archambault, J. and Friesen, J.D. 1993. Genetics of eukaryotic RNA polymerases I, II, and III. *Microbiol Rev* **57**(3): 703-724.

Armache, K.J., Kettenberger, H., and Cramer, P. 2003. Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**(12): 6964-6968.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**(8): 7131-7134.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. 2002. HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. *EMBO J* **21**(24): 6832-6841.

Aukerman, M.J. and Sakai, H. 2003. Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. *Plant Cell* **15**(11): 2730-2741.

Awrey, D.E., Weilbaecher, R.G., Hemming, S.A., Orlicky, S.M., Kane, C.M., and Edwards, A.M. 1997. Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem* **272**(23): 14747-14754.

Axtell, M.J., Jan, C., Rajagopalan, R., and Bartel, D.P. 2006. A two-hit trigger for siRNA biogenesis in plants. *Cell* **127**(3): 565-577.

Banerjee, D. and Slack, F. 2002. Control of developmental timing by small temporal RNAs: a paradigm for RNA-mediated regulation of gene expression. *Bioessays* **24**(2): 119-129.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**(6824): 120-124.

Baumberger, N. and Baulcombe, D.C. 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**(33): 11928-11933.

Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadal, O., Nogi, Y., and Thuriaux, P. 2008. Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol* **28**(5): 1596-1605.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. 2006. mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* **20**(14): 1885-1898.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818): 363-366.

Bertrand, C., Benhamed, M., Li, Y.F., Ayadi, M., Lemonnier, G., Renou, J.P., Delarue, M., and Zhou, D.X. 2005. Arabidopsis HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth. *J Biol Chem* **280**(2): 1465-1473.

Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. 2009. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*.

Birchler, J.A., Bhadra, M.P., and Bhadra, U. 2000. Making noise about silence: repression of repeated genes in animals. *Curr Opin Genet Dev* **10**(2): 211-216.

Blevins, T., Rajeswaran, R., Shivaprasad, P.V., Beknazariants, D., Si-Ammour, A., Park, H.S., Vazquez, F., Robertson, D., Meins, F., Jr., Hohn, T., and Pooggin, M.M. 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Res* **34**(21): 6233-6246.

Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. 1998. AGO1 defines a novel locus of Arabidopsis controlling leaf development. *EMBO J* **17**(1): 170-180.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. 2005. Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. *Cell* **123**(7): 1279-1291.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**(6): 1089-1103.

Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., and Hannon, G.J. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**(5906): 1387-1392.

Brink, R.A. 1956. A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. *Genetics* **41**(6): 872-889.
Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M., and Carroll, B.J. 2007. Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. *Proc Natl Acad Sci U S A* **104**(37): 14741-14746.
Brun, I., Sentenac, A., and Werner, M. 1997. Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* **16**(18): 5730-5741.
Buhler, M., Verdel, A., and Moazed, D. 2006. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**(5): 873-886.
Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. 2003. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr Biol* **13**(24): 2212-2217.
Cao, X. and Jacobsen, S.E. 2002. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol* **12**(13): 1138-1144.
Carles, C., Treich, I., Bouet, F., Riva, M., and Sentenac, A. 1991. Two additional common subunits, ABC10 alpha and ABC10 beta, are shared by yeast RNA polymerases. *J Biol Chem* **266**(35): 24092-24096.
Castanotto, D., Tommasi, S., Li, M., Li, H., Yanow, S., Pfeifer, G.P., and Rossi, J.J. 2005. Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. *Mol Ther* **12**(1): 179-183.
Cerutti, H. and Casas-Mollano, J.A. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet* **50**(2): 81-99.
Chan, S.W., Zhang, X., Bernatavichute, Y.V., and Jacobsen, S.E. 2006. Two-step recruitment of RNA-directed DNA methylation to tandem repeats. *PLoS Biol* **4**(11): e363.
Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. 2004. RNA silencing genes control de novo DNA methylation. *Science* **303**(5662): 1336.
Chandler, V.L. 2007. Paramutation: from maize to mice. *Cell* **128**(4): 641-645.
Chandler, V.L., Eggleston, W.B., and Dorweiler, J.E. 2000. Paramutation in maize. *Plant Mol Biol* **43**(2-3): 121-145.
Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev* **12**(24): 3857-3871.
Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T., and Chang, W.H. 2009. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci U S A* **106**(1): 127-132.
Chen, E.S., Zhang, K., Nicolas, E., Cam, H.P., Zofall, M., and Grewal, S.I. 2008. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* **451**(7179): 734-737.
Chen, X. 2008. MicroRNA metabolism in plants. *Curr Top Microbiol Immunol* **320**: 117-136.
Chen, Z.J. and Tian, L. 2007. Roles of dynamic and reversible histone acetylation in plant development and polyploidy. *Biochim Biophys Acta* **1769**(5-6): 295-307.

Cheong, J.H., Yi, M., Lin, Y., and Murakami, S. 1995. Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human hepatitis B virus X protein and may play a role in X transactivation. *Embo J* **14**(1): 143-150.

Chiannilkulchai, N., Stalder, R., Riva, M., Carles, C., Werner, M., and Sentenac, A. 1992. RPC82 encodes the highly conserved, third-largest subunit of RNA polymerase C (III) from Saccharomyces cerevisiae. *Mol Cell Biol* **12**(10): 4433-4440.

Choder, M. 2004. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem Sci* **29**(12): 674-681.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. 2002. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell* **110**(1): 33-42.

Chung, W.H., Craighead, J.L., Chang, W.H., Ezeokonkwo, C., Bareket-Samish, A., Kornberg, R.D., and Asturias, F.J. 2003. RNA polymerase II/TFIIF structure and conserved organization of the initiation complex. *Mol Cell* **12**(4): 1003-1013.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**(7184): 215-219.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. 2007. Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell* **27**(3): 449-461.

Cook, H.A., Koppetsch, B.S., Wu, J., and Theurkauf, W.E. 2004. The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell* **116**(6): 817-829.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12**(23): 3715-3727.

Cramer, P. 2004. RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev* **14**(2): 218-226.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**(5466): 640-649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**(5523): 1863-1876.

Csink, A.K., Linsk, R., and Birchler, J.A. 1994. The Lighten up (Lip) gene of Drosophila melanogaster, a modifier of retroelement expression, position effect variegation and white locus insertion alleles. *Genetics* **138**(1): 153-163.

Curaba, J. and Chen, X. 2007. Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., Hannon, G.J., and Brennecke, J. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**(7196): 798-802.

Dalmay, T., Hamilton, A., Rudd, S., Angell, S., and Baulcombe, D.C. 2000. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell* **101**(5): 543-553.

Das, A., Li, H., Liu, T., and Bellofatto, V. 2006. Biochemical characterization of Trypanosoma brucei RNA polymerase II. *Mol Biochem Parasitol* **150**(2): 201-210.

Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. 2002. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell* **10**(2): 409-415.

Daxinger, L., Kanno, T., Bucher, E., van der Winden, J., Naumann, U., Matzke, A.J., and Matzke, M. 2009. A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation. *EMBO J* **28**(1): 48-57.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**(7014): 231-235.

Dequard-Chablat, M., Riva, M., Carles, C., and Sentenac, A. 1991. RPC19, the gene for a subunit common to yeast RNA polymerases A (I) and C (III). *J Biol Chem* **266**(23): 15300-15307.

Deshpande, G., Calhoun, G., and Schedl, P. 2005. Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* **19**(14): 1680-1685.

Devaux, S., Kelly, S., Lecordier, L., Wickstead, B., Perez-Morga, D., Pays, E., Vanhamme, L., and Gull, K. 2007. Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. *Mol Biol Cell* **18**(4): 1293-1301.

Devaux, S., Lecordier, L., Uzureau, P., Walgraffe, D., Dierick, J.F., Poelvoorde, P., Pays, E., and Vanhamme, L. 2006. Characterization of RNA polymerase II subunits of Trypanosoma brucei. *Mol Biochem Parasitol* **148**(1): 60-68.

Diaz-Pendon, J.A., Li, F., Li, W.X., and Ding, S.W. 2007. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**(6): 2053-2063.

Dieci, G., Hermann-Le Denmat, S., Lukhtanov, E., Thuriaux, P., Werner, M., and Sentenac, A. 1995. A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J* **14**(15): 3766-3776.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**(19): 2301-2306.

Donaire, L., Barajas, D., Martinez-Garcia, B., Martinez-Priego, L., Pagan, I., and Llave, C. 2008. Structural and genetic requirements for the biogenesis of tobacco rattle virus-derived small interfering RNAs. *J Virol* **82**(11): 5167-5177.

Dugas, D.V. and Bartel, B. 2004. MicroRNA regulation of gene expression in plants. *Curr Opin Plant Biol* **7**(5): 512-520.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. 2007. Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat Genet* **39**(7): 848-856.

Dunoyer, P., Himber, C., and Voinnet, O. 2005. DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat Genet* **37**(12): 1356-1360.

Earley, K., Lawrence, R.J., Pontes, O., Reuther, R., Enciso, A.J., Silva, M., Neves, N., Gross, M., Viegas, W., and Pikaard, C.S. 2006. Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes Dev* **20**(10): 1283-1293.

Ebright, R.H. and Busby, S. 1995. The Escherichia coli RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**(2): 197-203.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem* **266**(1): 71-75.

Ekwall, K. 2004a. The RITS complex-A direct link between small RNA and heterochromatin. *Mol Cell* **13**(3): 304-305.

-. 2004b. The roles of histone modifications and small RNA in centromere function. *Chromosome Res* **12**(6): 535-542.

Ekwall, K., Nimmo, E.R., Javerzat, J.P., Borgstrom, B., Egel, R., Cranston, G., and Allshire, R. 1996. Mutations in the fission yeast silencing factors clr4+ and rik1+ disrupt the localisation of the chromo domain protein Swi6p and impair centromere function. *J Cell Sci* **109 (Pt 11)**: 2637-2648.

Ekwall, K. and Ruusala, T. 1994. Mutations in rik1, clr2, clr3 and clr4 genes asymmetrically derepress the silent mating-type loci in fission yeast. *Genetics* **136**(1): 53-64.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**(20): 2539-2544.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. 2001a. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**(6836): 494-498.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. 2001b. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev* **15**(2): 188-200.

Elgin, S.C. and Grewal, S.I. 2003. Heterochromatin: silence is golden. *Curr Biol* **13**(23): R895-898.

Erhard, K.F., Jr., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J., and Hollick, J.B. 2009. RNA polymerase IV functions in paramutation in Zea mays. *Science* **323**(5918): 1201-1205.

Fang, Y. and Spector, D.L. 2007. Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol* **17**(9): 818-823.

Farago, M., Nahari, T., Hammel, C., Cole, C.N., and Choder, M. 2003. Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell* **14**(7): 2744-2755.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., and Schoehn, G. 2007. Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol Cell* **25**(6): 813-823.

Ferri, M.L., Peyroche, G., Siaut, M., Lefebvre, O., Carles, C., Conesa, C., and Sentenac, A. 2000. A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70. *Mol Cell Biol* **20**(2): 488-495.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. 1996. Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc Natl Acad Sci U S A* **93**(16): 8449-8454.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**(6669): 806-811.

Flores, A., Briand, J.F., Gadal, O., Andrau, J.C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C., Thuriaux, P., Sentenac, A., and Werner, M. 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proc Natl Acad Sci U S A* **96**(14): 7815-7820.

Forstemann, K., Horwich, M.D., Wee, L., Tomari, Y., and Zamore, P.D. 2007. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. *Cell* **130**(2): 287-297.

Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. 2002. Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci U S A* **99**(22): 14584-14589.

Fritz, J.H., Girardin, S.E., and Philpott, D.J. 2006. Innate immune defense through RNA interference. *Sci STKE* **2006**(339): pe27.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. 2006. Chromosomal histone modification patterns--from conservation to diversity. *Trends Plant Sci* **11**(4): 199-208.

Fujioka, Y., Utsumi, M., Ohba, Y., and Watanabe, Y. 2007. Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol* **48**(9): 1243-1253.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. 2004. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat Cell Biol* **6**(8): 784-791.

Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A., and Thuriaux, P. 1997. A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* **17**(4): 1787-1795.

Gadal, O., Shpakovski, G.V., and Thuriaux, P. 1999. Mutants in ABC10beta, a conserved subunit shared by all three yeast RNA polymerases, specifically affect RNA polymerase I assembly. *J Biol Chem* **274**(13): 8421-8427.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. 2005. Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. *Curr Biol* **15**(16): 1494-1500.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**(3): 495-506.

Geiduschek, E.P. and Kassavetis, G.A. 2001. The RNA polymerase III transcription apparatus. *J Mol Biol* **310**(1): 1-26.

Gentry, D., Xiao, H., Burgess, R., and Cashel, M. 1991. The omega subunit of Escherichia coli K-12 RNA polymerase is not required for stringent RNA control in vivo. *J Bacteriol* **173**(12): 3901-3903.

Girton, J.R. and Johansen, K.M. 2008. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. *Adv Genet* **61**: 1-43.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**(5523): 1876-1882.

Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., and Choder, M. 2008. Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev* **22**(15): 2022-2027.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. 2002. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. *Cell* **111**(6): 803-814.

Grandemange, S., Schaller, S., Yamano, S., Du Manoir, S., Shpakovski, G.V., Mattei, M.G., Kedinger, C., and Vigneron, M. 2001. A human RNA polymerase II subunit is encoded by a recently generated multigene family. *BMC Mol Biol* **2**: 14.

Grewal, S.I. 2000. Transcriptional silencing in fission yeast. *J Cell Physiol* **184**(3): 311-318.

Grewal, S.I., Bonaduce, M.J., and Klar, A.J. 1998. Histone deacetylase homologs regulate epigenetic inheritance of transcriptional silencing and chromosome segregation in fission yeast. *Genetics* **150**(2): 563-576.

Grewal, S.I. and Elgin, S.C. 2007. Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**(7143): 399-406.

Grewal, S.I. and Moazed, D. 2003. Heterochromatin and epigenetic control of gene expression. *Science* **301**(5634): 798-802.

Grummt, I. 2003. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* **17**(14): 1691-1702.

Guilfoyle, T.J., Hagen, G., and Malcolm, S. 1984. Immunological studies on plant DNA-dependent RNA polymerases with antibodies raised against individual subunits. *J Biol Chem* **259**(1): 640-648.

Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., and Siomi, M.C. 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science* **315**(5818): 1587-1590.

Haag, J.R., Pontes, O., and Pikaard, C.S. 2009. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS ONE* **4**(1): e4110.

Hager, G.L., Holland, M.J., and Rutter, W.J. 1977. Isolation of ribonucleic acid polymerases I, II, and III from Saccharomyces cerevisiae. *Biochemistry* **16**(1): 1-8.

Hale, C.J., Stonaker, J.L., Gross, S.M., and Hollick, J.B. 2007. A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize. *PLoS Biol* **5**(10): e275.

Hamilton, A.J. and Baulcombe, D.C. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* **286**(5441): 950-952.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**(5532): 1146-1150.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. 1998. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev* **12**(3): 357-369.

Hausner, W., Lange, U., and Musfeldt, M. 2000. Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase. *J Biol Chem* **275**(17): 12393-12399.

Haynes, K.A., Caudy, A.A., Collins, L., and Elgin, S.C. 2006. Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Curr Biol* **16**(22): 2222-2227.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev* **23**(3): 318-330.

Hemming, S.A. and Edwards, A.M. 2000. Yeast RNA polymerase II subunit RPB9. Mapping of domains required for transcription elongation. *J Biol Chem* **275**(4): 2288-2294.

Hemming, S.A., Jansma, D.B., Macgregor, P.F., Goryachev, A., Friesen, J.D., and Edwards, A.M. 2000. RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J Biol Chem* **275**(45): 35506-35511.

Henderson, I.R. and Jacobsen, S.E. 2008. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Genes Dev* **22**(12): 1597-1606.

Henikoff, S. 2000. Heterochromatin function in complex genomes. *Biochim Biophys Acta* **1470**(1): O1-8.

Henz, S.R., Cumbie, J.S., Kasschau, K.D., Lohmann, J.U., Carrington, J.C., Weigel, D., and Schmid, M. 2007. Distinct expression patterns of natural antisense transcripts in Arabidopsis. *Plant Physiol* **144**(3): 1247-1255.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**(5718): 118-120.

Hirata, A., Kanai, T., Santangelo, T.J., Tajiri, M., Manabe, K., Reeve, J.N., Imanaka, T., and Murakami, K.S. 2008a. Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a Thermococcus kodakarensis mutant lacking subunit F is temperature-sensitive. *Mol Microbiol* **70**(3): 623-633.

Hirata, A., Klein, B.J., and Murakami, K.S. 2008b. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**(7180): 851-854.

Hollender, C. and Liu, Z. 2008. Histone deacetylase genes in Arabidopsis development. *J Integr Plant Biol* **50**(7): 875-885.
Hu, P., Wu, S., Sun, Y., Yuan, C.C., Kobayashi, R., Myers, M.P., and Hernandez, N. 2002. Characterization of human RNA polymerase III identifies orthologues for Saccharomyces cerevisiae RNA polymerase III subunits. *Mol Cell Biol* **22**(22): 8044-8055.
Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16**(1): 91-93.
Huang, Y. and Maraia, R.J. 2001. Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. *Nucleic Acids Res* **29**(13): 2675-2690.
Huettel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A.J., and Matzke, M. 2007. RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim Biophys Acta* **1769**(5-6): 358-374.
Huisinga, K.L. and Elgin, S.C. 2009. Small RNA-directed heterochromatin formation in the context of development: what flies might learn from fission yeast. *Biochim Biophys Acta* **1789**(1): 3-16.
Hull, M.W., McKune, K., and Woychik, N.A. 1995. RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev* **9**(4): 481-490.
Hunter, C., Sun, H., and Poethig, R.S. 2003. The Arabidopsis heterochronic gene ZIPPY is an ARGONAUTE family member. *Curr Biol* **13**(19): 1734-1739.
Ikeda, Y. and Kinoshita, T. 2009. DNA demethylation: a lesson from the garden. *Chromosoma* **118**(1): 37-41.
Imai, K., Imazawa, Y., Yao, Y., Yamamoto, K., Hisatake, K., Muramatsu, M., and Nogi, Y. 1999. The fission yeast rpa17+ gene encodes a functional homolog of AC19, a subunit of RNA polymerases I and III of Saccharomyces cerevisiae. *Mol Gen Genet* **261**(2): 364-373.
Imazawa, Y., Hisatake, K., Mitsuzawa, H., Matsumoto, M., Tsukui, T., Nakagawa, K., Nakadai, T., Shimada, M., Ishihama, A., and Nogi, Y. 2005. The fission yeast protein Ker1p is an ortholog of RNA polymerase I subunit A14 in Saccharomyces cerevisiae and is required for stable association of Rrn3p and RPA21 in RNA polymerase I. *J Biol Chem* **280**(12): 11467-11474.
Imazawa, Y., Imai, K., Fukushima, A., Hisatake, K., Muramatsu, M., and Nogi, Y. 1999. Isolation and characterization of the fission yeast gene rpa42+, which encodes a subunit shared by RNA polymerases I and III. *Mol Gen Genet* **262**(4-5): 749-757.
Ingles, C.J., Himmelfarb, H.J., Shales, M., Greenleaf, A.L., and Friesen, J.D. 1984. Identification, molecular cloning, and mutagenesis of Saccharomyces cerevisiae RNA polymerase genes. *Proc Natl Acad Sci U S A* **81**(7): 2157-2161.
Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. 2006. Argonaute slicing is required for heterochromatic silencing and spreading. *Science* **313**(5790): 1134-1137.
Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. 2002. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**(6880): 556-560.

Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J.C., Fritzler, M.J., and Chan, E.K. 2005. Disruption of GW bodies impairs mammalian RNA interference. *Nat Cell Biol* **7**(12): 1267-1274.

James, T.C. and Elgin, S.C. 1986. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol Cell Biol* **6**(11): 3862-3872.

Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D., and Corey, D.R. 2006. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nat Struct Mol Biol* **13**(9): 787-792.

Jasencakova, Z., Soppe, W.J., Meister, A., Gernand, D., Turner, B.M., and Schubert, I. 2003. Histone modifications in Arabidopsis- high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. *Plant J* **33**(3): 471-480.

Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. 2006. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* **23**(1): 71-81.

Jasiak, A.J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strasser, K., Martin, D.E., Soding, J., and Cramer, P. 2008. Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex. *J Biol Chem* **283**(39): 26423-26427.

Jen, C.H., Michalopoulos, I., Westhead, D.R., and Meyer, P. 2005. Natural antisense transcripts with coding capacity in Arabidopsis may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol* **6**(6): R51.

Jenuwein, T. and Allis, C.D. 2001. Translating the histone code. *Science* **293**(5532): 1074-1080.

Jin, H., Vacic, V., Girke, T., Lonardi, S., and Zhu, J.K. 2008. Small RNAs and the regulation of cis-natural antisense transcripts in Arabidopsis. *BMC Mol Biol* **9**(1): 6.

Johnson, L., Cao, X., and Jacobsen, S. 2002. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* **12**(16): 1360-1367.

Jullien, P.E., Kinoshita, T., Ohad, N., and Berger, F. 2006. Maintenance of DNA methylation during the Arabidopsis life cycle is essential for parental imprinting. *Plant Cell* **18**(6): 1360-1372.

Kalantidis, K., Schumacher, H.T., Alexiadis, T., and Helm, J.M. 2008. RNA silencing movement in plants. *Biol Cell* **100**(1): 13-26.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. 2008. Transient cyclical methylation of promoter DNA. *Nature* **452**(7183): 112-115.

Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddeloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. 2003. Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics* **163**(3): 1109-1122.

Kanno, T., Bucher, E., Daxinger, L., Huettel, B., Bohmdorfer, G., Gregor, W., Kreil, D.P., Matzke, M., and Matzke, A.J. 2008. A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat Genet* **40**(5): 670-675.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**(7): 761-765.

Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol* **14**(9): 801-805.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. 2007. A novel class of bacteria-induced small RNAs in Arabidopsis. *Genes Dev* **21**(23): 3123-3134.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Jr., Zhu, J.K., Staskawicz, B.J., and Jin, H. 2006. A pathogen-inducible endogenous siRNA in plant immunity. *Proc Natl Acad Sci U S A* **103**(47): 18002-18007.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. 2005. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**(5733): 467-469.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**(7196): 793-797.

Khazak, V., Estojak, J., Cho, H., Majors, J., Sonoda, G., Testa, J.R., and Golemis, E.A. 1998. Analysis of the interaction of the novel RNA polymerase II (pol II) subunit hsRPB4 with its partner hsRPB7 and with pol II. *Mol Cell Biol* **18**(4): 1935-1945.

Kim, D.H., Villeneuve, L.M., Morris, K.V., and Rossi, J.J. 2006. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nat Struct Mol Biol* **13**(9): 793-797.

Kimura, M., Ishiguro, A., and Ishihama, A. 1997. RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *J Biol Chem* **272**(41): 25851-25855.

Kimura, M., Suzuki, H., and Ishihama, A. 2002. Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Mol Cell Biol* **22**(5): 1577-1588.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. 2004. One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation. *Science* **303**(5657): 521-523.

Kloc, A. and Martienssen, R. 2008. RNAi, heterochromatin and the cell cycle. *Trends Genet* **24**(10): 511-517.

Kloc, A., Zaratiegui, M., Nora, E., and Martienssen, R. 2008. RNA interference guides histone modification during the S phase of chromosomal replication. *Curr Biol* **18**(7): 490-495.

Kolodziej, P. and Young, R.A. 1989. RNA polymerase II subunit RPB3 is an essential component of the mRNA transcription apparatus. *Mol Cell Biol* **9**(12): 5387-5394.

Kolodziej, P.A. and Young, R.A. 1991. Mutations in the three largest subunits of yeast RNA polymerase II that affect enzyme assembly. *Mol Cell Biol* **11**(9): 4669-4678.

Koonin, E.V., Makarova, K.S., and Elkins, J.G. 2007. Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota". *Biol Direct* **2**: 38.

Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell* **131**(7): 1260-1272.
Kuo, M.H. and Allis, C.D. 1998. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20**(8): 615-626.
Kurihara, Y., Takashi, Y., and Watanabe, Y. 2006. The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* **12**(2): 206-212.
Kurihara, Y. and Watanabe, Y. 2004. Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci U S A* **101**(34): 12753-12758.
Kusser, A.G., Bertero, M.G., Naji, S., Becker, T., Thomm, M., Beckmann, R., and Cramer, P. 2008. Structure of an archaeal RNA polymerase. *J Mol Biol* **376**(2): 303-307.
Kwapisz, M., Beckouet, F., and Thuriaux, P. 2008. Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet* **24**(5): 211-215.
Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. 2009. PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci U S A* **106**(3): 941-946.
Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. 2006. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* **25**(1): 118-128.
Lange, U. and Hausner, W. 2004. Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol Microbiol* **52**(4): 1133-1143.
Langer, D., Hain, J., Thuriaux, P., and Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* **92**(13): 5768-5772.
Lanzendorfer, M., Smid, A., Klinger, C., Schultz, P., Sentenac, A., Carles, C., and Riva, M. 1997. A shared subunit belongs to the eukaryotic core RNA polymerase. *Genes Dev* **11**(8): 1037-1047.
Larkin, R. and Guilfoyle, T. 1993. The second largest subunit of RNA polymerase II from Arabidopsis thaliana. *Nucleic Acids Res* **21**(4): 1038.
Larkin, R.M. and Guilfoyle, T.J. 1996. A 14-kDa Arabidopsis thaliana RNA polymerase III subunit contains two alpha-motifs flanked by a highly charged C terminus. *Gene* **172**(2): 211-215.
-. 1997. Reconstitution of yeast and Arabidopsis RNA polymerase alpha-like subunit heterodimers. *J Biol Chem* **272**(19): 12824-12830.
-. 1998. Two small subunits in Arabidopsis RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J Biol Chem* **273**(10): 5631-5637.
Larkin, R.M., Hagen, G., and Guilfoyle, T.J. 1999. Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. *Gene* **231**(1-2): 41-47.
Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**(5): 843-854.

Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. 2004. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell* **117**(1): 69-81.
Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. 2006. An ARGONAUTE4-Containing Nuclear Processing Center Colocalized with Cajal Bodies in Arabidopsis thaliana. *Cell* **126**(1): 93-106.
Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. 2005. Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. *Curr Biol* **15**(16): 1501-1507.
Liljelund, P., Mariotte, S., Buhler, J.M., and Sentenac, A. 1992. Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **89**(19): 9302-9305.
Lin, H. and Spradling, A.C. 1997. A novel group of pumilio mutations affects the asymmetric division of germline stem cells in the Drosophila ovary. *Development* **124**(12): 2463-2476.
Lin, Y., Nomura, T., Cheong, J., Dorjsuren, D., Iida, K., and Murakami, S. 1997. Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. *J Biol Chem* **272**(11): 7132-7139.
Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G., and Rutter, W.J. 1970. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* **170**(956): 447-449.
Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. 2003. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426**(6965): 465-469.
-. 2004. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11**(6): 576-577.
Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., and Martienssen, R. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**(6998): 471-476.
Lippman, Z. and Martienssen, R. 2004. The role of RNA interference in heterochromatic silencing. *Nature* **431**(7006): 364-370.
Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133**(3): 523-536.
Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. 2004a. Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**(5689): 1437-1441.
Liu, J., He, Y., Amasino, R., and Chen, X. 2004b. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in Arabidopsis. *Genes Dev* **18**(23): 2873-2878.
Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. 2003. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. *Science* **301**(5641): 1921-1925.

Liu, Q., Yao, X., Pi, L., Wang, H., Cui, X., and Huang, H. 2008. The ARGONAUTE10 gene modulates shoot apical meristem maintenance and leaf polarity establishment by repressing miR165/166 in Arabidopsis. *Plant J.*

Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C., and Clarke, J. 2006. SERRATE: a new player on the plant microRNA scene. *EMBO Rep* **7**(10): 1052-1058.

Long, J.A., Ohno, C., Smith, Z.R., and Meyerowitz, E.M. 2006. TOPLESS regulates apical embryonic fate in Arabidopsis. *Science* **312**(5779): 1520-1523.

Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. 2007. Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15**(10): 1237-1245.

Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. 2005. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev* **19**(24): 3004-3016.

Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. 2007. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol* **178**(7): 1133-1143.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**(6648): 251-260.

Luo, J. and Hall, B.D. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* **64**(1): 101-112.

Lynn, K., Fernandez, A., Aida, M., Sedbrook, J., Tasaka, M., Masson, P., and Barton, M.K. 1999. The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis development and has overlapping functions with the ARGONAUTE1 gene. *Development* **126**(3): 469-481.

Ma, J.B., Ye, K., and Patel, D.J. 2004. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429**(6989): 318-322.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. 2005. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434**(7033): 666-670.

Macrae, I.J., Li, F., Zhou, K., Cande, W.Z., and Doudna, J.A. 2006a. Structure of Dicer and mechanistic implications for RNAi. *Cold Spring Harb Symp Quant Biol* **71**: 73-80.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. 2006b. Structural basis for double-stranded RNA processing by Dicer. *Science* **311**(5758): 195-198.

Makino, Y., Yogosawa, S., Kayukawa, K., Coin, F., Egly, J.M., Wang, Z., Roeder, R.G., Yamamoto, K., Muramatsu, M., and Tamura, T. 1999. TATA-Binding protein-interacting protein 120, TIP120, stimulates three classes of eukaryotic transcription via a unique mechanism. *Mol Cell Biol* **19**(12): 7951-7960.

Mallory, A.C., Mlotshwa, S., Bowman, L.H., and Vance, V.B. 2003. The capacity of transgenic tobacco to send a systemic RNA silencing signal depends on the nature of the inducing transgene locus. *Plant J* **35**(1): 82-92.

Mann, C., Buhler, J.M., Treich, I., and Sentenac, A. 1987. RPC40, a unique gene for a subunit shared between yeast RNA polymerases A and C. *Cell* **48**(4): 627-637.

Mann, C., Micouin, J.Y., Chiannilkulchai, N., Treich, I., Buhler, J.M., and Sentenac, A. 1992. RPC53 encodes a subunit of Saccharomyces cerevisiae RNA polymerase C (III) whose inactivation leads to a predominantly G1 arrest. *Mol Cell Biol* **12**(10): 4314-4326.

Marmorstein, R. and Trievel, R.C. 2009. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim Biophys Acta* **1789**(1): 58-68.

Martinez-Calvillo, S., Saxena, A., Green, A., Leland, A., and Myler, P.J. 2007. Characterization of the RNA polymerase II and III complexes in Leishmania major. *Int J Parasitol* **37**(5): 491-502.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. 2009. RNA-mediated chromatin-based silencing in plants. *Curr Opin Cell Biol*.

Matzke, M.A., Primig, M., Trnovsky, J., and Matzke, A.J. 1989. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J* **8**(3): 643-649.

McKune, K., Richards, K.L., Edwards, A.M., Young, R.A., and Woychik, N.A. 1993. RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast* **9**(3): 295-299.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. 2005. A structural perspective of CTD function. *Genes Dev* **19**(12): 1401-1415.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. 2004. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol Cell* **15**(2): 185-197.

Meka, H., Daoust, G., Arnvig, K.B., Werner, F., Brick, P., and Onesti, S. 2003. Structural and functional homology between the RNAP(I) subunits A14/A43 and the archaeal RNAP subunits E/F. *Nucleic Acids Res* **31**(15): 4391-4400.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., and Salbert, G. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**(7183): 45-50.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. 2000. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J* **19**(19): 5194-5201.

Meyer, W.J., Schreiber, S., Guo, Y., Volkmann, T., Welte, M.A., and Muller, H.A. 2006. Overlapping functions of argonaute proteins in patterning and morphogenesis of Drosophila embryos. *PLoS Genet* **2**(8): e134.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., Chen, S., Hannon, G.J., and Qi, Y. 2008. Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell* **133**(1): 116-127.

Milkereit, P. and Tschochner, H. 1998. A specialized form of RNA polymerase I, essential for initiation and growth-dependent regulation of rRNA synthesis, is disrupted during transcription. *EMBO J* **17**(13): 3692-3703.

Millar, A.A. and Waterhouse, P.M. 2005. Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics* **5**(3): 129-135.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. 2001. Bacterial RNA polymerase subunit omega and eukaryotic

RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**(3): 892-897.

Mitsuzawa, H., Kanda, E., and Ishihama, A. 2003. Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res* **31**(16): 4696-4701.

Moazed, D., Buhler, M., Buker, S.M., Colmenares, S.U., Gerace, E.L., Gerber, S.A., Hong, E.J., Motamedi, M.R., Verdel, A., Villen, J., and Gygi, S.P. 2006. Studies on the mechanism of RNAi-dependent heterochromatin assembly. *Cold Spring Harb Symp Quant Biol* **71**: 461-471.

Moissiard, G., Parizotto, E.A., Himber, C., and Voinnet, O. 2007. Transitivity in Arabidopsis can be primed, requires the redundant action of the antiviral Dicer-like 4 and Dicer-like 2, and is compromised by viral-encoded suppressor proteins. *RNA* **13**(8): 1268-1278.

Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marin, M.I., Martinez-Macias, M.I., Ariza, R.R., and Roldan-Arjona, T. 2006. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc Natl Acad Sci U S A* **103**(18): 6853-6858.

Morris, K.V. 2008. RNA-mediated transcriptional gene silencing in human cells. *Curr Top Microbiol Immunol* **320**: 211-224.

Morris, K.V., Chan, S.W., Jacobsen, S.E., and Looney, D.J. 2004. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**(5688): 1289-1292.

Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A* **105**(8): 3145-3150.

Mosrin, C., Riva, M., Beltrame, M., Cassar, E., Sentenac, A., and Thuriaux, P. 1990. The RPC31 gene of Saccharomyces cerevisiae encodes a subunit of RNA polymerase C (III) with an acidic tail. *Mol Cell Biol* **10**(9): 4737-4743.

Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. 2004. Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* **119**(6): 789-802.

Moussian, B., Schoof, H., Haecker, A., Jurgens, G., and Laux, T. 1998. Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. *EMBO J* **17**(6): 1799-1809.

Mukherjee, K. and Chatterji, D. 1997. Studies on the omega subunit of Escherichia coli RNA polymerase--its role in the recovery of denatured enzyme activity. *Eur J Biochem* **247**(3): 884-889.

Murfett, J., Wang, X.J., Hagen, G., and Guilfoyle, T.J. 2001. Identification of Arabidopsis histone deacetylase HDA6 mutants that affect transgene expression. *Plant Cell* **13**(5): 1047-1061.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**(5514): 110-113.

Nakazawa, Y., Hiraguri, A., Moriyama, H., and Fukuhara, T. 2007. The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway. *Plant Mol Biol* **63**(6): 777-785.

Napoli, C., Lemieux, C., and Jorgensen, R. 1990. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* **2**(4): 279-289.

Ng, S.S., Yue, W.W., Oppermann, U., and Klose, R.J. 2009. Dynamic protein methylation in chromatin biology. *Cell Mol Life Sci* **66**(3): 407-422.

Nguyen, T.N., Schimanski, B., Zahn, A., Klumpp, B., and Gunzl, A. 2006. Purification of an eight subunit RNA polymerase I complex in Trypanosoma brucei. *Mol Biochem Parasitol* **149**(1): 27-37.

Nogi, Y., Yano, R., Dodd, J., Carles, C., and Nomura, M. 1993. Gene RRN4 in Saccharomyces cerevisiae encodes the A12.2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol Cell Biol* **13**(1): 114-122.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S.I. 2004. RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nat Genet* **36**(11): 1174-1180.

Nonet, M., Sweetser, D., and Young, R.A. 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. *Cell* **50**(6): 909-915.

Nouraini, S., Archambault, J., and Friesen, J.D. 1996. Rpo26p, a subunit common to yeast RNA polymerases, is essential for the assembly of RNA polymerases I and II and for the stability of the largest subunits of these enzymes. *Mol Cell Biol* **16**(11): 5985-5996.

Numata, K., Okada, Y., Saito, R., Kiyosawa, H., Kanai, A., and Tomita, M. 2007. Comparative analysis of cis-encoded antisense RNAs in eukaryotes. *Gene* **392**(1-2): 134-141.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M.C. 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev* **18**(14): 1655-1666.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**(5): 613-622.

Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. 2008. Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* **180**(1): 207-218.

Ooi, S.K. and Bestor, T.H. 2008. The colorful history of active DNA demethylation. *Cell* **133**(7): 1145-1148.

Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. 2001. Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J Biol Chem* **276**(13): 10097-10102.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. 1997. Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent. *Cell* **90**(3): 479-490.

-. 2002. RNAi related mechanisms affect both transcriptional and posttranscriptional transgene silencing in Drosophila. *Mol Cell* **9**(2): 315-327.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. 2004. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**(5658): 669-672.

Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. 2003. Control of leaf morphogenesis by microRNAs. *Nature* **425**(6955): 257-263.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. 2005. Nuclear processing and export of microRNAs in Arabidopsis. *Proc Natl Acad Sci U S A* **102**(10): 3691-3696.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol* **12**(17): 1484-1495.

Parker, J.S., Roe, S.M., and Barford, D. 2005. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434**(7033): 663-666.

Partridge, J.F., DeBeauchamp, J.L., Kosinski, A.M., Ulrich, D.L., Hadler, M.J., and Noffsinger, V.J. 2007. Functional separation of the requirements for establishment and maintenance of centromeric heterochromatin. *Mol Cell* **26**(4): 593-602.

Pelissier, T. and Wassenegger, M. 2000. A DNA target of 30 bp is sufficient for RNA-directed DNA methylation. *RNA* **6**(1): 55-65.

Penate, X., Lopez-Farfan, D., Landeira, D., Wentland, A., Vidal, I., and Navarro, M. 2009. RNA pol II subunit RPB7 is required for RNA pol I-mediated transcription in Trypanosoma brucei. *EMBO Rep* **10**(3): 252-257.

Peng, J.C. and Karpen, G.H. 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9**(1): 25-35.

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. 2004. SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev* **18**(19): 2368-2379.

Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. 2002. The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc Natl Acad Sci U S A* **99**(23): 14670-14675.

Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. 2000. The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* **19**(20): 5473-5482.

Poethig, R.S., Peragine, A., Yoshikawa, M., Hunter, C., Willmann, M., and Wu, G. 2006. The function of RNAi in plant development. *Cold Spring Harb Symp Quant Biol* **71**: 165-170.

Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. 2006. The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* **126**(1): 79-92.

Pontes, O. and Pikaard, C.S. 2008. siRNA and miRNA processing: new functions for Cajal bodies. *Curr Opin Genet Dev* **18**(2): 197-203.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. 2005. Reinforcement of silencing at

transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* **19**(17): 2030-2040.

Prescott, E.M., Osheim, Y.N., Jones, H.S., Alen, C.M., Roan, J.G., Reeder, R.H., Beyer, A.L., and Proudfoot, N.J. 2004. Transcriptional termination by RNA polymerase I requires the small subunit Rpa12p. *Proc Natl Acad Sci U S A* **101**(16): 6068-6073.

Probst, A.V., Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., Lawrence, R.J., Pikaard, C.S., Murfett, J., Furner, I., Vaucheret, H., and Mittelsten Scheid, O. 2004. Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* **16**(4): 1021-1034.

Proshkina, G.M., Shematorova, E.K., Proshkin, S.A., Zaros, C., Thuriaux, P., and Shpakovski, G.V. 2006. Ancient origin, functional conservation and fast evolution of DNA-dependent RNA polymerase III. *Nucleic Acids Res* **34**(13): 3615-3624.

Qi, Y., Denli, A.M., and Hannon, G.J. 2005. Biochemical specialization within Arabidopsis RNA silencing pathways. *Mol Cell* **19**(3): 421-428.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., and Hannon, G.J. 2006. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* **443**(7114): 1008-1012.

Qu, F., Ye, X., and Morris, T.J. 2008. Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1. *Proc Natl Acad Sci U S A* **105**(38): 14732-14737.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): 593-599.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**(2): 192-203.

Richards, E.J. and Elgin, S.C. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**(4): 489-500.

Richmond, T.J. and Davey, C.A. 2003. The structure of DNA in the nucleosome core. *Nature* **423**(6936): 145-150.

Riddle, N.C. and Elgin, S.C. 2008. A role for RNAi in heterochromatin formation in Drosophila. *Curr Top Microbiol Immunol* **320**: 185-209.

Romano, N. and Macino, G. 1992. Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. *Mol Microbiol* **6**(22): 3343-3353.

Ronemus, M.J., Galbiati, M., Ticknor, C., Chen, J., and Dellaporta, S.L. 1996. Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* **273**(5275): 654-657.

Rose, S.D., Kim, D.H., Amarzguioui, M., Heidel, J.D., Collingwood, M.A., Davis, M.E., Rossi, J.J., and Behlke, M.A. 2005. Functional polarity is introduced by Dicer processing of short substrate RNAs. *Nucleic Acids Res* **33**(13): 4140-4156.

Rosenheck, S. and Choder, M. 1998. Rpb4, a subunit of RNA polymerase II, enables the enzyme to transcribe at temperature extremes in vitro. *J Bacteriol* **180**(23): 6187-6192.

Rubbi, L., Labarre-Mariotte, S., Chedin, S., and Thuriaux, P. 1999. Functional characterization of ABC10alpha, an essential polypeptide shared by all three forms of eukaryotic DNA-dependent RNA polymerases. *J Biol Chem* **274**(44): 31485-31492.

Runner, V.M., Podolny, V., and Buratowski, S. 2008. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol Cell Biol* **28**(6): 1883-1891.

Sadhale, P.P. and Woychik, N.A. 1994. C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Mol Cell Biol* **14**(9): 6164-6170.

Saez-Vasquez, J. and Pikaard, C.S. 1997. Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc Natl Acad Sci U S A* **94**(22): 11869-11874.

Saito, K., Ishizuka, A., Siomi, H., and Siomi, M.C. 2005. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biol* **3**(7): e235.

Sakurai, H., Mitsuzawa, H., Kimura, M., and Ishihama, A. 1999. The Rpb4 subunit of fission yeast Schizosaccharomyces pombe RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol Cell Biol* **19**(11): 7511-7518.

Sampath, V., Balakrishnan, B., Verma-Gaur, J., Onesti, S., and Sadhale, P.P. 2008. Unstructured N terminus of the RNA polymerase II subunit Rpb4 contributes to the interaction of Rpb4.Rpb7 subcomplex with the core RNA polymerase II of Saccharomyces cerevisiae. *J Biol Chem* **283**(7): 3923-3931.

Sampath, V. and Sadhale, P. 2005. Rpb4 and Rpb7: a sub-complex integral to multi-subunit RNA polymerases performs a multitude of functions. *IUBMB Life* **57**(2): 93-102.

Saze, H. and Kakutani, T. 2007. Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J* **26**(15): 3641-3652.

Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. 2008. Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana. *Science* **319**(5862): 462-465.

Schotta, G., Ebert, A., Dorn, R., and Reuter, G. 2003. Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Semin Cell Dev Biol* **14**(1): 67-75.

Schramm, L. and Hernandez, N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**(20): 2593-2620.

Selitrennik, M., Duek, L., Lotan, R., and Choder, M. 2006. Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. *Eukaryot Cell* **5**(12): 2092-2103.

Shankaranarayana, G.D., Motamedi, M.R., Moazed, D., and Grewal, S.I. 2003. Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. *Curr Biol* **13**(14): 1240-1246.

Sheffer, A., Varon, M., and Choder, M. 1999. Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. *Mol Cell Biol* **19**(4): 2672-2680.
Sheth, U. and Parker, R. 2003. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**(5620): 805-808.
Shpakovski, G.V. 1994. The fission yeast Schizosaccharomyces pombe rpb6 gene encodes the common phosphorylated subunit of RNA polymerase and complements a mutation in the corresponding gene of Saccharomyces cerevisiae. *Gene* **147**(1): 63-69.
Shpakovski, G.V., Acker, J., Wintzerith, M., Lacroix, J.F., Thuriaux, P., and Vigneron, M. 1995. Four subunits that are shared by the three classes of RNA polymerase are functionally interchangeable between Homo sapiens and Saccharomyces cerevisiae. *Mol Cell Biol* **15**(9): 4702-4710.
Shpakovski, G.V., Gadal, O., Labarre-Mariotte, S., Lebedenko, E.N., Miklos, I., Sakurai, H., Proshkin, S.A., Van Mullem, V., Ishihama, A., and Thuriaux, P. 2000. Functional conservation of RNA polymerase II in fission and budding yeasts. *J Mol Biol* **295**(5): 1119-1127.
Shpakovski, G.V. and Shematorova, E.K. 1999. Rpc19 and Rpc40, two alpha-like subunits shared by nuclear RNA polymerases I and III, are interchangeable between the fission and budding yeasts. *Curr Genet* **36**(4): 208-214.
Siaut, M., Zaros, C., Levivier, E., Ferri, M.L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. 2003. An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. *Mol Cell Biol* **23**(1): 195-205.
Sijen, T., Vijn, I., Rebocho, A., van Blokland, R., Roelofs, D., Mol, J.N., and Kooter, J.M. 2001. Transcriptional and posttranscriptional gene silencing are mechanistically related. *Curr Biol* **11**(6): 436-440.
Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes Dev* **18**(20): 2437-2468.
Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. 2000. EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in C. elegans. *Curr Biol* **10**(4): 169-178.
Smid, A., Riva, M., Bouet, F., Sentenac, A., and Carles, C. 1995. The association of three subunits with yeast RNA polymerase is stabilized by A14. *J Biol Chem* **270**(22): 13534-13540.
Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**(5): 1507-1521.
Song, J.J., Liu, J., Tolia, N.H., Schneiderman, J., Smith, S.K., Martienssen, R.A., Hannon, G.J., and Joshua-Tor, L. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10**(12): 1026-1032.
Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. 2004. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* **305**(5689): 1434-1437.

Song, L., Han, M.H., Lesicka, J., and Fedoroff, N. 2007. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* **104**(13): 5437-5442.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. 2002. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J* **21**(23): 6549-6559.

Sosunov, V., Zorov, S., Sosunova, E., Nikolaev, A., Zakeyeva, I., Bass, I., Goldfarb, A., Nikiforov, V., Severinov, K., and Mustaev, A. 2005. The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res* **33**(13): 4202-4211.

Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., and Werner, M. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**(13): 4920-4933.

Sridhar, V.V., Kapoor, A., Zhang, K., Zhu, J., Zhou, T., Hasegawa, P.M., Bressan, R.A., and Zhu, J.K. 2007. Control of DNA methylation and heterochromatic silencing by histone H2B deubiquitination. *Nature* **447**(7145): 735-738.

Steiner, N.C., Hahnenberger, K.M., and Clarke, L. 1993. Centromeres of the fission yeast Schizosaccharomyces pombe are highly variable genetic loci. *Mol Cell Biol* **13**(8): 4578-4587.

Stettler, S., Mariotte, S., Riva, M., Sentenac, A., and Thuriaux, P. 1992. An essential and specific subunit of RNA polymerase III (C) is encoded by gene RPC34 in Saccharomyces cerevisiae. *J Biol Chem* **267**(30): 21390-21395.

Stockinger, E.J., Mao, Y., Regier, M.K., Triezenberg, S.J., and Thomashow, M.F. 2001. Transcriptional adaptor and histone acetyltransferase proteins in Arabidopsis and their interactions with CBF1, a transcriptional activator involved in cold-regulated gene expression. *Nucleic Acids Res* **29**(7): 1524-1533.

Studitsky, V.M., Walter, W., Kireeva, M., Kashlev, M., and Felsenfeld, G. 2004. Chromatin remodeling by RNA polymerases. *Trends Biochem Sci* **29**(3): 127-135.

Suganuma, T. and Workman, J.L. 2008. Crosstalk among Histone Modifications. *Cell* **135**(4): 604-607.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. 1999. The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* **99**(2): 123-132.

Takahashi, K., Murakami, S., Chikashige, Y., Funabiki, H., Niwa, O., and Yanagida, M. 1992. A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol Biol Cell* **3**(7): 819-835.

Tamaru, H. and Selker, E.U. 2001. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* **414**(6861): 277-283.

Tan, Q., Li, X., Sadhale, P.P., Miyao, T., and Woychik, N.A. 2000a. Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant. *Mol Cell Biol* **20**(21): 8124-8133.

Tan, Q., Linask, K.L., Ebright, R.H., and Woychik, N.A. 2000b. Activation mutants in yeast RNA polymerase II subunit RPB3 provide evidence for a structurally

conserved surface required for activation in eukaryotes and bacteria. *Genes Dev* **14**(3): 339-348.

Tan, Q., Prysak, M.H., and Woychik, N.A. 2003. Loss of the Rpb4/Rpb7 subcomplex in a mutant form of the Rpb6 subunit shared by RNA polymerases I, II, and III. *Mol Cell Biol* **23**(9): 3329-3338.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. 2003. Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. *Proc Natl Acad Sci U S A* **100**(15): 8823-8827.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. 2007. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**(11): 1025-1040.

Tharun, S. and Parker, R. 2001. Targeting an mRNA for decapping: displacement of translation factors and association of the Lsm1p-7p complex on deadenylated yeast mRNAs. *Mol Cell* **8**(5): 1075-1083.

Thon, G., Cohen, A., and Klar, A.J. 1994. Three additional linkage groups that repress transcription and meiotic recombination in the mating-type region of Schizosaccharomyces pombe. *Genetics* **138**(1): 29-38.

Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. 1995. A mutation in the C31 subunit of Saccharomyces cerevisiae RNA polymerase III affects transcription initiation. *EMBO J* **14**(2): 351-359.

Thuriaux, P., Mariotte, S., Buhler, J.M., Sentenac, A., Vu, L., Lee, B.S., and Nomura, M. 1995. Gene RPA43 in Saccharomyces cerevisiae encodes an essential subunit of RNA polymerase I. *J Biol Chem* **270**(41): 24252-24257.

Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. 2004. RISC assembly defects in the Drosophila RNAi mutant armitage. *Cell* **116**(6): 831-841.

Tomari, Y., Du, T., and Zamore, P.D. 2007. Sorting of Drosophila small silencing RNAs. *Cell* **130**(2): 299-308.

Tournier, B., Tabler, M., and Kalantidis, K. 2006. Phloem flow strongly influences the systemic spread of silencing in GFP Nicotiana benthamiana plants. *Plant J* **47**(3): 383-394.

Treich, I., Riva, M., and Sentenac, A. 1991. Zinc-binding subunits of yeast RNA polymerases. *J Biol Chem* **266**(32): 21971-21976.

Ujvari, A. and Luse, D.S. 2006. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* **13**(1): 49-54.

Ulmasov, T. and Guilfoyle, T.J. 1992. Sequence of the fifth largest subunit of RNA polymerase II from plants. *J Biol Chem* **267**(32): 23165-23169.

Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. 1995. Arabidopsis expresses two genes that encode polypeptides similar to the yeast RNA polymerase I and III AC40 subunit. *Gene* **167**(1-2): 203-207.

-. 1996. Association between 36- and 13.6-kDa alpha-like subunits of Arabidopsis thaliana RNA polymerase II. *J Biol Chem* **271**(9): 5085-5094.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* **313**(5785): 320-324.

Vaistij, F.E., Jones, L., and Baulcombe, D.C. 2002. Spreading of RNA targeting and DNA methylation in RNA silencing requires transcription of the target gene and a putative RNA-dependent RNA polymerase. *Plant Cell* **14**(4): 857-867.

van Rij, R.P., Saleh, M.C., Berry, B., Foo, C., Houk, A., Antoniewski, C., and Andino, R. 2006. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. *Genes Dev* **20**(21): 2985-2995.

Vazquez, F. 2006. Arabidopsis endogenous small RNAs: highways and byways. *Trends Plant Sci* **11**(9): 460-468.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. 2004. Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* **16**(1): 69-79.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303**(5658): 672-676.

Verma-Gaur, J., Rao, S.N., Taya, T., and Sadhale, P. 2008. Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation. *Eukaryot Cell* **7**(6): 1009-1018.

Vlachonasios, K.E., Thomashow, M.F., and Triezenberg, S.J. 2003. Disruption mutations of ADA2b and GCN5 transcriptional adaptor genes dramatically affect Arabidopsis growth, development, and gene expression. *Plant Cell* **15**(3): 626-638.

Voinnet, O. 2005. Non-cell autonomous RNA silencing. *FEBS Lett* **579**(26): 5858-5871.

Volpe, T., Schramke, V., Hamilton, G.L., White, S.A., Teng, G., Martienssen, R.A., and Allshire, R.C. 2003. RNA interference is required for normal centromere function in fission yeast. *Chromosome Res* **11**(2): 137-146.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**(5588): 1833-1837.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. 1993. Arabidopsis thaliana DNA methylation mutants. *Science* **260**(5116): 1926-1928.

Voutsina, A., Riva, M., Carles, C., and Alexandraki, D. 1999. Sequence divergence of the RNA polymerase shared subunit ABC14.5 (Rpb8) selectively affects RNA polymerase III assembly in Saccharomyces cerevisiae. *Nucleic Acids Res* **27**(4): 1047-1055.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S., and Handa, H. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**(3): 343-356.

Wallace, J.A. and Orr-Weaver, T.L. 2005. Replication of heterochromatin: insights into mechanisms of epigenetic inheritance. *Chromosoma* **114**(6): 389-402.

Wallrath, L.L. and Elgin, S.C. 1995. Position effect variegation in Drosophila is associated with an altered chromatin structure. *Genes Dev* **9**(10): 1263-1277.

Wang, X.H., Aliyari, R., Li, W.X., Li, H.W., Kim, K., Carthew, R., Atkinson, P., and Ding, S.W. 2006. RNA interference directs innate immunity against viruses in adult Drosophila. *Science* **312**(5772): 452-454.

Wang, X.J., Gaasterland, T., and Chua, N.H. 2005. Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana. *Genome Biol* **6**(4): R30.

Wang, X.J., Reyes, J.L., Chua, N.H., and Gaasterland, T. 2004. Prediction and identification of Arabidopsis thaliana microRNAs and their mRNA targets. *Genome Biol* **5**(9): R65.

Wang, Z. and Roeder, R.G. 1997. Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev* **11**(10): 1315-1326.

Wassenegger, M., Heimes, S., Riedel, L., and Sanger, H.L. 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**(3): 567-576.

Waterhouse, P.M., Graham, M.W., and Wang, M.B. 1998. Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. *Proc Natl Acad Sci U S A* **95**(23): 13959-13964.

Waterhouse, P.M., Wang, M.B., and Lough, T. 2001. Gene silencing as an adaptive defence against viruses. *Nature* **411**(6839): 834-842.

Weinberg, M.S., Villeneuve, L.M., Ehsani, A., Amarzguioui, M., Aagaard, L., Chen, Z.X., Riggs, A.D., Rossi, J.J., and Morris, K.V. 2006. The antisense strand of small interfering RNAs directs histone methylation and transcriptional gene silencing in human cells. *RNA* **12**(2): 256-262.

Werner, A. and Berdal, A. 2005. Natural antisense transcripts: sound or silence? *Physiol Genomics* **23**(2): 125-131.

Werner, F. 2007. Structure and function of archaeal RNA polymerases. *Mol Microbiol* **65**(6): 1395-1404.

Werner, F. and Weinzierl, R.O. 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* **10**(3): 635-646.

Werner, M., Chaussivert, N., Willis, I.M., and Sentenac, A. 1993. Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem* **268**(28): 20721-20724.

Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriaux, P. 1992. Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol Cell Biol* **12**(3): 1087-1095.

White, S.A. and Allshire, R.C. 2008. RNAi-mediated chromatin silencing in fission yeast. *Curr Top Microbiol Immunol* **320**: 157-183.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**(4): 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**(5): 630-634.

Wind, M. and Reines, D. 2000. Transcription elongation factor SII. *Bioessays* **22**(4): 327-336.

Woychik, N.A. and Hampsey, M. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**(4): 453-463.

Woychik, N.A., Lane, W.S., and Young, R.A. 1991. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* **266**(28): 19053-19055.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* **4**(3): 313-323.

Woychik, N.A. and Young, R.A. 1989. RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol Cell Biol* **9**(7): 2854-2859.

-. 1990. RNA polymerase II subunit RPB10 is essential for yeast cell viability. *J Biol Chem* **265**(29): 17816-17819.

Wu, K., Zhang, L., Zhou, C., Yu, C.W., and Chaikam, V. 2008. HDA6 is required for jasmonate response, senescence and flowering in Arabidopsis. *J Exp Bot* **59**(2): 225-234.

Xie, Z., Allen, E., Wilken, A., and Carrington, J.C. 2005. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* **102**(36): 12984-12989.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. 2004. Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol* **2**(5): E104.

Yamamoto, R.T., Nogi, Y., Dodd, J.A., and Nomura, M. 1996. RRN3 gene of Saccharomyces cerevisiae encodes an essential RNA polymerase I transcription factor which interacts with the polymerase independently of DNA template. *EMBO J* **15**(15): 3964-3973.

Yang, L., Liu, Z., Lu, F., Dong, A., and Huang, H. 2006a. SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J* **47**(6): 841-850.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. 2006b. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res* **34**(2): 667-675.

Yoshikawa, M., Peragine, A., Park, M.Y., and Poethig, R.S. 2005. A pathway for the biogenesis of trans-acting siRNAs in Arabidopsis. *Genes Dev* **19**(18): 2164-2175.

Young, R.A. and Davis, R.W. 1983. Yeast RNA polymerase II genes: isolation with antibody probes. *Science* **222**(4625): 778-782.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R., and Chen, X. 2005. Methylation as a crucial step in plant microRNA biogenesis. *Science* **307**(5711): 932-935.

Yura, T. and Ishihama, A. 1979. Genetics of bacterial RNA polymerases. *Annu Rev Genet* **13**: 59-97.

Zaros, C., Briand, J.F., Boulard, Y., Labarre-Mariotte, S., Garcia-Lopez, M.C., Thuriaux, P., and Navarro, F. 2007. Functional organization of the Rpb5 subunit shared by the three yeast RNA polymerases. *Nucleic Acids Res* **35**(2): 634-647.

Zaros, C. and Thuriaux, P. 2005. Rpc25, a conserved RNA polymerase III subunit, is critical for transcription initiation. *Mol Microbiol* **55**(1): 104-114.

Zaychikov, E., Martin, E., Denissova, L., Kozlov, M., Markovtsov, V., Kashlev, M., Heumann, H., Nikiforov, V., Goldfarb, A., and Mustaev, A. 1996. Mapping of catalytic residues in the RNA polymerase active center. *Science* **273**(5271): 107-109.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. 1999. Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98**(6): 811-824.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. 2007. Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J* **26**(6): 1691-1701.

Zhimulev, I.F. and Belyaeva, E.S. 2003. Intercalary heterochromatin and genetic silencing. *Bioessays* **25**(11): 1040-1051.

Zilberman, D., Cao, X., and Jacobsen, S.E. 2003. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299**(5607): 716-719.

Zofall, M. and Grewal, S.I. 2006. RNAi-mediated heterochromatin assembly in fission yeast. *Cold Spring Harb Symp Quant Biol* **71**: 487-496.

CHAPTER TWO

PLANT NUCLEAR RNA POLYMERASE IV MEDIATES siRNA-DIRECTED DNA METHYLATION AND HETEROCHROMATIN FORMATION

Published in *Cell*, vol. 120;5 613-622

My contributions to this work:

In this work, I designed and performed all of the DNA methylation experiments that demonstrated the requirement for the *NRPD2a/NRPD2* and *NRPD1a/NRPD1* genes in siRNA-directed DNA methylation and which placed Pol IV in the 24-nt heterochromatic siRNA pathway, along with RDR2, DCL3 and DRM2. This included methylation-sensitive southern blot analyses and Chop-PCR analyses in Figure 4A-C. I obtained and confirmed the *dcl3*, *rdr2* and mutants used in Figure 4, and helped with screening of *nrpd1a/nrpd1* mutant lines used in Figure 4. I also assisted with the editing of the paper and response to reviewer comments and suggestions.

Cell, Vol. 120, 613–622, March 11, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.02.007

Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation

Yasuyuki Onodera,[1,2,4] Jeremy R. Haag,[1,4]
Thomas Ream,[1,4] Pedro Costa Nunes,[3]
Olga Pontes,[1] and Craig S. Pikaard[1,*]
[1]Biology Department
Washington University
1 Brookings Drive
St. Louis, Missouri 63130
[2]Graduate School of Agriculture
Faculty of Agriculture
Hokkaido University
Kita 9, Nishi 9, Kita-ku
Sapporo 060-8589
Japan
[3]Secção de Genética
Centro de Botanica e Engenharia Biologica
Instituto Superior de Agronomia
Tapada da Ajuda
1349-017 Lisboa
Portugal

Summary

All eukaryotes have three nuclear DNA-dependent RNA polymerases, namely, Pol I, II, and III. Interestingly, plants have catalytic subunits for a fourth nuclear polymerase, Pol IV. Genetic and biochemical evidence indicates that Pol IV does not functionally overlap with Pol I, II, or III and is nonessential for viability. However, disruption of the Pol IV catalytic subunit genes *NRPD1* or *NRPD2* inhibits heterochromatin association into chromocenters, coincident with losses in cytosine methylation at pericentromeric 5S gene clusters and *AtSN1* retroelements. Loss of CG, CNG, and CNN methylation in Pol IV mutants implicates a partnership between Pol IV and the methyltransferase responsible for RNA-directed de novo methylation. Consistent with this hypothesis, 5S gene and *AtSN1* siRNAs are essentially eliminated in Pol IV mutants. The data suggest that Pol IV helps produce siRNAs that target de novo cytosine methylation events required for facultative heterochromatin formation and higher-order heterochromatin associations.

Introduction

In eukaryotes, three nuclear DNA-dependent RNA polymerases (RNAPs) transcribe genomic DNA into RNA. RNA polymerase I (Pol I) transcribes the ribosomal RNA (rRNA) genes clustered at nucleolus organizer regions (Grummt, 2003); RNA polymerase II (Pol II) transcribes the vast majority of genes, including protein-coding genes (Woychik and Hampsey, 2002), and RNA polymerase III (Pol III) transcribes genes encoding short (<400 nt) structural RNAs that include tRNAs and 5S rRNA (Schramm and Hernandez, 2002).

RNA polymerases I, II, and III are composed of 12–17 proteins, including subunits sharing sequence and structural homology with the eubacterial RNA polymerase subunits β', β, α^{I}, α^{II}, and ω (Archambault and Friesen, 1993; Cramer et al., 2001; Zhang et al., 1999). RNA Pol I, II, and III (designated RPA, RPB, and RPC in yeast and N [nuclear] RPA, NRPB, and NRPC in *Arabidopsis*) largest subunits are homologous to eubacterial β' and are encoded by different genes, *(N)RPA1*, *(N)RPB1*, and *(N)RPC1*. Likewise, the second-largest subunits of Pol I, II, and III are β homologs encoded by *(N)RPA2*, *(N)RPB2*, and *(N)RPC2*. Together, the largest and second-largest subunits form the catalytic center in which RNA synthesis occurs (Cramer et al., 2000; Zhang et al., 1999), with α^{I}, α^{II}, and ω serving regulatory or assembly functions.

Surprisingly, analysis of the *Arabidopsis thaliana* genome sequence revealed evidence for a fourth class of RNA polymerase in addition to Pol I, II, and III (CSP and Jonathan Eisen, discussed in *Arabidopsis* Genome Initiative [2000]). Specifically, two class IV largest and second-largest subunit genes were predicted, implying the existence of a nuclear RNA polymerase IV (Pol IV) distinct from eubacterial-type RNAPs of chloroplasts, from mitochondrial polymerase, or from RNA-dependent RNA polymerases (RdRP).

Here, we present evidence that RNA Pol IV is located within the nucleus and plays a role in heterochromatin formation. Dispersal of chromocenters in Pol IV mutants is correlated with the loss of cytosine methylation from pericentromeric 5S gene clusters and *AtSN1* retroelements. By contrast, methylation of constitutively heterochromatic 180 bp centromere core repeats is not appreciably affected in Pol IV mutants. We propose that Pol IV is required for the production of siRNAs that direct de novo methylation of repetitive elements that are subject to facultative heterochromatin formation, thereby facilitating higher-order heterochromatin associations.

Results

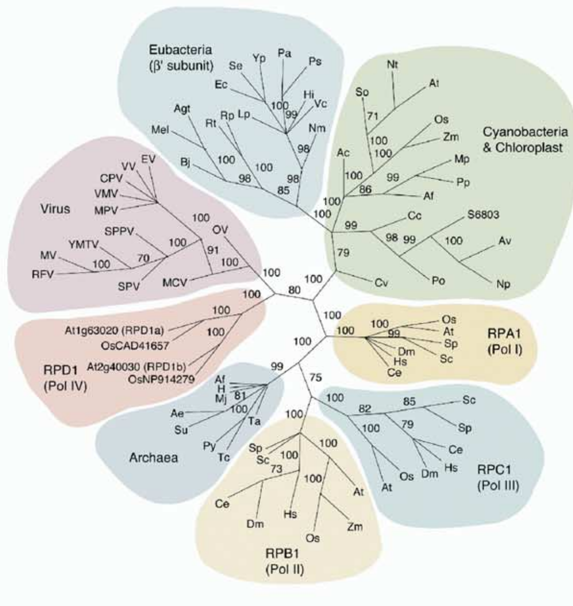
Genes for RNA Pol IV

An unrooted phylogenetic tree of DNA-dependent RNA polymerase (RNAP) largest subunits (Figure 1A) reveals distinct clades for eubacteria, cyanobacteria and chloroplasts, archaea, DNA viruses, and eukaryotic RNA polymerases I (RPA1), II (RPB1), and III (RPC1). *Arabidopsis thaliana* (At) Pol I, II, and III largest subunits group with their orthologs from rice (Os), yeast (Sp and Sc), *C. elegans* (Ce), *Drosophila* (Dm), and human (Hs). Unlike other eukaryotes, *Arabidopsis* and rice have additional genes (*NRPD1a* and *b*) that form a clade for a putative Pol IV.

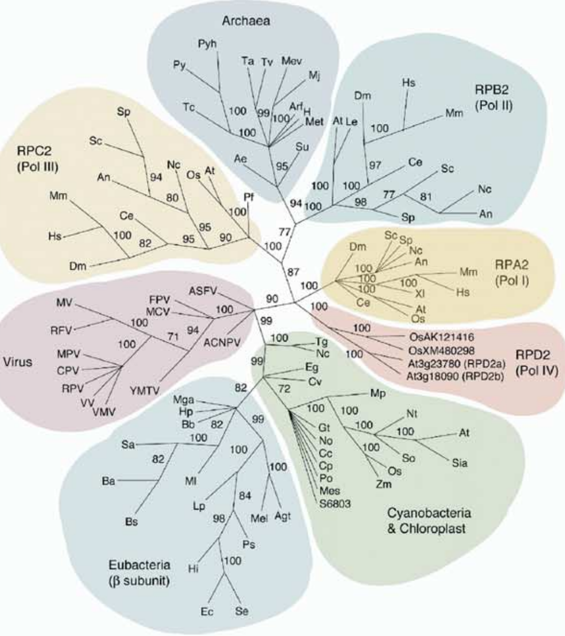
An unrooted tree of RNAP second-largest subunits resembles the tree for the largest subunits (Figure 1B). Again, in addition to clades for *RPA2* (Pol I), *RPB2* (Pol II), and *RPC2* (Pol III), a plant-specific *NRPD2* (Pol IV)

*Correspondence: pikaard@biology.wustl.edu
[4]These authors contributed equally to this work.

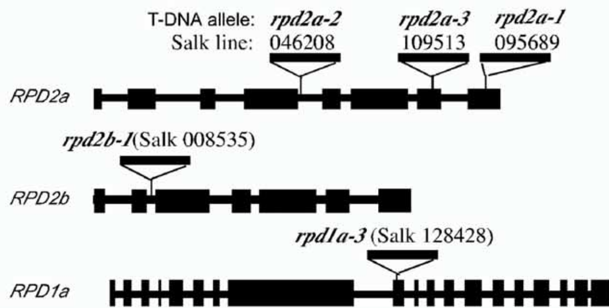
A RNAP largest subunits



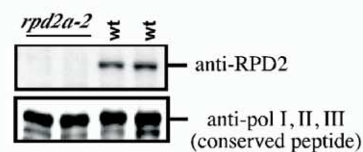
B RNAP second-largest subunits



C T-DNA disrupted alleles



D Immunoblot



E RPD2 immunolocalization

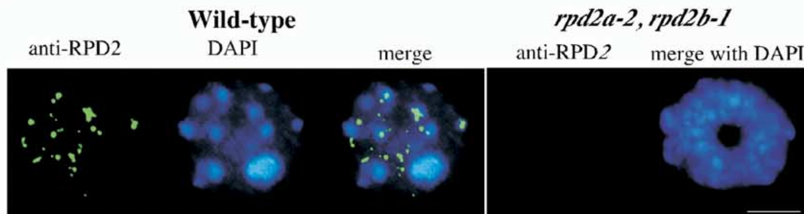


Figure 1. Evidence for RNA Pol IV in Plants

(A and B) Unrooted neighbor-joining phylogenies based on conserved domains A, C, D, and F of DNA-dependent RNA polymerase largest subunits and conserved domains A, C, D, F, G, H, and I of DNA-dependent RNA polymerase second-largest subunits. Bootstrap values are given for branch nodes. Species designations and GenBank accession numbers for the sequences analyzed are provided in Tables S1 and S2.

(C) Diagrams of T-DNA-disrupted *nrpd2* and *nrpd1* alleles. Exons are denoted by black rectangles.

(D) Immunoblot showing no detectable NRPD2 protein in two *nrpd2a-2* mutant individuals, unlike their wild-type siblings. A control immunoblot utilized an antibody raised against a peptide conserved in Pol I, II, and III second-largest subunits.

(E) NRPD2 localizes to the nucleus. On the left is a wild-type interphase nucleus showing immunolocalization of NRPD2 relative to ten DAPI-positive chromocenters. On the right is a homozygous *nrpd2a-1 nrpd2b-1* nucleus. The dark, DAPI-negative region is the nucleolus. The wild-type and mutant plants were progeny of homozygous siblings. The size bar corresponds to 5 μm.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

clade exists. In both *Arabidopsis* and rice, there are two *NRPD2* genes (*NRPD2a* and *NRPD2b*) that were apparently duplicated after monocots and dicots diverged.

Multiple alignments revealed that NRPD2 proteins closely resemble their Pol I–III homologs, whereas NRPD1 sequences frequently lack amino acids that are invariant in Pol I–III largest subunits, including amino acids near the active site (see Figures S1–S4 in the Supplemental Data available with this article online). Therefore, we focused our studies on NRPD2 but also subjected *nrpd1a* mutants to a subset of the same assays. *NRPD1b* was ignored because existing annotation suggested that this gene lacks essential C-terminal domains.

Only *NRPD2a* appears to be expressed in *Arabidopsis*, based on existing EST (cDNA) sequences and by our inability to amplify *NRPD2b* RNA using RT-PCR or 5′ RACE. By contrast, *NRPD2a* sequences were readily amplified by PCR and by primer extension (Figure S5) to yield a full-length mRNA sequence (GenBank accession number AY862891).

Salk lines 046208, 109513, and 095689 contain the T-DNA-disrupted mutant alleles *nrpd2a-2*, *nrpd2a-3*, and *nrpd2a-1*, respectively. Salk lines 008535 and 128428 contain the *nrpd2b-1* and *nrpd1a-3* alleles (Figure 1C). Plants homozygous for these alleles were identified by PCR or Southern blot analysis of segregating families. The *nrpd2a* and *nrpd1a* alleles are all recessive and cause equivalent molecular phenotypes (data below and data not shown).

NRPD2 Expression and Nuclear Localization

RNA and protein blot analyses showed that NRPD2a is expressed throughout the plant but is most highly expressed in flowers and roots (data not shown). In homozygous *nrpd2a-2* mutants, no NRPD2 protein is detectable (Figure 1D), indicating that *nrpd2a-2* is a null allele. Immunolocalization of NRPD2 showed it to be a nuclear protein that is concentrated in numerous distinct foci (Figure 1E). Examination of 56 interphase nuclei revealed 10–15 NRPD2 signals in 71% of the nuclei and fewer than ten signals in 29% of the nuclei. In the nucleus shown, there are ten prominent DAPI-positive heterochromatic chromocenters, which are made up of centromeric repeats for the ten chromosomes, dispersed pericentromeric repeats, and four NORs (nucleolus organizer regions) (Fransz et al., 2002). Approximately 15 NRPD2 signals of varying size are apparent in Figure 1E, five of which are located at chromocenters and five of which are at the edges of chromocenters. Similar association of NRPD2 with chromocenters was observed in all nuclei.

Genetic Analysis of NRPD Mutants

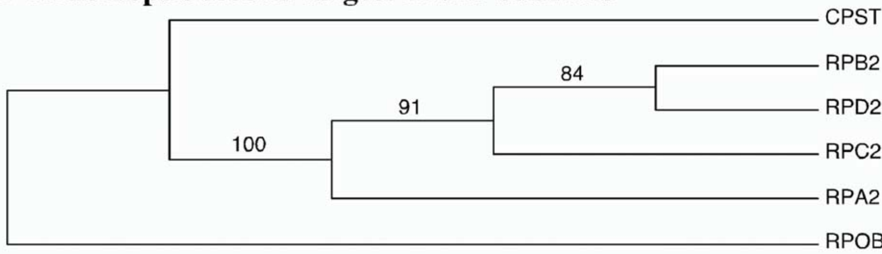
To rule out any possible functional redundancy of *NRPD2a* and *NRPD2b*, we generated lines homozygous for both the *nrpd2a-2* and *nrpd2b-1* alleles, which was laborious, because the genes are linked (∼10 cM genetic distance). We first crossed *nrpd2a-2* and *nrpd2b-1* homozygotes to generate F1 individuals that were hemizygous for each allele. The F1 was then outcrossed with a wild-type plant such that all resulting progeny had a wild-type chromosome 3 and either an *nrpd2a-2* or an *nrpd2b-1* allele but not both, unless a meiotic recombination event occurred between the two genes. We then identified the latter rare recombinants that had one wild-type chromosome 3 and one chromosome 3 bearing both the *nrpd2a-2* and *nrpd2b-1* alleles, allowed these to self-fertilize, and genotyped their progeny. Plants homozygous for both *nrpd2a -2* and *nrpd2b-1* (referred to as *nrpd2* double mutants or simply *nrpd2* in the remainder of the paper) were recovered, demonstrating that *NRPD2* is nonessential for viability. Siblings that were homozygous for the wild-type *NRPD2* gene were also identified and used as controls in subsequent assays. This genetic strategy is likely to have segregated away any potential T-DNAs unlinked to *NRPD2*, but, if such T-DNAs persist, they are as likely in the wild-type control plants as in their double mutant siblings.

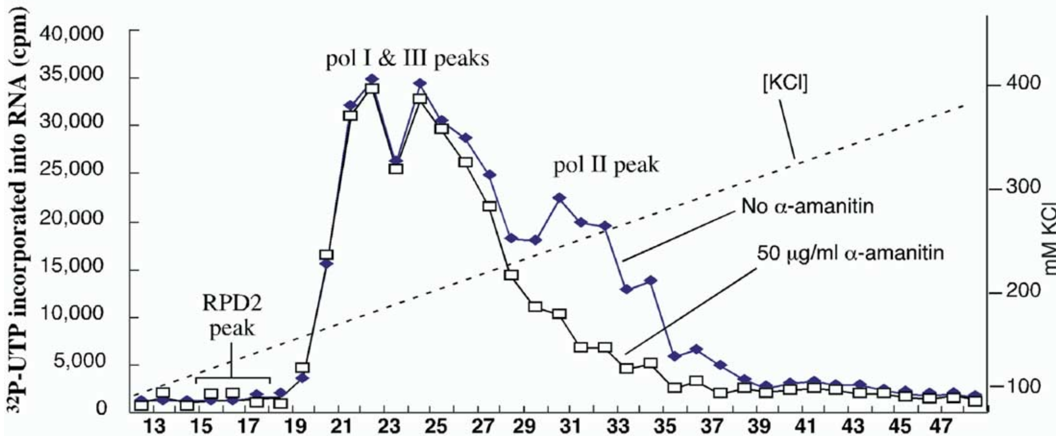
We tested whether NRPD2 might be functionally redundant with the NRPA2, NRPB2, or NRPC2 subunits of Pol I–III by asking if any of these subunits were nonessential. We identified hemizygous individuals bearing T-DNA insertions in *NRPA2*, *NRPB2*, or *NRPC2* and genotyped 60–80 of their progeny. Only homozygous wild-type and hemizygous progeny were obtained; no homozygous mutants were recovered (data not shown). These results indicate that *NRPA2*, *NRPB2*, and *NRPC2* are essential genes, unlike *NRPD2a* and *NRPD2b*, and that *NRPD2* genes do not complement *nrpa2*, *nrpb2*, or *nrpc2* mutations. The *nrpd2* double mutation also failed to induce haploinsufficiency in plants hemizygous for *nrpa2*, *nrpb2*, or *nrpc2* mutations, consistent with the interpretation that NRPD2 does not overlap functionally with Pol I, II, or III.

NRPD2 Does Not Copurify with DNA-Dependent RNA Polymerases I–III

Among *Arabidopsis* RNAP second-largest subunits, NRPD2 is most similar to NRPB2 (Figure 2A). Therefore, we asked if NRPD2 copurified with RNA Pol II activity, as might be expected if NRPD2 is an alternative Pol II subunit. Nuclear extract was fractionated by anion exchange chromatography, and fractions were tested for DNA-dependent RNA polymerase activity (Figure 2B) and for the presence of NRPD2, NRPB2, or a 24 kDa polymerase subunit (RPB5) that is shared by Pol I, II, and III (Larkin et al., 1999; Saez-Vasquez and Pikaard, 2000).

The DNA-dependent RNA polymerase assay measures the incorporation of radioactive nucleotide triphosphates into RNA using sheared template DNA, which allows polymerase initiation from broken DNA ends in a promoter-independent fashion (Schwartz and Roeder, 1974). Duplicate reactions were performed with and without α-amanitin, a potent inhibitor of RNA Pol II, and mean values were plotted (Figure 2B). Comparison of the RNA polymerase activity profiles reveals a peak of activity that is inhibited by α-amanitin (fractions 29–37), indicative of Pol II (Figure 2B). As expected, NRPB2 eluted in these fractions (Figure 2C). By contrast, NRPD2 eluted in fractions 15–18, suggesting that NRPD2 is not an alternative Pol II subunit. Immunoblotting of column fractions using an antibody against the 24 kDa subunit that is shared by Pol I, II, and III revealed a good correspondence between the presence of the





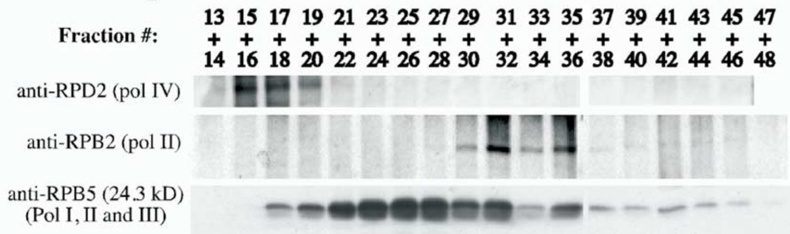


Figure 2. NRPD2 Does Not Cofractionate with Pol II or with DNA-Dependent RNA Polymerase Activity

(A) Neighbor-joining tree (with bootstrap values based on 1000 replications) for second-largest subunits of *Arabidopsis* chloroplast RNAP and RNA polymerases I, II, and III. The *E. coli* RpoB subunit serves as the outgroup.

(B) Fractionation of DNA-dependent RNA polymerase activity by DEAE-Sepharose chromatography. Fractions eluted with a linear KCl gradient were tested for RNA polymerase activity both with and without α-amanitin.

(C) Immunoblot detection of NRPD2, NRPB2, and NRPB5 in fractions eluted from the DEAE column.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

24 kDa subunit and RNAP activity. Surprisingly, the peak fractions for NRPD2a displayed no detectable RNAP activity. We conclude that NRPD2 is not an alternative subunit of a conventional DNA-dependent RNA polymerase.

Heterochromatin Association Is Impaired in nrpd2 Mutants

In *nrpd2* mutants, we noted an increased number and decreased size of DAPI-positive heterochromatic foci in interphase nuclei relative to wild-type siblings (Figure 1E), prompting further investigation. Histone H3 dimethylated on lysine 9 ($H3^{dimethyl}K9$) is a marker of heterochromatin (Richards and Elgin, 2002) that colocalizes with chromocenters in wild-type nuclei (Figure 3A). However, in *nrpd2* mutant siblings, the $H3^{dimethyl}K9$ signals are dispersed and colocalize with the numerous, small DAPI-positive foci (Figure 3A; Table S3).

Chromocenters involving NORs are relatively resistant to dispersal (Figure 3B). It is noteworthy that there are four NORs in a diploid nucleus, located at the tips of chromosomes 2 and 4. However, 36% of wild-type and 19% of *nrpd2* interphase nuclei show only two NOR fluorescence in situ hybridization (FISH) signals (as in Figure 3B) due to association of pairs of NORs and their linked centromeres. Nuclei with either three or four NOR FISH signals are also observed in wild-type and *nrpd2* mutants, but only *nrpd2* mutants frequently

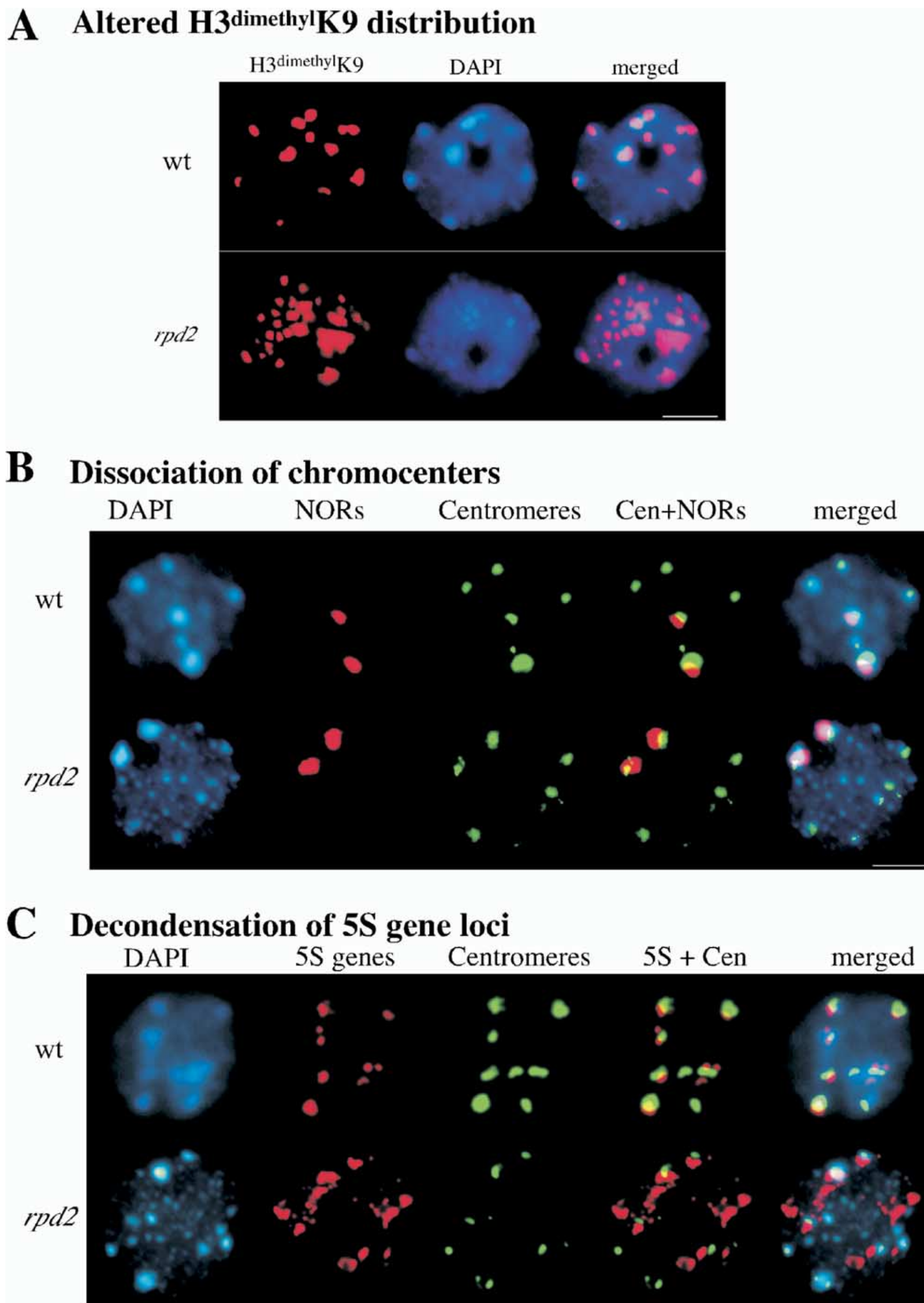


Figure 3. Heterochromatin Is Disrupted in *nrpd2* Mutants

(A) Immunolocalization of histone H3 dimethylated on lysine 9 in interphase cells of wild-type and the *nrpd2a-2 nrpd2b-1* mutant. Chromatin was counterstained with DAPI.

(B) Chromocenters containing NORs are relatively resistant to dispersal in *nrpd2a-2 nrpd2b-1* mutants. Centromeres and NORs (45S rRNA gene loci) were detected by FISH. Chromatin was counterstained with DAPI.

(C) 5S gene loci become decondensed and dissociated from centromeres in *nrpd2a-2 nrpd2b-1*double mutants. 5S genes and centromeres were detected by FISH. Wild-type and mutant plants were progeny of homozygous siblings. Size bars in all panels correspond to 5 μm. Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

(23%) show >4 NOR signals (Table S3), presumably due to dissociation of facultative heterochromatin subdomains of the ~4 Mbp NORs.

5S rRNA gene repeats are tandemly arranged in pericentromeric regions of chromosomes 3, 4, and 5 in *Arabidopsis* ecotype Col-0 such that dual FISH typically reveals substantial overlap of 5S and 180 bp centromere repeat signals in wild-type cells (Figure 3C).

However, in *nrpd2* double mutant siblings, the 5S genes are typically decondensed and show significantly less (p = 0.0012) colocalization with centromeres, consistent with the interpretation that pericentromeric facultative heterochromatin is dispersed away from the constitutively heterochromatic centromeres (see Table S3 for quantitation).

Pol IV Participates in the siRNA-Chromatin Modification Pathway

Heterochromatin disruption and 5S gene dispersal in Pol IV mutants suggested a possible loss of cytosine methylation (Soppe et al., 2002). To determine if *nrpd2* or *nrpd1a* mutants affect 5S gene cytosine methylation, we performed Southern blotting using methylation-sensitive restriction endonucleases. HpaII and MspI cut CCGG motifs, but HpaII will not cut if the inner C is methylated, and MspI will not cut if the outer C is methylated (McClelland et al., 1994). HaeIII recognizes GGCC but won't cut if the inner C is methylated. Digestion of 5S genes with these three enzymes reports on methylation at CG (HpaII), CNG (MspI), and CNN (in the ecotype Col-0, the 5S HaeIII site is a CNN site). The Southern blots reveal ladders of bands at ∼500 bp intervals (Figure 4A), the size of a 5S gene repeat (Campell et al., 1992). High levels of methylation cause most of the hybridization signal to be near the top of the ladder, whereas loss of methylation results in more signal near the bottom.

5S gene methylation at HpaII, MspI, and HaeIII sites is decreased in *nrpd1a-3* and *nrpd2* mutants (Figure 4A, lanes 3, 5, 18, 20, 22, and 24) relative to their wild-type siblings (lanes 2, 4, 19, 21, 23, and 25), with HaeIII digestion showing the largest effect. Comparison of *nrpd1* and *nrpd2* to the DNA methylation mutants *ddm1*, *met1*, *cmt3*, and *drm1drm2* showed that HpaII digestion of 5S genes in *nrpd1* and *nrpd2* mutants occurred to the same extent as in a *drm1drm2* double mutant (compare lanes 3, 5, and 6) but to a lesser extent than in a *ddm1* (lane 10) or *met1* (lane 11) mutant. DRM2 is responsible for de novo methylation in all sequence contexts (CG, CNG, and CNN); DDM1 is involved in maintenance of methylation in all sequence contexts, and MET1 is primarily responsible for maintenance of CG methylation (reviewed in Bender [2004]). DRM1 has no known function. CMT3 is primarily responsible for maintenance of CNG methylation, so a *CMT3* mutant has little effect on HpaII digestion (lane 7) but has a profound effect on MspI digestion (lane 16). Collectively, the results indicate that Pol IV affects 5S gene methylation in all sequence contexts (CG, CNG, and CNN). Interestingly, the highly methylated 180 bp centromere repeats are unaffected by *nrpd1* and *nrpd2* mutations (Figure 4B), suggesting that Pol IV does not affect global cytosine methylation levels but acts on only a subset of methylated genomic sequences.

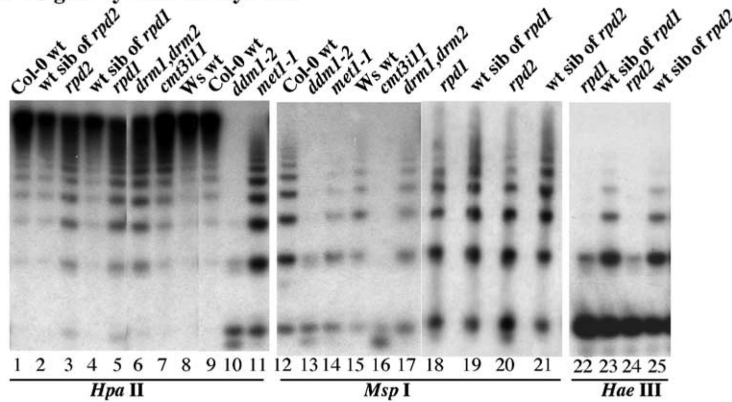
Methylation of *AtSN1*, a well-characterized retroelement family (Hamilton et al., 2002; Xie et al., 2004), was assayed using HaeIII digestion followed by PCR (Figure 4C) (Hamilton et al., 2002). If HaeIII sites are methylated, the DNA is not cut and can be amplified. However, if CNN methylation is lost at any of three HaeIII sites (see diagram), HaeIII digestion precludes PCR amplification. In wild-type Col-0, Ler, or Ws (the genetic backgrounds for the mutants tested), *AtSN1* elements are heavily methylated and resistant to HaeIII cleavage. Methylation is unaffected by *met1* or *cmt3* mutants but is substantially reduced in a *drm1 drm2* double mutant, as expected for CNN methylation. HaeIII methylation is also disrupted in mutants of the heterochromatic siRNA pathway, including *rdr2* (RNA-dependent RNA polymerase 2), *hen1* (Hua enhancer 1), or *dcl3* (Dicer-like 3), consistent with published results (Xie et al., 2004). By contrast, *AtSN1* methylation is not diminished in a mutant of *DCL1*, the dicer responsible for miRNA production. Importantly, *AtSN1* methylation is also reduced in both *nrpd1* and *nrpd2* mutants. The loss of *AtSN1* methylation in both siRNA pathway mutants and *nrpd* mutants suggests that Pol IV might also affect siRNAs. Consistent with this hypothesis, 5S gene and *AtSN1* siRNAs are significantly reduced or eliminated in *nrpd2* and *nrpd1* mutants (Figures 4D and 4E) as in *hen1*, *rdr2*, *drm*, or *ago4* mutants, confirming prior studies (Herr et al., 2005; Xie et al., 2004; Zilberman et al., 2004). By contrast, mutations of the RNA-dependent RNA polymerases *rdr1* or *rdr6* (*sgs2*, also known as *sde1*) had no effect, though *rdr6* is known to function in RNA silencing of transgenes (Baulcombe, 2004). Interestingly, 5S siRNA levels were actually increased in *ddm1* and *met1* mutants (Figure 4D), indicating that disrupted maintenance of cytosine methylation is not the explanation for loss of 5S siRNAs in *nrpd1* and *nrpd2* mutants.

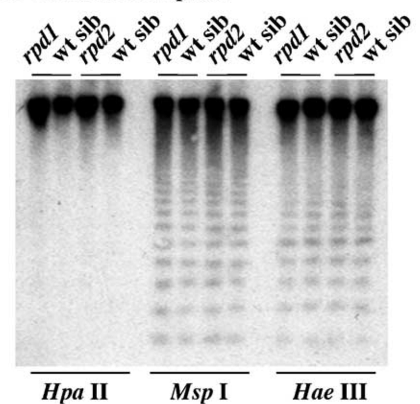
Importantly, miRNA levels are unaffected in *nrpd* mutants, as shown by comparison of miR163, 159, 164, 171, and 172 levels in mutant and wild-type siblings (Figure 4F), indicating that Pol IV acts only in the siRNA pathway and not in the miRNA pathway.

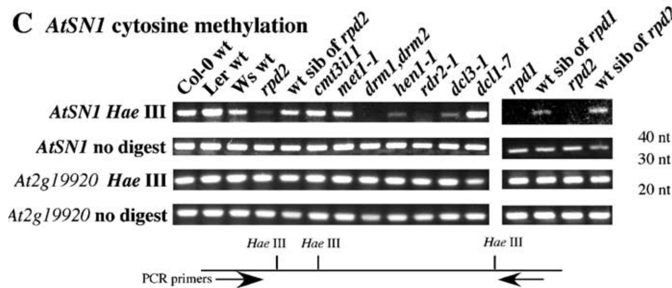
Discussion

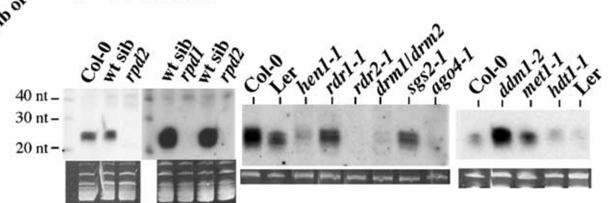
Loss of NRPD1 or NRPD2 function causes the loss of cytosine methylation at pericentromeric 5S genes and *AtSN1* retroelements yet has no discernible effect on centromere repeat methylation. These observations suggest that Pol IV primarily affects facultative heterochromatin rather than constitutive heterochromatin, consistent with the localization of NRPD2 at foci that overlap or are adjacent to chromocenters but are not fully coincident with chromocenters. We propose that Pol IV acts on genes that cycle between decondensed, euchromatic states and condensed, chromocenter-associated heterochromatic states, playing a key role in the amplification of siRNAs that direct cytosine methylation to these genes when they become activated (Aufsatz et al., 2002; Wassenegger, 2000).

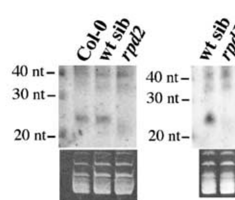
Interestingly, the total amount of H3^{dimethyl}K9, a reliable marker of heterochromatin, does not appear to be reduced in Pol IV mutant nuclei. Instead, the H3^{dimethyl}K9 is simply dispersed into a larger number of heterochromatic foci. Collectively, these data, combined with data showing disruption of chromocenters in *ddm1* and *met1* mutants (Soppe et al., 2002), suggest that loss of cytosine methylation from either pericentromeric repeats or centromeric repeats is sufficient to disrupt











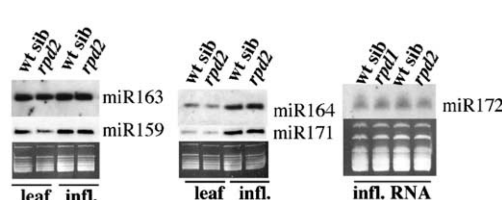


Figure 4. NRPD1 and NRPD2 Are Required for 5S Gene and *AtSN1* Cytosine Methylation and siRNA Accumulation

(A) Analysis of 5S gene repeats in *nrpd1a-3* and *nrpd2a-2 nrpd2b-1* double mutants relative to wild-type siblings and methylation mutants. Genomic DNA digested with HpaII, MspI, or HaeIII was hybridized to a 5S gene probe. *nrpd1*, *nrpd2*, *ddm1*, and *met1* mutants are in the Col-0 genetic background; *drm1drm2* and *cmt3* are in the WS background.

(B) Methylation of 180 bp centromere repeats is apparently unaffected in *nrpd1* and *nrpd2* mutants relative to wild-type siblings.

(C) *nrpd1* and *nrpd2* mutations cause decreased *AtSN1* cytosine methylation. PCR was used to amplify a portion of an *AtSN1* retroelement that includes three HaeIII sites. Undigested DNA and a gene lacking HaeIII sites served as PCR controls.

(D) 5S siRNAs in *nrpd1*, *nrpd2*, and mutants affecting siRNA production. Small RNA blots were probed for 5S siRNA sequences. Ethidium-stained gel bands serve as loading controls. The *hdt1* mutant is an ecotype Col-0 line with a T-DNA insertion in a nucleolar histone deacetylase; it serves as a T-DNA control in the blot at far right.

(E) *AtSN1* siRNAs are reduced or eliminated in *nrpd1* and *nrpd2* mutants.

(F) miRNAs 159, 163, 164, and 171 are unaffected in *nrpd1* and *nrpd2* mutants.

Arabidopsis pol IV subunit names are abbreviated from NRPD to RPD in this and all subsequent figures.

higher-order heterochromatin association into chromocenters. One possibility is that methylcytosine binding domain proteins and/or their associated proteins might act as linkers or bridges that help bring together dispersed heterochromatin domains.

At 5S genes, Pol IV affects cytosine methylation in all sequence contexts (CG, CNG, and CNN). Importantly, CG, CNG, and CNN de novo methylation is accomplished by DRM methyltransferase activity (Cao et al., 2003; Cao and Jacobsen, 2002). DRM is also responsible for siRNA-directed DNA methylation (in all sequence contexts) in *Arabidopsis* (Cao et al., 2003). We have shown that Pol IV and DRM activities are both needed for CNN methylation at *AtSN1* retroelements, as are genes of the siRNA pathway. These facts, combined with our demonstration that 5S and *AtSN1* siRNAs are essentially eliminated in Pol IV mutants, are most parsimonious with the hypothesis that Pol IV is involved in production of siRNAs that guide DRM-mediated cytosine methylation to repeated sequences complementary to the siRNAs (Chan et al., 2004). This would explain why loss of cytosine methylation in Pol IV mutants is most apparent at CNN (HaeIII in our experiments) sites, which would be dependent on continuous de novo

methylation due to the lack of a dedicated CNN maintenance methyltransferase (reviewed in Bender [2004]). By contrast, preexisting methylation at CG and CNG sites would be perpetuated by the MET1 and CMT3 maintenance methyltransferases, explaining the lesser effect of Pol IV or *drm* mutations on HpaII and MspI-sensitive 5S gene methylation (Figure 4A).

One could argue that DNA methylation is upstream of siRNA production, as suggested by the decrease in *AtSN1* siRNAs in *ddm1* and *met1* mutants (Lippman et al., 2003). However, this hypothesis does not fit with the fact that *ddm1* and *met1* cause dramatic decreases in 5S gene methylation yet actually increase 5S siRNA levels, possibly due to derepression of silenced 5S genes, thereby increasing the number of transcripts from which to generate dsRNAs and siRNAs. By contrast, Pol IV and *drm* mutations cause only modest decreases in total methylation yet essentially eliminate 5S siRNAs.

So how can loss of de novo methylation in a *drm* mutant eliminate siRNAs (Figure 4D) if siRNAs are upstream of de novo methylation? This apparent paradox might be explained if initial, primary siRNAs direct de novo methylation events that then trigger a massive amplification of siRNAs, and more extensive methylation, by a mechanism requiring Pol IV. Presumably, it is this second wave that yields the high levels of siRNAs and methylation that we detect. One possibility is that methylated DNA serves as the template for Pol IV-mediated transcription of aberrant RNAs. Another possibility is that methylation stalls elongating polymerases, as suggested by studies in *Neurospora* (Rountree and Selker, 1997), providing RDR2 with an opportunity to make dsRNAs from incomplete transcripts and leading to local production of aberrant RNAs or siRNAs that prime Pol IV transcription. Testing such hypotheses will be priorities for future studies.

Experimental Procedures

Plant Strains

Arabidopsis mutants *hen1-1*, *rdr2-1*, *dcl3-1*, and *dcl1-7* were provided by Jim Carrington. *met1-1* was provided by Eric Richards. *cmt3i11* was provided by Judith Bender. *sgs2-1* (alias *sde1*; *rdr6*) was provided by Herve Vaucheret. Salk T-DNA insertion lines and other mutants were obtained from the *Arabidopsis* Biological Resource Center (ABRC).

RNA and Immunoblot Analysis of NRPD2

RNA was isolated as described previously (Chen et al., 1998). RNA blots were hybridized to a probe generated by random priming of the *NRPD2a* 5′ RACE cDNA product using standard methods (Sambrook and Russell, 2001). For immunoblotting, plant tissue was homogenized in SDS sample buffer (125 mM Tris-HCl [pH 6.8], 2% SDS, 10% glycerol, and 0.7 M β-mercaptoethanol) and 40 μg of protein, determined using a BCA (bicinchoninic acid) protein assay kit (PIERCE), subjected to SDS-PAGE on a 7.5% gel, and electroblotted to a PVDF membrane. Anti-NRPD2 and anti-NRPB2 antisera were raised in rabbits against peptides DMDIDVKDLEEFEA and MEYNEYEPEEPQYVE of NRPD2a (At3g23780) and *A. thaliana* NRPB2 (At4g21710), respectively. Anti-Pol I+II+III rabbit antiserum was raised against peptide GDKFSSRHGQKG, which is conserved in Pol I, II, and III second-largest subunits. Sera were affinity purified using peptides covalently linked to NHS-activated Sepharose resin (Pharmacia Biotech). Columns were washed with 3–5 column volumes of PBS (pH 7.0), 0.05% Tween-20; antibodies were eluted using 0.1 M glycine-HCl (pH 3.0) neutralized by addition of Tris-HCl (pH 8.0) and stored at −80°C. Antisera were diluted 1:250 for probing immunoblots. The secondary antibody, diluted 1:5000, was peroxidase-linked donkey anti-rabbit IgG (Amersham). Immunoblots were visualized by chemiluminescence (ECL Western Blotting Detection kit; Amersham).

Screening of T-DNA Knockout Lines

T-DNA insertions in *NRPD2a*, *NRPD2b*, and *NRPD1a* were verified by PCR and sequencing using a T-DNA left border primer (5′-CGTCCGCAATGTGTTATTAAG-3′) and primers specific for *NRPD2a*, *NRPD2b*, or *NRPD1a* as suggested by the suppliers of the Salk lines. Screening by Southern blot analysis was according to standard methods (Sambrook and Russell, 2001).

Anion Chromatography and DNA-Dependent RNA Polymerase Assay

Arabidopsis plants were grown for 10 days at 25°C in 3 liter flasks containing 1 liter of liquid 1× Gamborg B5 medium, 1× Gamborg vitamins (Sigma), and 2% sucrose shaken at moderate speed. Tissue (200 g) was homogenized, and crude nuclear proteins were fractionated by DEAE-Sepharose chromatography and tested for RNA polymerase activity as described previously (Saez-Vasquez and Pikaard, 1997).

Phylogenetic Analyses

RNAP subunits were identified by blastp searches using *E. coli* RPOC and RPOB, *S. cerevisiae* RPB1 and RPB2, and *A. thaliana* NRPD1a and NRPD2a protein sequences. Sequences were aligned, using Clustal X (version 1.81). Conserved sequences were highlighted using BOXSHADE. (http://bioweb.pasteur.fr/seqanal/interfaces/boxshade.html). Phylogenetic analysis was by the neighbor-joining method, with 1000 bootstrap replications, using PAUP (version 4.0b10).

Cytosine Methylation Assays

Genomic DNA (100 ng) was digested with HpaII, MspI, or HaeIII. Following agarose gel electrophoresis, DNA was blotted to uncharged nylon membranes. Probes were generated by random priming, and blots were hybridized using standard methods (Sambrook and Russell, 2001).

AtSN1 methylation assays used ~100 ng of DNA digested with HaeIII (or undigested for controls). Approximately 5% of digestion reaction DNA was then used for each PCR reaction. PCR conditions were 2 min at 94°C, followed by 35 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 30 s. Primer sequences for *AtSN1* were the following: 5′-ACTTAATTAGCACTCAAATTAAACAAAATAAGT-3′ and 5′-TTTAAACATAAGAAGAAGTTCCTTTTTCATCTAC-3′. The *At2g19920* control was amplified using 5′-TCACCCGAACAGTTGGAAGAA GAG-3′ and 5′-GTGAGGAACCGGTCCATTATTGCT-3′. PCR products were subjected to agarose gel electrophoresis.

In Situ Hybridization and Immunolocalization

Emerging leaves of 21-day-old plants were fixed in ethanol:acetic acid (3:1, v/v). Nuclei were prepared as described (Schwarzacher and Mosgoeller, 2000). FISH using biotin-dUTP or digoxygenin-dUTP labeled 180 bp *A. thaliana* pericentromeric repeat, 5S gene or 45S rRNA gene intergenic spacer sequence probes was as described previously (Pontes et al., 2004).

For immunolocalization experiments, nuclei were fixed in 4% paraformaldehyde. H3^{dimethyl}K9 was localized using published methods (Houben et al., 1996) with antibody purchased from Upstate Biotechnology. For NRPD2, slides were permeabilized with 10% DMSO, 3% NP-40 in PBS, before blocking with 1% BSA in PBS. Primary antibodies were diluted 1:100 in PBS, 1% BSA, and slides were incubated overnight at 4°C. Secondary antibodies were conjugated to rhodamine or fluorescein (Sigma). Chromatin was counterstained with DAPI in antifade buffer (Vector Laboratories). Nuclei were examined using a Nikon Eclipse E600 epifluorescence microscope and images collected using a Q-Imaging Retiga EX digital camera.

siRNA and miRNA Detection

RNA was isolated using the mirVana miRNA isolation kit (Ambion). RNA (2–6 μg) was resolved by denaturing polyacrylamide gel electrophoresis on a 20% (w/v) gel. Gels were electroblotted (20 mA/cm^2 for 2 hr) to Magnacharge nylon membranes (0.22 μm; Osmonics) using a semidry transfer apparatus. An end-labeled RNA ladder was used as a molecular weight marker (Decade Marker System, Ambion). The AtSN1 riboprobe was synthesized from a NdeI-linearized plasmid DNA template (Zilberman et al., 2003). All other riboprobes were generated according to the mirVana probe construction kit (Ambion) using oligonucleotides specific for a given small RNA and labeling by T7 polymerase transcription in the presence of α-^{32}P CTP. DNA oligonucleotides for 5S and miRNA probes were the following: siR1003T7 (5S) (5′-AGACCGTGAGGCCAAACTTGG CATcctgtctc-3′; small letters are complementary to the T7 promoter oligonucleotide), miR159T7 (5′-TTTGGATTGAAGGGAGCTC TAcctgtctc-3′), miR163T7 (5′-TTGAAGAGGACTTGGAACTTCGAT cctgtctc-3′), and miR164T7 (5′-TGGAGAAGCAGGGCACGTGCA cctgtctc-3′). Unincorporated nucleotides were removed using Performa DTR Gel Filtration Cartridges (Edgebiosystems). Blot hybridization was in 50% formamide, 0.25 M Na_2HPO_4 (pH 7.2), 0.25 M NaCl, 7% SDS at 42°C (14–16 hr) followed by two 15 min washes at 37°C in 2× SSC, two 15 min washes at 37°C in 2× SSC, 0.1% SDS, and a 10 min wash in 0.5× SSC, 1% SDS.

Supplemental Data

Supplemental Data include five figures, three tables, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at http://www.cell.com/cgi/content/full/120/5/613/DC1/.

Acknowledgments

We are indebted to David Baulcombe for sharing his lab's unpublished evidence that *sde4* is *nrpd1a*. We thank the Carrington, Richards, Vaucheret, and Bender labs for mutant seed stocks; the Carrington lab for cloned *AtSN1*; Thomas Guilfoyle for the anti-RPB5 antibody; M.E. Gifford for phylogenetic analysis advice; Douglas Chalker for use of his fluorescence microscope; and our labmate Keith Earley for RNAP assay advice. We thank Dr. Wanda Viegas for comentoring P.C.N. This research was supported by NIH grant R01-GM60380 and USDA grant 99-35301-7865 (to C.S.P.). P.C.N. was supported by a predoctoral fellowship from the Portuguese Fundação para a Ciência e Tecnologia (SFRH/BD/6520/2001) and the Luso-American Foundation (grant 237/2004).

Received: December 23, 2004
Revised: January 24, 2005
Accepted: February 4, 2005
Published online: February 10, 2005

References

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature *408*, 796–815.

Archambault, J., and Friesen, J.D. (1993). Genetics of eukaryotic RNA polymerases I, II, and III. Microbiol. Rev. *57*, 703–724.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002). RNA-directed DNA methylation in *Arabidopsis*. Proc. Natl. Acad. Sci. USA Suppl. *99*, 16499–16506.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*, 356–363.

Bender, J. (2004). Chromatin-based silencing mechanisms. Curr. Opin. Plant Biol. *7*, 521–526.

Campell, B.R., Song, Y., Posch, T.E., Cullis, C.A., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. Gene *112*, 225–228.

Cao, X., and Jacobsen, S.E. (2002). Role of the *Arabidopsis* DRM methyltransferases in de novo DNA methylation and gene silencing. Curr. Biol. *12*, 1138–1144.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr. Biol. *13*, 2212–2217.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chen, Z.J., Comai, L., and Pikaard, C.S. (1998). Gene dosage and stochastic effects determine the severity and direction of uniparental rRNA gene silencing (nucleolar dominance) in *Arabidopsis* allopolyploids. Proc. Natl. Acad. Sci. USA *95*, 14891–14896.

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. Science *288*, 640–649.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *292*, 1863–1876.

Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. Proc. Natl. Acad. Sci. USA *99*, 14584–14589.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes Dev. *17*, 1691–1702.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. *21*, 4671–4679.

Herr, A., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science, in press. Published online February 3, 2005. 10.1126/science.1106910.

Houben, A., Belyaev, N.D., Turner, B.M., and Schubert, I. (1996). Differential immunostaining of plant chromosomes by antibodies recognizing acetylated histone H4 variants. Chromosome Res. *4*, 191–194.

Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999). *Arabidopsis thaliana* RNA polymerase II subunits related to yeast and human RPB5. Gene *231*, 41–47.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. *1(3)* e67 DOI: 10.1371/journal.pbio.0000067.

McClelland, M., Nelson, M., and Raschke, E. (1994). Effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases. Nucleic Acids Res. *22*, 3640–3659.

Pontes, O., Neves, N., Silva, M., Lewis, M.S., Madlung, A., Comai, L., Viegas, W., and Pikaard, C.S. (2004). Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis* suecica genome. Proc. Natl. Acad. Sci. USA *101*, 18240–18245.

Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489–500.

Rountree, M.R., and Selker, E.U. (1997). DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. Genes Dev. *11*, 2383–2395.

Saez-Vasquez, J., and Pikaard, C.S. (1997). Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. Proc. Natl. Acad. Sci. USA *94*, 11869–11874.

Saez-Vasquez, J., and Pikaard, C.S. (2000). RNA polymerase I holoenzyme-promoter interactions. J. Biol. Chem. *275*, 37173–37180.

Sambrook, J., and Russell, D.R. (2001). Molecular Cloning: A Laboratory Manual, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev. *16*, 2593–2620.

Schwartz, L.B., and Roeder, R.G. (1974). Purification and subunit structure of deoxyribonucleic acid-dependent ribonucleic acid polymerase I from mouse myeloma, MOPC 315. J. Biol. Chem. *249*, 5898–5906.

Schwarzacher, H.G., and Mosgoeller, W. (2000). Ribosome biogenesis in man: current views on nucleolar structures and function. Cytogenet. Cell Genet. *91*, 243–252.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. EMBO J. *21*, 6549–6559.

Wassenegger, M. (2000). RNA-directed DNA methylation. Plant Mol. Biol. *43*, 203–220.

Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453–463.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kassachau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. *2(5):*, e104 DOI: 10.1371/journal.pbio.0020104.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell *98*, 811–824.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr. Biol. *14*, 1214–1220.

Accession Numbers

The GenBank accession number for the NRPD2a mRNA sequence determined for this paper is AY862891.

Note Added in Proof

In the early online version of the article, the genes NRPD1a, NRPD1b, NRPD2a, and NRPD2b were named RPD1a, RPD1b, RPD1a, and RPD2b, respectively. We have changed the names due to a nomenclature conflict.

Supplemental Data

Plant Nuclear RNA Polymerase IV Mediates siRNA and DNA Methylation-Dependent Heterochromatin Formation

Yasuyuki Onodera, Jeremy R. Haag,
Thomas Ream, Pedro Costa Nunes,
Olga Pontes, and Craig S. Pikaard

I. Phylogenetic Analyses

Species whose subunit sequences are included in the unrooted trees of Figure 1 are the following: Ac, *Adiantum capillus-veneris*; ACNPV, *Autographa californica nucleopolyhedrovirus*; Af, *Anthoceros formosae*; Agt, *Agrobacterium tumefaciens*; An, *Aspergillus nidulans*; Ap, *Aquifex pyrophilus*; Arf, *Archaeoglobus fulgidus*; ASFV, *African swine fever virus*; At, *Arabidopsis thaliana*; Av, *Anabaena variabilis*; Ba, *Bacillus anthracis*; Bb, *Borrelia burgdorferi*; Bj, *Bradyrhizobium japonicum*; Bs, *Bacillus subtilis*; Cc, *Cyanidium caldarium*; Ce, *Caenorhabditis elegans*; Cp, *Cyanophora paradoxa*; CPV, *Cowpox virus*; Cv, *Chlorella vulgaris*; Dm, *Drosophila melanogaster*; Ec, *Escherichia coli*; Eg, *Euglena gracilis*; EV, *Ectromelia virus*; FPV, *Fowlpox virus*; Gt, *Guillardia theta*; H, *Halobacterium salinarum*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Hs, *Homo sapiens*; Le, *Lycopersicon esculentum*; Lp, *Legionella pneumophila*; MCV, *Molluscum contagiosum virus*; Mel, *Mesorhizobium loti*; Mes, *Mesostigma viride*; Met, *Methanothermobacter thermautotrophicus*; Mev, *Methanococcus vannielii*; Mg, *Mycoplasma genitalium*; Mga, *Mycoplasma gallisepticum*; Mj, *Methanocaldococcus jannaschii*; Ml, *Mycobacterium leprae*; Mm, *Mus musculus*; Mp, *Marchantia polymorpha*; MPV, *Monkeypox virus*; Mt, *Mycobacterium tuberculosis*; MV, *Myxoma virus*; Nc, *Neurospora crassa*; Nca, *Neospora caninum*; Nm, *Neisseria meningitidis*; No, *Nephroselmis olivacea*; Np, *Nostoc punctiforme*; Nt, *Nicotiana tabacum*; Os, *Oryza sativa*; OV, *Orf virus*; Pa, *Pseudomonas aeruginosa*; Pf, *Plasmodium falciparum*; Po, *Porphyra purpurea*; Pp, *Physcomitrella patens*; Ps, *Pseudomonas syringae*; Py, *Pyrococcus abyssi*; Pyh, *Pyrococcus horikoshii*; RFV, *Rabbit fibroma virus*; Rp, *Rickettsia prowazekii*; RPV, *Rabbitpox virus*; Rt, *Rickettsia typhi*; S6803, *Synechocystis sp. PCC 6803*; Sa, *Staphylococcus aureus*; Sc, *Saccharomyces cerevisiae*; Se, *Salmonella enterica*; Sia, *Sinapis alba*; So, *Spinacia oleracea*; Sp, *Schizosaccharomyces pombe*; SPPV, *Sheeppox virus*; SPV, *Swinepox virus*; Su, *Sulfolobus acidocaldarius*; Ta, *Thermoplasma acidophilum*; Tc, *Thermococcus celer*; Tg, *Toxoplasma gondii*; Tv, *Thermoplasma volcanium*; Vc, *Vibrio cholerae*; VMV, *Variola major virus*; VV, *Vaccinia virus*; Xl, *Xenopus laevis*; YMTV, *Yaba monkey tumor virus*; Yp, *Yersinia pestis*; Zm, *Zea mays*.

Additional Methods for Phylogenetic Analyses

Second-largest subunits in some of the archaea and largest subunits in archaea and chloroplasts display a split domain architecture (Bergsland and Haselkorn, 1991; Puhler et al., 1989; Schneider and Hasekorn, 1988). In these cases, sequences were joined and aligned in Clustal X (version 1.81) to fit the domain architecture of *E. coli* and *S. cerevisiae* protein sequences in order to facilitate phylogenetic comparisons. The annotated sequence for At2g40030 (*RPD1b*) present in Genbank lacks conserved C-terminal domains G and H, and was not studied functionally due to the presumption that it would be non-functional. However, our own analysis of the genomic sequence using TWINSCAN (http://www.genes.cs.wustl.edu) revealed part of domain G in what is currently annotated as an intergenic region and the remainder of the predicted protein can be found in a predicted neighboring gene, *At2g40040*, suggesting that the existing annotation is incorrect. We used our own annotation for *A. thaliana* RPD1b in the phylogenetic analysis shown in Figure 1. The annotated sequence for *O. sativa* RPD1a (CAD41657) also appeared to be

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

inaccurate in parts after alignment, so the genomic sequence was analyzed using FGENESH+ (www.softberry.com) with *O. sativa* RPD1b as a reference sequence in order to perform gene finding with similarity. The sequences were aligned and a final prediction for *O. sativa* RPD1a was used in the phylogenetic analyses.

Arabidopsis RPD1a is 30% identical (42% similar) to rice OsCAD41657, but only 14% identical (23% similar) to Arabidopsis *RPD1b*. The higher similarity among orthologs between species than among paralogs within a species indicates that two *RPD1* genes existed prior to the divergence of monocots and dicots ~200 million years ago (Wolfe et al., 1989).

The Arabidopsis RPD2a protein is 84% identical to the predicted Arabidopsis RPD2b open reading frame and 55% identical to rice OsAK121416.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Tables S1 and S2. GenBank Accessions for the DNA-Dependent RNA Polymerase Largest Subunits Analyzed in Figure 1

Supplemental Table 1 - RNAP Largest Subunit Sequences

Category	Genbank Accession	Abbreviation	Organism	Gene/Locus	Protein
Pol IV	NM_104980	At1g63020	Arabidopsis thaliana	At1g63020	RPD1a
	NM_129561	At2g40030	Arabidopsis thaliana	At2g40030	RPD1b
	XP_473570	OsXP473570	Oryza sativa	CAD41657	RPD1a
	NP_914279	OsNP914279	Oryza sativa	AP004365	RPD1b
Pol I	NM_115626	AtRpaI	Arabidopsis thaliana	At3g57660	RPA1
	J03530	ScRpaI	Saccharomyces cerevisiae	YSCPOLAI	
	NM_079019	DmRpaI	Drosophila melanogaster		
	AAC99959	HsRpaI	Homo sapiens		
	NP_496872	CeRpaI	Caenorhabditis elegans		
	NP_496872	OsRpaI	Oryza sativa		
	JS0080	SpRpaI	Schizosaccharomyces pombe		
Pol II	NM_119746	AtRpb1	Arabidopsis thaliana	At4g35800	RPB1
	NM_078569	DmRpb1	Drosophila melanogaster		
	X03128	ScRpb1	Saccharomyces cerevisiae	SCRPO21	
	CAA45125	HsRpb1	Homo sapiens		
	NP_500523	CeRpb1	Caenorhabditis elegans		
	AAQ08515	ZmRpb1	Zea mays		
	XP_493925	OsRpb1	Oryza sativa		
	NP_595673	SpRpb1	Schizosaccharomyces pombe		
Pol III	NP_595673	AtRpc1	Arabidopsis thaliana	At5g60040	RPC1
	X03129	ScRpc1	Saccharomyces cerevisiae	SCRPO31	
	AF021351	HsRpc1	Homo sapiens		
	NM_132843	DmRpc1	Drosophila melanogaster		
	NP_501127	CeRpc1	Caenorhabditis elegans		
	NP_501127	OsRpc1	Oryza sativa		
	O94666	SpRpc1	Schizosaccharomyces pombe		
Eubacteria	AAC43086	EcRpoC	Escherichia coli K12	rpoC	RPOC
	NP_457916	SeRpoC	Salmonella enterica		
	NP_252959	PaRpoC	Pseudomonas aeruginosa		
	YP_026389	BaRpoC	Bacillus anthracis		
	NP_215182	MtRpoC	Mycobacterium tuberculosis		
	NP_073010	MgRpoC	Mycoplasma genitalium		
	NP_438672	HiRpoC	Haemophilus influenzae		
	NP_220532	RpRpoC	Rickettsia prowazekii		
	CAA61517	SaRpoC	Staphylococcus aureus		
	CAA52958	ApRpoC	Aquifex pyrophilus		
	NP_994402	YpRpoC	Yersinia pestis biovar Medievalis str. 91001		
	NP_229983	VcRpoC	Vibrio cholerae O1 biovar eltor str. N16961		
	ZP_00123798	PsRpoC	Pseudomonas syringae pv. syringae B728a		
	YP_094367	LpRpoC	Legionella pneumophila subsp. pneumophila str. Philadelphia 1		
	NP_282991	NmRpoC	Neisseria meningitidis Z24		
	NP_102111	MelRpoC	Mesorhizobium loti MAFF303099		
	NP_772049	BjRpoC	Bradyrhizobium japonicum USDA 110		
	NP_354930	AgtRpoC	Agrobacterium tumefaciens str. C58		
	YP_067097	RtRpoC	Rickettsia typhi str. Wilmington		
Archaea	CAA47723	Tc	Thermococcus celer	rpoA1	
	CAA47724	Tc	Thermococcus celer	rpoA2	
	NP_126306	Py	Pyrococcus abyssi	rpoA1	
	NP_126307	Py	Pyrococcus abyssi	rpoA2	
	NP_248036	Mj	Methanocaldococcus jannaschii DSM 2661	rpoA1	
	NP_248037	Mj	Methanocaldococcus jannaschii DSM 2661	rpoA2	
	NP_444249	H	Halobacterium	rpoA1	
	P15354	H	Halobacterium	rpoA2	
	NP_148215	Ae	Aeropyrum pernix	rpoA1	
	NP_148214	Ae	Aeropyrum pernix	rpoA2	
	NP_070713	Af	Archaeoglobus fulgidus	rpoA1	
	NP_070714	Af	Archaeoglobus fulgidus	rpoA2	
	CAA48281	Ta	Thermoplasma acidophilum	rpoA1	
	CAA48282	Ta	Thermoplasma acidophilum	rpoA2	
	P11512	Su	Sulfolobus acidocaldarius	rpoA1	
	P11514	Su	Sulfolobus acidocaldarius	rpoA2	
Virus	NP_044030	MCV	Molluscum contagiosum virus		
	O57204	VV	Vaccinia virus		
	AAF14956	MV	Myxoma virus		
	AAF17950	RFV	Rabbit fibroma virus		
	AAR07427	YMTV	Yaba monkey tumor virus		
	T28521	VMV	Variola major virus		
	CAD90647	CPV	Cowpox		
	AAL69807	SPV	Swinepox virus		
	NP_659643	SPPV	Sheeppox virus		
	AAL40548	MPV	Monkeypox virus		
	AAM92386	EV	Ectromelia virus		
	NP_957833	OV	Orf virus		

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Category	Genbank Accession	Abbreviation	Organism	Gene/Locus	Protein
Cyanobacteria & Chloroplast	AAL40548	AtCPST	A. thaliana (CPST)	rpoC1	
	NP_051049	AtCPST	A. thaliana (CPST)	rpoC2	
	CAA60277	ZmCPST	Zea mays (CPST)	rpoC1	
	CAA60278	ZmCPST	Zea mays (CPST)	rpoC2	
	BAC55418	AfCPST	Anthoceros formosae (hornwort) CPST	rpoC1	
Category	**Genbank Accession**	**Abbreviation**	**Organism**	**Gene/Locus**	**Protein**
Cyanobacteria & Chloroplast	NP_904221	PpCPST	Physcomitrella patens (moss) CPST	rpoC2	
	P06273	MpCPST	Marchantia polymorpha (liverwort) CPST	rpoC1	
	NP_039277	MpCPST	Marchantia polymorpha (liverwort) CPST	rpoC2	
	AAP29383	AcCPST	Adiantum capillus-veneris (fern)CPST	rpoC1	
	NP_848050	AcCPST	Adiantum capillus-veneris (fern)CPST	rpoC2	
	CAA77411	NtCPST	Nicotiana tabacum (CPST)	rpoC1	
	NP_054486	NtCPST	Nicotiana tabacum (CPST)	rpoC2	
	P11705	SoCPST	Spinacia oleracea (CPST)	rpoC1	
	NP_054922	SoCPST	Spinacia oleracea (CPST)	rpoC2	
	AAC08137	PoCPST	Porphyra purpurea chloroplast (red algae)	rpoC1	
	NP_053860	PoCPST	Porphyra purpurea chloroplast (red algae)	rpoC2	
	P56300	CvCPST	Chlorella vulgaris chloroplast	rpoC1	
	NP_045895	CvCPST	Chlorella vulgaris chloroplast	rpoC2	
	NP_045032	CcCPST	Cyanidium caldarium (CPST)	rpoC1	
	NP_045033	CcCPST	Cyanidium caldarium (CPST)	rpoC2	
	ZP_00160830	Av	Anabaena variabilis ATCC 29413	rpoC1	
	ZP_00160831	Av	Anabaena variabilis ATCC 29413	rpoC2	
	ZP_00111112	Np	Nostoc punctiforme PCC 73102	rpoC1	
	ZP_00111113	Np	Nostoc punctiforme PCC 73102	rpoC2	
	VIMSS11977	S6803	Synechocystis sp. PCC 6803	rpoC1	
	NP_440684	S6803	Synechocystis sp. PCC 6803	rpoC2	
	NP_039374	OsCPST	Oryza sativa (CPST)	rpoC1	
	NP_039375	OsCPST	Oryza sativa (CPST)	rpoC2	

Supplemental Table 2 - RNAP 2nd Largest Subunit Sequences

Category	Genbank Accession	Abbreviation	Organism	Gene/Locus	Protein
Pol IV	NM_113282	At3g23780	Arabidopsis thaliana	At3g23780	RPD2a
	NM_112691	At3g18090	Arabidopsis thaliana	At3g18090	RPD2b
	AK121416	OsAK121416	Oryza sativa	AK121416	RPD2-like
	XM_480298	OsXM480298	Oryza sativa	XM_480298	RPD2-like
Pol I	NM_102734	AtRpa2	Arabidopsis thaliana	At1g29940	RPA2
	M62804	ScRpa2	Saccharomyces cerevisiae	YSCRPA135	RPA135
	AAF51503	DmRpa2	Drosophila melanogaster		
	Q9H9Y6	HsRpa2	Homo sapiens		RPA135
	NP_595819	SpRpa2	Schizosaccharomyces pombe		RPA2
	XP_329740	NcRpa2	Neurospora crassa		
	EAA59242	AnRpa2	Aspergillus nidulans		
	AAH60656	MmRpa2	Mus musculus		
	AAH59304	XlRpa2	Xenopus laevis		
	NP_492476	CeRpa2	Caenorhabditis elegans		
	NP_922143	OsRpa2	Oryza sativa		
Pol II	NM_118291	AtRpb2	Arabidopsis thaliana	At4g21710	RPB2
	P08266	DmRpb2	Drosophila melanogaster		
	P08518	ScRpb2	Saccharomyces cerevisiae		
	AAH23503	HsRpb2	Homo sapiens		
	Q10578	CeRpb2	Caenorhabditis elegans		
	S35548	SpRpb2	Schizosaccharomyces pombe		
	XP_324477	NcRpb2	Neurospora crassa		
	S65068	LeRpb2	Lycopersicon esculentum		
	EAA61953	AnRpb2	Aspergillus nidulans		
	NP_722493	MmRpb2	Mus musculus		
Pol III	NM_123882	AtRpc2	Arabidopsis thaliana	At5g45140	RPC2
	AAB59324	ScRpc2	Saccharomyces cerevisiae		
	CAA35185	DmRpc2	Drosophila melanogaster		
	AAM18214	HsRpc2	Homo sapiens		
	NP_593690	SpRpc2	Schizosaccharomyces pombe		
	EAA65727	AnRpc2	Aspergillus nidulans		
	XP_328211	NcRpc2	Neurospora crassa		
	NP_081699	MmRpc2	Mus musculus		
	NP_498192	CeRpc2	Caenorhabditis elegans		
	XP_470900	OsRpc2	Oryza sativa		
Eubacteria	NC_000913	EcRpoB	Escherichia coli K12	rpoB	RPOB
	NP_807130	SeRpoB	Salmonella enterica	rpoB	
	NP_252960	PaRpoB	Pseudomonas aeruginosa	rpoB	
	YP_052605	BaRpoB	Bacillus anthracis	rpoB	
	NP_302273	MlRpoB	Mycobacterium leprae	rpoB	
	AAP56563	MgaRpoB	Mycoplasma gallisepticum	rpoB	
	NP_438673	HiRpoB	Haemophilus influenzae	rpoB	
	AAC69338	LpRpoB	Legionella pneumophila	rpoB	
	NP_207989	HpRpoB	Helicobacter pylori 26695	rpoB	
	NP_387988	BsRpoB	Bacillus subtilis	rpoB	
	NP_645314	SaRpoB	Staphylococcus aureus	rpoB	
	YP_067096	RtRpoB	Rickettsia typhi str. Wilmington	rpoB	
	NP_212523	BbRpoB	Borrelia burgdorferi B31	rpoB	

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Archaea	CAA32924	Su	Sulfolobus acidocaldarius	rpoB	
	CAA47722	Tc	Thermococcus celer	rpoB	
	NP_248034	Mj	Methanocaldococcus jannaschii DSM 2661	rpoB2	
	NP_248035	Mj	Methanocaldococcus jannaschii DSM 2661	rpoB1	
	NP_281214	H	Halobacterium	rpoB2	
	NP_281213	H	Halobacterium	rpoB1	
	NP_148216	Ae	Aeropyrum pernix K1	rpoB	
	NP_126305	Py	Pyrococcus abyssi GE5	rpoB	
	NP_143407	Pyh	Pyrococcus horikoshii OT3	rpoB	
	CAA51726	Mev	Methanococcus vannielii	rpoB2	
	CAA51727	Mev	Methanococcus vannielii	rpoB1	
	NP_070711	Arf	Archaeoglobus fulgidus DSM 4304	rpoB2	
	NP_070712	Arf	Archaeoglobus fulgidus DSM 4304	rpoB1	
	NP_276179	Met	Methanothermobacter thermautotrophicus str. Delta H"	rpoB2	
	NP_276180	Met	Methanothermobacter thermautotrophicus str. Delta H"	rpoB1	
	NP_111701	Tv	Thermoplasma volcanium GSS1	rpoB	
	NP_393870	Ta	Thermoplasma acidophilum DSM 1728	rpoB	
Viruses	AAC55257	MCV	Molluscum contagiosum virus		
	AAO89423	VV	Vaccinia virus		
	AAF15002	MV	Myxoma virus		
	AAF17997	RFV	Rabbit fibroma virus		
	AAR07472	YMTV	Yaba monkey tumor virus		
	T28566	VMV	Variola major virus		
	AAM13599	CPV	Cowpox virus		
	S78061	ASFV	African swine fever virus		
	AAL40593	MPV	Monkeypox virus		
	YP_006777	RPV	Rabbitpox virus		
	CAE52727	FPV	Fowlpox virus (isolate HP-438[Munich])		
	AAA66680	ACNPV	Autographa californica nucleopolyhedrovirus		
Cyanobacteria & Chloroplast	BAA84377	AtCPST	Arabidopsis thaliana (CPST)	rpoB	RPOB
	Q9TL06	NoCPST	Nephroselmis olivacea (CPST)	rpoB	
	P11703	SoCPST	Spinacia oleracea (CPST)	rpoB	
	P06271	NtCPST	Nicotiana tabacum (CPST)	rpoB	
	P46818	SiaCPST	Sinapis alba (CPST)	rpoB	
	CAA60276	ZmCPST	Zea mays (CPST)	rpoB	
	NP_039373	OsCPST	Oryza sativa (CPST)	rpoB	
	RNLVB	MpCPST	Marchantia polymorpha (liverwort) CPST	rpoB	
	Q9MUS5	MesCPST	Mesostigma viride (CPST)	rpoB	
	BAA57969	CvCPST	Chlorella vulgaris (green algae) CPST	rpoB	
	CAA50138	EgCPST	Euglena gracilis (CPST)	rpoB	
	AAC35676	GtCPST	Guillardia theta (CPST)	rpoB	
	AAC08138	PoCPST	Porphyra purpurea (CPST)	rpoB	
	NP_045031	CcCPST	Cyanidium caldarium (CPST)	rpoB	
	NP_043230	CpPST	Cyanophora paradoxa (PST)	rpoB	
	AAD17842	TgPST	Toxoplasma gondii (PST)	rpoB	
	AAF14261	NcaPST	Neospora caninum (PST)	rpoB	
	NP_440685	S6803	Synechocystis sp. PCC 6803	rpoB	
Other	NP_701431	Pf	Plasmodium falciparum 3D7		

RNA polymerase subunits are categorized according to clade designations.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

II. Protein Alignments

Supplemental Figure 1. Multiple Alignment of RPD1 with DNA-Dependent RNA Polymerase Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 1. Alignment of RNAP Largest Subunits

```
At_RPD1    -------------MEDDCEELQVPVGTLTSIGFSISNNNDRDKMSVLEV------------
At_RPB1    ------------MDTRFPFSPAEVSKVRVVQFGILSPDEIRQMSVIHVEHSETTEK--GK
At_RPC1    ----METKMEIEFTKKPYIEDVGPLKIKSINFSVLSDLEVMKAAEVQVWNIGLYDHS-FK
At_RPA1    MAHAQTTEVCLSFHRSLLFPMGASQVVESVRFSFMTEQDVRKHSFLKVTSPILHDNV-GN
Sc_RPB1    -------------MVGQQYSSAPLRTVKEVQFGLFSPEEVRAISVAKIRFPETMDETQTR
Ec_RPOC    ----------MKDLLKFLKAQTKTEEFDAIKIALASPDMIRSWSFGEVKKPETINYRTFK
consensus             -     K   F   A V   VKSIQFSILSPDEVRKMSVL  V    PET D     K
           Conserved domain A
At_RPD1    EAPNQVTDSRLGLPNPDSVCRTCGSKDRKVCEGHFGVINFAYSIINPYFLKEVAALLNKI
At_RPB1    PKVGGLSDTRLGTIDRKVKCETCMAN-MAECPGHFGYLELAKPMYHVGFMKTVLSIMRCV
At_RPC1    PYENGLLDPRMGPPNKKSICTTCEGN-FQNCPGHYGYLKLDLPVYNVGYFNFILDILKCI
At_RPA1    PFPGGLYDLKLGPKDDKQACNSCGQL-KLACPGHCGHIELVFPIYHPLLFNLLFNFLQRA
Sc_RPB1    AKIGGLNDPRLGSIDRNLKCQTCQEG-MNECPGHFGHIDLAKPVFHVGFIAKIKKVCECV
Ec_RPOC    PERDGLFCARIFGPVKDYECLCGKYK-RLK---HRGVICEKCGVEVTQTKVRRERMGHIE
consensus  P   GGL  D  RLG  PDKK    C  TC      R    CPGHFG  IELA  PVYHVGFI    I    IL  CI

At_RPD1    CPGCKYIRKKQFQITEDQPERCRYCT----------------LNTGYPLMKFRVTTKEVF
At_RPB1    CFNCSKILADEVCRSLFRQAMKIK-------------NPKNRLKKILDACKNKTKCDGGD
At_RPC1    CKRCSNMLLDEKLYEDHLRKMRNPRM---------EPLKKTELAKAVVKKCSTMASQRII
At_RPA1    CFFCHHFMAKPEDVERAVSQLKLIIKGDIVSAKQLESNTPTKSKSSDESCESVVTTDSSE
Sc_RPB1    CMHCGKLLLDEHN-ELMRQALAIKDS-------------KKRFAAIWTLCKTKMVCET--
Ec_RPOC    LASPTAHIWFLKS-LPSRIGLLLDMP-----------LRDIERVLYFESYVVIEGGMTNL
consensus  C   CS  IL  DE     E    R  ALKI                    K  RL     LE  CKSKM  TDE

At_RPD1    RRSGIVVEVNEESLMKLKKRGVLTLP----------------------------------
At_RPB1    DIDDVQSHSTDEPVKKSRGGCGAQQPKLTIEG----------------------------
At_RPC1    TCKKCGYLNGMVKKIAAQFGIGISHDRSKIHG----------------------------
At_RPA1    ECEDSDVEDQRWTSLQFAEVTAVLKNFMRLSSKSCSRCKGINPKLEKPMFGWVRMRAMKD
Sc_RPB1    ---DVPSE-DDPTQLVSRGGCGNTQPTIRKDG----------------------------
Ec_RPOC    ERQQILTEEQYLDALEEFGDEFD-------------------------------------
consensus  E    DI  SE  QD  T  L     RGG  GIT  P  IKI  G

At_RPD1    -----------------------------------PDYWSFLPQDSNIDESCLKPTRRII
At_RPB1    ----------------------------------MKMIAEYKIQRKKNDEPDQLPEPAER
At_RPC1    ----------------------------------GEIDECKSAISHTKQST---AAINPL
At_RPA1    SDVGANVIRGLKLKKSTSSVENPDGFDDSGIDALSEVEDGDKETREKSTEVAAEFEEHNS
Sc_RPB1    ----------------------------------LKLVGSWKKDRATGDAD----EPELR
Ec_RPOC    ------------------------------------AKMGAEAIQALLKSMDLEQECEQL
consensus                                           I     FK    R      DE         E     L
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
                                                        Conserved domain B
At_RPD1    THAQVYALLLGIDQRLIKKDIP----------------MFNSLGLTSFPVTPNGYRVTEI
At_RPB1    KQTLGADRVLSVLKRISDADCQLLGFNPKFA-------RPDWMILEVLPIPPPPVRPSVM
At_RPC1    TYVLDPNLVLGLFKRMSDKDCELL---YIAY-------RPENLIITCMLVPPLSIRPSVM
At_RPA1    KRDLLPSEVRNILKHLWQNEHEFCSFIGDLWQSGSEKIDYSMFFLESVLVPPTKFRPPTT
Sc_RPB1    --VLSTEEILNIFKHISVKDFTSLGFNEVFS-------RPEWMILTCLPVPPPPVRPSIS
Ec_RPOC    REELNETNSETKRKKLTKRIKLLEAFVQSGN-------KPEWMILTVLPVLPPDLRPLVP
consensus  K  L    VL I KRLS KD  LLGF            RPEWMILT LPVPPP VRPSVM

At_RPD1    VHQFNGARLIFDERTRIYKKLVGFEGNTLELSSRVMECMQYSRLFSETVSSSKDS-----
At_RPB1    MDATSRSEDDLTHQLAMIIRHNENLKRQEKNGAPAHIISEFTQLLQFHIATYFDNELPGQ
At_RPC1    IGGIQSNENDLTARLKQIILGNASLHKILSQPTSSPKNMQVWDTVQIEVARYINSEVRG-
At_RPA1    GGD-SVMEHPQTVGLNKVIESNNILGNACTNKLDQSKVIFRWRNLQESVNVLFDSKTAT-
Sc_RPB1    FNESQRGEDDLTFKLADILKANISLETLEHNGAPHHAIEEAESLLQFHVATYMDNDIAGQ
Ec_RPOC    LDGGRFATSDLNDLYRRVINRNNRLKRLLDLAAPDIIVRNEKRMLQEAVDALLDNGRRGR
consensus  I G Q AE DLT RLR IIK N  L RIL NGAP   IMQ  RLLQE VATYFDSEI G

                          Conserved domain C
At_RPD1    -----ANPYQKKSDTPKLCGLR-FMKDVLLGKRSDHTFRTVVVGDPSLKLNEIGIPESIA
At_RPB1    PRATQKSGRPIKSICSRLKAKEGRIRGNLMGKRVDFSARTVITPDPTINIDELGVPWSIA
At_RPC1    -CQNQPEEHPLSGILQRLKGKGGRFRANLSGKRVEFTGRTVISPDPNLKITEVGIPILMA
At_RPA1    ----VQSQRDSSGICQLLEKKEGLFRQKMMGKRVNHACRSVISPDPYIAVNDIGIPPCFA
Sc_RPB1    PQALQKSGRPVKSIRARLKGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGVPKSIA
Ec_RPOC    -AITGSNKRPLKSLADMIKGKQGRFRQNLLGKRVDYSGRSVITVGPYLRLHQCGLPKKMA
consensus      Q S RPLKSI  RLKGKEGRFRGNLMGKRVDFSARTVISPDP LKL EIGIP SIA

At_RPD1    KRLQVSEHLNQCNKERLVTSFVPTLLDNKE------------------MHVRRGDRLVAI
At_RPB1    LNLTYPETVTPYNIERLKELVDYGPHPPGK--------TGAKYIIRDDGQRLDLRYLKKS
At_RPC1    QILTFPECVSRHNIEKLRQCVRNGPNKYPG---------ARNVRYPDGSSRTLVGDYRKR
At_RPA1    LKLTYPERVTPWNVEKLREAIINGPDIHPGATHYSDKSSTMKLPSTEKARRAIARKLLSS
Sc_RPB1    KTLTYPEVVTPYNIDRLTQLVRNGPNEHPG----------AKYVIRDSGDRIDLRYSKRA
Ec_RPOC    LELFKPFIYGKLELRGLATTIKAAKKMVER------------------------------
consensus  L LTYPE VTPYNIERLR  VRNGP   PG           K    D G R  LR LKK
                                        Conserved domain D
At_RPD1    QVNDLQTG---------DKIFRSLMDGDTVLMNRPPSIHQHSLIAMTVRILPTTSVVSLN
At_RPB1    SDQHLELG---------YKVERHLQDGDFVLFNRQPSLHKMSIMGHRIRIMP-YSTFRLN
At_RPC1    IADELAIG---------CIVDRHLQEGDVVLFNRQPSLHRMSIMCHRARIMP-WRTLRFN
At_RPA1    RGATTELGKTCDINFEGKTVHRHMRDGDIVLVNRQPTLHKPSLMAHKVRVLKGEKTLRLH
Sc_RPB1    GDIQLQYG---------WKVERHIMDNDPVLFNRQPSLHKMSMMAHRVKVIP-YSTFRLN
Ec_RPOC    -EEAVVWD-----------ILDEVIREHPVLLNRAPTLHRLGIQAFEPVLIE-GKAIQLH
consensus   D  L LG          KVERHLMDGD VLFNRQPSLHKMSIMAHRVRIIP YSTFRLN
                    * * * (active site)
At_RPD1    PICCLPFRGDFDGDCLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTAA
At_RPB1    LSVTSPYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQANRPVMGIVQDTLLGC
At_RPC1    ESVCNPYNADFDGDEMNMHVPQTEEARTEAITLMG-------------------------
At_RPA1    YANCSTYNADFDGDEMNVHFPQDEISRAEAYNIVNANNQYARPSNGEPLRALIQDHIVSS
Sc_RPB1    LSVTSPYNADFDGDEMNLHVPQSEETRAELSQLCAVPLQIVSPQSNKPCMGIVQDTLCGI
Ec_RPOC    PLVCAAYNADFDGDQMAVHVPLTLEAQLEARALMMSTNNILSPANGEPIIVPSQDVVLG-
consensus   SVCSPYNADFDGDEMNMHVPQSEEARAEA  LMAV  QIVSPQNGRPLMGIVQDTLLG
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
                                                                   Conserved
At_RPD1    YLVNVEKNCYLNRAQMQQLQM-------------------------YCPFQLPPPAIIKA
At_RPB1    RKI-TKRDTFIEKDVFMNTLM----------------------WWEDFDGKVPAPAILKP
At_RPC1    -------DTFYDRAAFSLICS--------------------YMGDGMDSIDLPTPTILKP
At_RPA1    VLL-TKRDTFLDKDHFNQLLFSSGVTDMVLSTFSGRSGKKVMVSASDAELLTVTPAILKP
Sc_RPB1    RKL-TLRDTFIELDQVLNMLY----------------------WVPDWDGVIPTPAIIKP
Ec_RPOC    -------LYYMTRDCVN------------------------AKGEGMVLTGPKEAERLYR
consensus    L T RDTFIDRD FNNLL                        D D  LPTPAILKP
            domain E
At_RPD1    SPSSTEPQWTGMQLFGMLFPPGFD-YTYPLNNVVV-------------------------
At_RPB1    -----RPLWTGKQVFNLIIPKQINLLRYSAWHADTETG----------------------
At_RPC1    -----IELWTGKQIFSVLLRPNASIRVYVTLNVKEKNFKKG-------------------
At_RPA1    -----VPLWTGKQVITAVLNQITKGHPPFTVEKATKLPVDFFKCRSREVKPNSGDLTKKK
Sc_RPB1    -----KPLWSGKQILSVAIPNGIHLQRF-----DEGTT----------------------
Ec_RPOC    -----SGLASLHARVKVRITEYEKDANG-----ELV------------------------
consensus        PLWTGKQIFGVLIP    L  Y     D

At_RPD1    SNGELLSFSEGSAWLRDGEGNFIERLLKHDKGKVLD---IIYSAQEMLSQWLLMRGLSVS
At_RPB1    ------FITPGDTQVRIERGELLAGTLCKKTLGT--------SNGSLVHVIWEEVGPDAA
At_RPC1    EHGFDETMCINDGWVYFRNSELISGQLGKATLALDIFPLGNGNKDGLYSILLRDYNSHAA
At_RPA1    EIDESWKQNLNEDKLHIRKNEFVCGVIDKAQFAD----------YGLVHTVHELYGSNAA
Sc_RPB1    ------LLSPKDNGMLIIDGQIIFGVVEKKTVGS--------SNGGLIHVVTREKGPQVC
Ec_RPOC    AKTSLKDTTVGRAILWMIVPKGLPYSIVNQALGKK-------AISKMLNTCYRILGLKPT
consensus         ISIGDA L I  GELI GVL K TLG         S  GLLHVV RD G  AA

At_RPD1    LADLYLSSDLQSRKNLTEEISYGLREAEQVCNKQQLMVESWRDFLAVNGEDKEEDSVSDL
At_RPB1    RKFLGHTQWLVNYWLLQNGFTIGIGDTIADSSTMEKINETISNAKTAVKDLIRQFQGKEL
At_RPC1    AVCMNRLAKLSARWIGIHGFSIGIDDVQPGEELSKERKDSIQFGYDQCHRKIEEFNRGNL
At_RPA1    GNLLSVFSRLFTVFLQTHGFTCGVDDLIILKDMDEERTKQLQECENVGERVLRKTFGIDV
Sc_RPB1    AKLFGNIQKVVNFWLLHNGFSTGIGDTIADGPTMREITETIAEAKKKVLDVTKEAQANLL
Ec_RPOC    VIFADQIMYTGFAYAARSGASVGIDDMVIP----EKKHEIISEAEAEVAEIQEQFQS---
consensus  A  L  I KL   WLL  GFSIGIDDLI      EEI ESI EA    V DVIEEFQG DL

At_RPD1    ARFCYE--------------RQKSATLSELAVSAFKDAYR------DVQALAYRYGDQSN
At_RPB1    DPEP----------------GRTMRDTFENRVNQVLNKAR-----DDAGSSAQKSLAETN
At_RPC1    QLKA----------------GLDGAKSLEAEITGILNTIR-----EATGKACMSGLHWRN
At_RPA1    DVQIDPQDMRSRIERILYEDGESALASLDRSIVNYLNQCSSKGVMNDLLSDGLLKTPGRN
Sc_RPB1    TAKH----------------GMTLRESFEDNVVRFLNEAR-----DKAGRLAEVNLKDLN
Ec_RPOC    --------------------GLVTAGERYNKVIDIWAAAN-----DRVSKAMM---DSFN
consensus                      GLT A S E  VV FLN AR     DDVGK AL   L   N
                                                       18 aa deleted
               Conserved domain F
At_RPD1    SFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAAWNDPNSPLRGAKGKD
At_RPB1    NLKAMVTAGSKGSFINISQMTACVGQQNVEGKRIPFGFDGRTLPHFTKDDYGPESR----
At_RPC1    SPLIMSQCGSKGSPINISQMVACVGQQTVNGHRAPDGFIDRSLPHFPRMSKSPAAK----
At_RPA1    CISLMTISGAKGSKVNFQQISSHLGQQDLEGKRVPRMVSGKTLPCFHPWDWSPRAG----
Sc_RPB1    NVKQMVMAGSKGSFINIAQMSACVGQQSVEGKRIAFGFVDRTLPHFSKDDYSPESK----
Ec_RPOC    SIYMMADSGARGSAAQIRQLAGMRGLMAKPDGS--------------------IIE----
consensus  SI IMS AGSKGS INI QMSACVGQQ VEGKRIP GF DRTLPHF K DYSP AK
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
                                   Bridge helix
At_RPD1    STTTESYVPYGVIENSFLTGLNPLESFVHSVTSRDSSFSGNADLP--GTLSRRLMFFMRD
At_RPB1    ----------GFVENSYLRGLTPQEFFFHAMGGREGLIDTAVKTSETGYIQRRLVKAMED
At_RPC1    ----------GFVANSFYSGLTATEFFFHTMGGREGLVDTAVKTASTGYMSRRLMKALED
At_RPA1    ----------GFISDRFLSGLRPQEYYFHCMAGREGLVDTAVKTSRSGYLQRCLMKNLES
Sc_RPB1    ----------GFVENSYLRGLTPQEFFFHAMGGREGLIDTAVKTAETGYIQRRLVKALED
Ec_RPOC    ----------TPITANFREGLNVLQYFISTHGARKGLADTALKTANSGYLTRRLVDVAQD
consensus            GFIENSFLSGLTPQEFFFHTMGGREGLIDTAVKTA TGYLQRRLMKALED

At_RPD1    IYAAYDGTVRNSFGNQLVQFTYETDGPVEDITG---------------------------
At_RPB1    IMVKYDGTVRNSLG-DVIQFLYGEDGMDAVWIESQKLDSLKMKKSEFDRTFKYEIDDENW
At_RPC1    LLVHYDNTVRNASG-CILQFTYGDDGMDPALME---------------------------
At_RPA1    LKVNYDCTVRDADG-SIIQFQYGEDGVDVHRSS---------------------------
Sc_RPB1    IMVHYDNTTRNSLG-NVIQFIYGEDGMDAAHIEKQSLDTIGGSDAAFEKRYRVDLLNTDH
Ec_RPOC    LVVTEDDCGTHEGI-MMTPVIEGGDVKEPLRDR---------------------------
consensus  IMV YD TVRNS G  IIQFIYGEDGMD   IE

At_RPD1    -----------------------------------------------EALGSLSACALSE
At_RPB1    NPTYLSDEHLEDLKGIRELRDVFDAEYSKLETDRFQLGTEIATNGDSTWPLPVNIKRHIW
At_RPC1    --------------------------------------------GKDGAPLNFNRLFLKV
At_RPA1    -------------------------------------------FIEKFKELTINQDMVLQ
Sc_RPB1    TLDPSLLESGSEILGDLKLQVLLDEEYKQLVKDRFLR-EVFVDGEANWPLPVNIRRIIQ
Ec_RPOC    ------------------------------------------------VLGRVTAEDVLK
consensus                                              GE    PL VN   LI

At_RPD1    AAYSALDQPIS----------LLETSPLLNLKNVLECGSKKGQREQTMSLYLSEYLSK--
At_RPB1    NAQKTFKIDLRKISDMHPVEIVDAVDKLQERLLVVPGDDALSVEAQKNATLFFNILLRST
At_RPC1    QATCPPRSHHTYLS-------SEELSQKFEEELVRHDKSRVCTDAFVKSLREFVSLLG--
At_RPA1    KCSEDMLSG--------------ASSYISDLPISLKKGAEKFVEAMPMNERIASKFVR--
Sc_RPB1    NAQQTFHIDHTKPSDLTIKDIVLGVKDLQENLLVLRGKNEIIQNAQRDAVTLFCCLLRSR
Ec_RPOC    PGTADILVPRN-----------TLLHEQWCDLLEENSVDAVKVRSVVSCDTDFGVCAH--
consensus  NAQ  I I  T  S        V ALS L E LLVL      V VEAQ    L LF  LLR

                                Conserved domain G
At_RPD1    ------------KKHGFEYGSLEIKNHLEKLSFSEIVSTSMIIFSPSSNTKVPLSPWVCH
At_RPB1    LASKRVLEEYKLSREAFEWVIGEIESRFLQSLVAPGEMIGCVAAQSIGEPATQMT--LNT
At_RPC1    ------------VKSASPPQVLYKASGVTDKQLEAGTAIGTIGAQSIGEPGTQMT--LKT
At_RPA1    ----------------QEELLKLVKSKFFASLAQPGEPVGVLAAQSVGEPSTQMT--LNT
Sc_RPB1    LATRRVLQEYRLTKQAFDWVLSNIEAQFLRSVVHPGEMVGVLAAQSIGEPATQMT--LNT
Ec_RPOC    ----------------------CYGRDLARGHIINKGEAIGVIAAQSIGEPGTQLT--MRT
consensus              K AFEWVL  IKS F  SLV PGE IGVIAAQSIGEPATQMT  LNT

At_RPD1    FHISEKVLKRKQLSAESVVSSLN-EQYKSRNRELKLDIVDLDIQNTNHCSSDDQAMKDDN
At_RPB1    FHYAGVSAKNVTLGVPRLREIIN-VAKRIKTPSLSVYLTPEASKSKEGAKTVQCALEYTT
At_RPC1    FHFAGVASMNITQGVPRINEIIN-ASKNISTPVISAELENPLELTS--ARWVKGRIEKTT
At_RPA1    FHLAGRGEMNVTLGIPRLQEILMTAAANIKTPIMTCPLLKG--KTKEDANDITDRLRKIT
Sc_RPB1    FHFAGVASKKVTSGVPRLKEILN-VAKNMKTPSLTVYLEPGHAADQEQAKLIRSAIEHTT
Ec_RPOC    FHIGGA---DITGGLPRVADLFE--ARRPKEPAILAEISGI-------------------
consensus  FHFAGVA KNVTLGVPRL EILN  AKNIKTP LSVEL      T E AK I  AIE TT
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
              192 aa deleted
At_RPD1   VCITVTVVEASKHSVLELDAIRLVLIPFL---------------------LDSPVKG---
At_RPB1   LRSVTQATEVWYDPDPMSTIIEEDFEFVR---------------------SYYEMPDEDV
At_RPC1   LGQVAESIEVLMTSTSASVRIILDN-------------------------KIIEEACLSI
At_RPA1   VADIIKSMELSVVPYTVYENEVCSIHKLKINLYKPEHYPKHTDITEEDWEETMRAVFLRK
Sc_RPB1   LKSVTIASEIYYDPDPRSTVIPEDEEIIQLHFS-------------------LLDEEAEQ
Ec_RPOC   ---------VSFGKETKGK-----------------------------------------
consensus L  V   IEVSY PDP S  I  D   I                        I

At_RPD1   ------------------------------------------------------------
At_RPB1   SPDKISPWLLRIELNREMMVDKKLSMADIAEKINLEFDDDLTCIFNDDNAQKLILRIRIM
At_RPC1   TPWSVKNSILKTPRIKLNDNDIRVLDTG--------------------------------
At_RPA1   LEDAIETHMKMLHRIRGIHNDVTGPIAGNETDNDDSVSGKQNEDDGDDDGEGTEVDDLGS
Sc_RPB1   SFDQQSPWLLRLELDRAAMNDKDLTMGQVGERIKQTFKNDLFVIWSEDNDEKLIIRCRVV
Ec_RPOC   ------------------------------------------------------------
consensus S D I   LLRL    R   ND  L MA                   DD      I

At_RPD1   -------------------------------------------------------DQGIK
At_RPB1   NDE--------------------------------GPKGELQDESAEDDVFLKKIESNML
At_RPC1   ---------------------------------------------LDITPVVDKSRAHFN
At_RPA1   DAQKQKKQETDEMDYEENSEDETNEPSSISGVEDPEMDSENEDTEVSKEDTPEPQEESME
Sc_RPB1   R-----------------------------------PKSLDAETEAEEDHMLKKIENTML
Ec_RPOC   ------------------------------------------------------------
consensus                                           D   E D  L K E  M

At_RPD1   KVNILWTDRPKAPKRNGNHLAGELYLKVTM------------------------------
At_RPB1   TEMALRGIPDINKVFIKQVRKSRFDEEGGF------------------------------
At_RPC1   LHNLKNGIKTVERVVVAEDMDKSKQIDG--------------------------------
At_RPA1   PQKEVKGVKNVKEQSKKKRRKFVRAKSDRHIFVKGEGEKFEVHFKFATDDPHILLAQIAQ
Sc_RPB1   ENITLRGVENIERVVMMKYDRKVPSPTGEYVK----------------------------
Ec_RPOC   -----------RRLVITPVDGSDP------------------------------------
consensus     LRGIK I RVVI     K      G

At_RPD1   YGDRGKRNCWTA--------------------------------------------LLET
At_RPB1   KTSEEWMLDTEG--------------------------------------------VNLL
At_RPC1   --KTKWKLFVEG--------------------------------------------TNLL
At_RPA1   QTAQKVYIQNSGKIERCTVANCGDPQVIYHGDNPKERREISNDEKKASPALHASGVDFPA
Sc_RPB1   --EPEWVLETDG--------------------------------------------VNLS
Ec_RPOC   --YEEMIPKWRQ--------------------------------------------LNVF
consensus     EWML  EG                                            LNL

At_RPD1   CLPIMDMIDWGRSHPDNIRQCCSVYGIDAGRSIFVANLESAVSDTGKEILREHLLLVADS
At_RPB1   AVMCHEDVDPKRTTSNHLIEIIEVLGIEAVRRALLDELRVVISFDGSYVNYRHLAILCDT
At_RPC1   AVMGTPGINGRTTTSNNVVEVSKTLGIEAARTTIIDEIGTVMGNHGMSIDIRHMMLLADV
At_RPA1   LWEFQDKLDVRYLYSNSIHDMLNIFGVEAARETIIREINHVFKSYGISVSIRHLNLIADY
Sc_RPB1   EVMTVPGIDPTRIYTNSFIDIMEVLGIEAGRAALYKEVYNVIASDGSYVNYRHMALLVDV
Ec_RPOC   EGERVERGDVISDGPEAPHDILRLRGVHAVTRYIVNEVQDVYRLQGVKINDKHIEVIVRQ
consensus  VM HD ID RRT SN IIDIL VLGIEAAR  II EI   VI    GI IN RHL LLAD
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
                                       Conserved domain H
At_RPD1    LSVTGEFVALNAKGWSKQRQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALA
At_RPB1    MTYRGHLMAITRHGIN-----RNDTGPLMRCSFEETVDILLDAAAYAETDCLRGVTENIM
At_RPC1    MTYRGEVLGIQRTGIQ-----KMDKSVLMQASFERTGDHLFSAAASGKVDNIEGVTECVI
At_RPA1    MTFSGGYRPMSRMGGI-----AESTSPFCRMTFETATKFIVQAATYGEKDTLETPSARIC
Sc_RPB1    MTTQGGLTSVTRHGFN-----RSNTGALMRCSFEETVEILFEAGASAELDDCRGVSENVI
Ec_RPOC    ML---SRDLLGITKAS-----LATESFISAASFQETTRVLTEAAVAGKRDELRGLKENVI
consensus  MTY G LLAITR G N      R TTSPLMRASFEETTDILLDAAA GERDDLRGVSENVI
              41 aa deleted

At_RPD1    WGKVPGFGTGDQFEIIISPKVHGFTTPVDVYDLLSSTKTMRRTNSAPKSDKATVQPFGLL
At_RPB1    LGQLAPIGTGDCELYLNDE-MLKNAIELQLPSYMDGLEFGMTPARSPVSGTPYHEGMMSP
At_RPC1    MGIPMKLGTGILKVLQRTDDLPK-------------LKYGPDPIIS--------------
At_RPA1    LGLPALSGTGCFDLMQRVEL----------------------------------------
Sc_RPB1    LGQMAPIGTGAFDVMIDEESLVKYMPEQKITEIEDGQDGGVTPYSN-------ESGLVNA
Ec_RPOC    VGRLIPAGTGYAYHQDRMRRRAAG------------------------------------
consensus  LG LAPIGTG  DLMIR E L K      I         G  P


At_RPD1    HSAFLKDIKVLDGKGIPMSLLRTIFTWKNIELLSQSLKRILHSYEINELLNERDEGLVKM
At_RPB1    NYLLSPNMRLSPMSDAQFSPYVGGMAFSPSSSPGYSPSSPGYSPTSPGYSPTSPGYSPTS
At_RPC1    ------------------------------------------------------------
At_RPA1    ------------------------------------------------------------
Sc_RPB1    DLDVKDELMFSPLVDSGSNDAMAG-GFTAYGGVDYG------EATSP---FAAYGEAPTS
Ec_RPOC    ----------------------------------------------------EAPAAPQV
consensus         I                  F                                P

At_RPD1    VLQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFEDFSYHKCVLGATKIIAPK
At_RPB1    PGYSPTSPTYSPSSPGYSPTSPAYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYS
At_RPC1    ------------------------------------------------------------
At_RPA1    ------------------------------------------------------------
Sc_RPB1    PGFGVSSPGFSPTSPTYSPTSPAYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYS
Ec_RPOC    TAEDASASLAELLNAGLGGSDNE-------------------------------------
consensus   G   P              T

At_RPD1    KMNFYKSKYLKNGTLESGGFSENP------------------------------------
At_RPB1    PTSPSYSPTSPAYSPTSPAYSPTSPAYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSP
At_RPC1    ------------------------------------------------------------
At_RPA1    ------------------------------------------------------------
Sc_RPB1    PTSPSYSPMSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSP
Ec_RPOC    ------------------------------------------------------------
consensus        S       S  S   YS

At_RPD1    ------------------------------------------------------------
At_RPB1    SYSPTSPAYSPTSPGYSPTSPSYSPTSPSYGPTSPSYNPQSAKYSPSIAYSPSNARLSPA
At_RPC1    ------------------------------------------------------------
At_RPA1    ------------------------------------------------------------
Sc_RPB1    AYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPNYSPTSPSYSPTSPGYSPGSPAYSP
Ec_RPOC    ------------------------------------------------------------
consensus
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
At_RPD1    ------------------------------------------------------------
At_RPB1    SPYSPTSPNYSPTSPSYSPTSPSYSPSSPTYSPSSPYSSGASPDYSPSAGYSPTLPGYSP
At_RPC1    ------------------------------------------------------------
At_RPA1    ------------------------------------------------------------
Sc_RPB1    ------------------------------------------------------------
Ec_RPOC    ------------------------------------------------------------
consensus

At_RPD1    ------------------------------
At_RPB1    SSTGQYTPHEGDKKDKTGKKDASKDDKGNP
At_RPC1    ------------------------------
At_RPA1    ------------------------------
Sc_RPB1    -------------KQDEQKHNENENSR---
Ec_RPOC    ------------------------------
consensus
```

The alignment was performed using ClustalX and then edited by hand using MacClade 4.03 prior to being exported to BOXSHADE for shading. Positions with identical amino acids are indicated by green shading, whereas similar amino acids are indicated by yellow shading. Previously published (Cramer et al., 2001) alignments and structural features were considered during the editing process. Regions of the *E. coli* β' subunit that do not align with the eukaryotic RNAPs were deleted, as indicated below the alignments. Conserved domains (Jokerst et al., 1989) are indicated with letters and bold lines above the alignments. The active site (metal A site; Cramer et al., 2001), is indicated by asterisks. Also noted is the bridge domain, which traverses the cleft in the polymerase near the active site. Domain assignments are according to Cramer et al. (2001). Protein sequences compared are: At_RPD1 (Pol IV), At_RPB1 (Pol II), At_RPC1 (Pol III), At_RPA1 (Pol I), Sc_RPB1 (Pol II), and Ec_RPOC (β′ subunit).

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Figure S2. Multiple Alignment of RPD2 with DNA-Dependent RNA Polymerase Second-Largest Subunits of *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 2. Alignment of RNAP 2nd Largest subunits

```
At_RPD2   MPDMDIDVKDLEEFEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
At_RPB2   --------MEYNEYEPEP-QYVEDDDDEEITQEDAWAVISAYFEEKGLVRQQLDSFDEFI
At_RPC2   ---MGLDQEDLDLTNDDHFIDKEKLSAPIKSTADKFQLVPEFLKVRGLVKQHLDSFNYFI
At_RPA2   --------------------------MVVNAKDSTVPTMEDFKELHNLVTHHIESFDYMT
Sc_RPB2   ----MSDLANSEKYYDED-PYGFEDESAPITAEDSWAVISAFFREKGLVSQQLDSFNQFV
Ec_RpoB   --------------------MVYSYTEKKRIRKDFGKRPQVLDVPY-LLSIQLDSFQKFI
consensus       D  D E Y  E       D     T  D W VIS FFE KGLVSQQLDSFNYFI

At_RPD2   EHGLQNVFQSFGEMLVEPSFDVVK--KKDNDWRYATVKFGEVTVEKPTFFSDD-KELEFL
At_RPB2   QNTMQEIVDESADIEIRPESQHNPGHQSDFAETIYKISFGQIYLSKPMMTESDGETATLF
At_RPC2   NVGIHKIVKANSRITS-------------TVDPSIYLRFKKVRVGEPSIINVN-TVENIN
At_RPA2   LKGLDVMFNRIKPVSVYDPN--------TENELSIWLENPLVFAPQKESFKSTSRKEPLL
Sc_RPB2   DYTLQDIICEDSTLILEQLAQHTT--ESDNISRKYEISFGKIYVTKPMVNESDGVTHALY
Ec_RpoB   EQDPEGQYGLEAAFRSVFPIQSYS--------GNSELQYVSYRLGEPV----------FD
consensus   GLQ I                     D  E    L FG VYV KP    SD     L

            Conserved domain A
At_RPD2   PWHARLQNMTYSARIKVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSIP
At_RPB2   PKAARLRNLTYSAPLYVDVTKRVIK----------KGHDG--EEVTETQDFTKVFIGKVP
At_RPC2   PHMCRLADMTYAAPIFVNIEYVHGS-----------------HGNKAKSAKDNVIIGRMP
At_RPA2   PFECRQAKISYTGTFMADVCFKYND---------------------GVVVRDKFDFGQFP
Sc_RPB2   PQEARLRNLTYSSGLFVDVKKRTYEAIDVPGRELKYELIA--EESEDDSESGKVFIGRLP
Ec_RpoB   VQECQIRGVTYSAPLRVKLRLVIYEREAPEGT-------------VKDIKEQEVYMGEIP
consensus P EARLRNLTYSAPLFVDV  RVFD                 E    DV K KVFIGRIP

                                          Conserved domain B
At_RPD2   VMVKSILCKTSEKG-KENCKKGDCAFDQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP--
At_RPB2   IMLRSSYCTLFQNSEKDLTELGECPYDQGGYFIINGSEKVLIAQEKMSTNHVYVFKKRQP
At_RPC2   IMLRSCRCVLHGKDEEELARLGECPLDPGGYFIIKGTEKVLLIQEQLSKNRIIIDSDK--
At_RPA2   IMLMSKLCSLKGADCRKLLKCKESTSEMGGYFILNGIERVFRCVIAPKRNHPTSMIRNSF
Sc_RPB2   IMLRSKNCYLSEATESDLYKLKECPFDMGGYFIINGSEKVLIAQERSAGNIVQVFKKAAP
Ec_RpoB   LMTDN------------------------GTFVINGTERVIVSQLHRSPGVFFDSDKGKT
consensus IMLRS  C L    EKDL KLGECPFD GGYFIINGSEKVLIAQE MS N VFI  K

At_RPD2   ----------WTVSFRSENKRNRFIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFF
At_RPB2   NKYAYVGEVRSMAENQNRPPSTMFVRMLARASAKGGSSGQYIRCTLPYIRTEIPIIIVFR
At_RPC2   ---------------KGNINASVTSSTEMTKSKTVIQMEKEKIYLFLHRFVKKIPIIIVLK
At_RPA2   RDRKEGYSSKAVVTRCVRDDQSSVTVKLYYLRNGSARVGFWIVGREYLLPVGLVLKALTN
Sc_RPB2   SPISHVAEIRSALEKGSRFISTLQVKLYGRE----GSSARTIKATLPYIKQDIPIVIIFR
Ec_RpoB   HSSGKVLYNARIIPYRGSWLDFEFDPKDN-----------LFVRIDRRR--KLPATIILR
consensus      V   R IV F  R  ST FV KL R     G  G  IV TL YI  EIPIIIIFR
```


Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
At_RPD2    ALGVSSDKEAMDLIAFDGDDASITNSLIASIHVADAVCEAFRCG---NNALTYVEQQIKS
At_RPB2    ALGFVADKDILEHICYDFADTQMMELLRPSLEEAFVIQNQLVALDYIGKRGATVGVTKEK
At_RPC2    AMGMESDQEIVQMVGRDPRFSASLLPSIEECVSEGVNTQKQALDYLEAKVKKISYGTPPE
At_RPA2    SCDEEIYESLNCCYSEHYGRGDGAIGTQLVRERAKIILDEVRDLGLFTREQCRKHLG-QH
Sc_RPB2    ALGIIPDGEILEHICYDVNDWQMLEMLKPCVEDGFVIQDRETALDFIGRRGTALGIKKEK
Ec_RpoB    ALNYTTEQILDLFFEKV----LFTNDLDHGPYISETLRVDPTNDRLSALVEIYRM-MRPG
consensus  ALGI SD EILE I YD  D  ML  L   IE A VI D    L L AK     V I K
                             107 aa deleted

                                                          Conserved domain C
At_RPD2    TKFPPAESVDECLHLYLFPGLQSLKKKARFLGYMVKCLLNSYAGKRKCENRDSFRNKRIE
At_RPB2    RIKYARDILQKEMLPHVGIGEHCETKKAYYFGYIIHRLLLCALGRRPEDDRDHYGNKRLD
At_RPC2    KDGRALSILRDLFLAHVPVPDNNFRQKCFYVGVMLRRMIEAMLNKDAMDDKDYVGNKRLE
At_RPA2    FQPVLDGVAEAVLRDYLFVHLDNDHDKFNLLIFIIQKLYSLVDQTSLPDNPDSLQNQEIL
Sc_RPB2    RIQYAKDILQKEFLPHITQLEGFESRKAFFLGYMINRLLLCALDRKDQDDRDHFGKKRLD
Ec_RpoB    EPPTREAAESLFENLFFSEDRYDL┐--KDDIIDVMKKLIDIRNGKGEVDDIDHLGNRRIR
consensus  R   A DIL    LL HL V E  E|KKAFFLGYMIKRLL   LGKR  DDRDHFGNKRID
                                     24 aa deleted

At_RPD2    LAGELLEREIRVHLAHARRKMTRAMQKHLSGDG---------DLKPIEHYLDASVITNGL
At_RPB2    LAGPLLGGLFRMLFRKLTRDVRSYVQKCVDNGK---------EVN-LQFAIKAKTITSGL
At_RPC2    LSGQLISLLFEDLFKTMLSEAIKNVDHILNKPIRAS----RFDFSQCLNKDSRYSISLGL
At_RPA2    VPGHVITIYLKEKLEEWLRKCKSLLKDELDNTNSKFSFESLADVKKLINKNPPRSIGTSI
Sc_RPB2    LAGPLLAQLFKTLFKKLTKDIFRYMQRTVEEAH---------DFN-MKLAINAKTITSGL
Ec_RpoB    SVGEMAENQFRVGLVRVERAVKERLSLGDL------------DTLMPQDMINAKPISAAV
consensus  LAG LL   LFRVLFKKL RDVKR LQK LD            DV   L    I AKSITSGL

At_RPD2    SRAFSTGAWSH-PFRKMERVSGVVANLGRANPLQTLIDLRRTRQQ----VLYTGKVGDAR
At_RPB2    KYSLATGNWG--QANAAGTRAGVSQVLNRLTYASTLSHLRRLNSP----IGREGKLAKPR
At_RPC2    ERTLSTGNFDI-KRFRMHRKG-MTQVLTRLSFIGSMGFITKISPQ----FEKSRKVSGPR
At_RPA2    ETLLKTGALKTQSGLDLQQRAGYTVQAERLNFLRFLSFFRAVHRGA---SFAGLRTTTVR
Sc_RPB2    KYALATGNWGE-QKKAMSSRAGVSQVLNRYTYSSTLSHLRRTNTP----IGRDGKLAKPR
Ec_RpoB    KEFFGSSQ--------------LSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVR
consensus  K   LATGNW        M   RAGVSQVL RLNFLSTLSHLRRI        I RDGKLA PR

           Conserved domain D
At_RPD2    YPHPSHWGRVCFLSTPDGENCGLVKNMSLLGLVSTQSLES--VVEKLFACGMEELMDDTC
At_RPB2    QLHNSQWGMMCPAETPEGQACGLVKNLALMVYITVGSAAYPILEFLEEWGTENFEEISPS
At_RPC2    SLQPSQWGMLCPCDTPEGESCGLVKNLALMTHVTTDEEEGPLVAMCYKLGVTDLEVLSAE
At_RPA2    KLLPESWGFLCPVHTPDGTPCGLLNHMTRTSRITSQFDSKGNIRDFLKIRKSVVDVLTGA
Sc_RPB2    QLHNTHWGLVCPAETPEGQACGLVKNLSLMSCISVGTDPMPIITFLSEWGMEPLEDYVPH
Ec_RpoB    DVHPTHYGRVCPIETPEGPNIGLINSLSVYAQTNEY------------------------
consensus   LHPSHWGMVCPIETPEG  CGLVKNLSLMG ITT SD   PII       G    EEVLS

At_RPD2    TPL--FGKHKVLLNGDWVGLCADSESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTD
At_RPB2    VI---PQATKIFVNGMWVGVHRDPDMLVKTLRRLRRRVDVNTEVGVVRDIRLKELRIYTD
At_RPC2    ELHTPDSFLVILNGLILGKHSRPQYFANSLRRLRRAGKIGEFVSVFTNEKQHCVYVASDV
At_RPA2    GMV--PSLPKLVRAGPPKVIHVLLDGQVVGTLSSNLVTKVVSYIRRLKVEAPSVIPEDLE
Sc_RPB2    QS---PDATRVFVNGVWHGVHRNPARLMETLRTLRRKGDINPEVSMIRDIREKELKIFTD
Ec_RpoB    ------------------------------------------------------------
consensus   I   P   KILVNGIW GVHR   D   V  LRS RR  DV  EV IIRD     ELRIFTD
```

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
At_RPD2    AGRLLRPLLVVEN------------LQKLKQEKPSQYP-------------FDHLLDHGI
At_RPB2    YGRCSRPLFIVDN------QKLLIKKRDIYALQQRESAEEDG---------WHHLVAKGF
At_RPC2    GRVCRPLVIADKG------------ISRVKQHHMKELQDGVR--------TFDDFIRDGL
At_RPA2    VGYVPTSMGGSYPG------------LYLASCPARFIRPVKN-----------ISIPSDN
Sc_RPB2    AGRVYRPLFIVEDDESLGHKELKVRKGHIAKLMATEYQDIEGGFEDVEEYTWSSLLNEGL
Ec_RpoB    -GFLETPYRKV--------------------------------------------TDGV
consensus  AGRL RPL IVE                I      RE   D            F   LI DGL

At_RPD2    LELIGIEEEEDCNTAWGIKQLLKEPK------------------------IYTHCELDLS
At_RPB2    IEYIDTEEEETTMISMTISDLVQARLRPEE----------------AYTENYTHCEIHPS
At_RPC2    IEYLDVNEENNALVCLRAEAAK---------------------------ADTTHIEIEPF
At_RPA2    IELIGPFEQVANPINIIFISTFP----------------------------ATHEEIHPT
Sc_RPB2    VEYIDAEEEESILIAMQPEDLEPAEANEENDLDVDPAK---RIRVSHHATTFTHCEIHPS
Ec_RpoB    VTDEIHYLSAIEEGNYVIAQANSNLDEEGHFVEDLVTCRSKGESSLFSRDQVDYMDVSTQ
consensus  IEYID EEEE   LI M I   L                             YTHCEIHPS

             Conserved domain E
At_RPD2    FLLGVSCAVVPFANHDHGRRVLYQSQKHCQQAIGFSSTNPNIRCDTLSQQLFYPQKPLFK
At_RPB2    LILGVCASIIPFPDHNQSPRNTYQS-AMGKQAMGIYVTNYQFRMDTLAYVLYYPQKPLVT
At_RPC2    TILGVVAGLIPYPHHNQSPRNTYQC-AMGKQAMGNIAYNQLNRMDTLLYLLVYPQRPLLT
At_RPA2    GMISVVANLTPWSDHNQSPRNMYQC-QMAKQTMAYSTQALQFRADQKIYHLQTPQSPVVR
Sc_RPB2    MILGVAASIIPFPDHNQSPRNTYQS-AMGKQAMGVFLTNYNVRMDTMANILYYPQKPLGT
Ec_RpoB    QVVSVGASLIPFLEHDDANRALMGA-NMQRQAVPT------LRAD----------KPLVG
consensus   ILGV ASLIPFPDHNQSPRNTYQS AMGKQAMG    TN N RMDTL YLLYYPQKPLVT

                     Conserved domain F   **(active site)
At_RPD2    TLASECLKKEVLFNGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDA
At_RPB2    TRAMEHLHFRQLPAGINAIVAISCYSGYNQEDSVIMNQSSIDRGFFRSLFFRSYRDEEKK
At_RPC2    TRTIELVGYDKLGAGQNATVAVMSFSGYDIEDAIVMNKSSLDRGFGRCIVMKKIVAMSQK
At_RPA2    TKTYTTYSIDENPTGTNAIVAVLAHTGFDMEDAMILNKSSVERGMCHGQIYQTENIDLSD
Sc_RPB2    TRAMEYLKFRELPAGQNAIVAIACYSGYNQEDSMIMNQSSIDRGLFRSLFFRSYMDQEKK
Ec_RpoB    TGMERAVAV-ELALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRFTTIHI--QELACVS
consensus  TRAME L  FDELPAGQNAIVAVL YSGYNQEDSIIMNKSSIDRGMFRSI FRSY   E  K
              82 aa deleted

                                       Conserved domain G
At_RPD2    KDSEKRKKMDELVQFGKTHSKIGKVDSLEDDGFPFIGANMSTGDIVIGRCTESG------
At_RPB2    MGTLVKEDFGRPDRGSTMGMRHGSYDKLDDDGLAPPGTRVSGEDVIIGKTTPISQDEAQG
At_RPC2    YDNCTADRILIPQR---TGPDAEKMQILDDDGLATPGEIIRPNDIYINKQVPVDTVTKFT
At_RPA2    QNS----RFDSGSKSFRRSTNKAEHFRIDADGLPSVGQKLYPDEPYCSIYDEVTN-----
Sc_RPB2    YGMSITETFEKPQRTNTLRMKHGTYDKLDDDGLIAPGVRVSGEDVIIGKTTPISPDEEEL
Ec_RpoB    RDTKLGPEEITADIPNVG---EAALSKLDESGIVYIGAEVTGGDILVGKVTPKGETQL--
consensus   DS  I ERFD  P  R       K  G LDKLDDDGL   PG RVSGEDIIIGK TPIS
                                                            9 aa deleted

                                                 Conserved domain H
At_RPD2    ----------ADHSIKLKHTERGIVQKVVLSS-NDEGKNFAAVSLRQVRSPCLGDKFSSM
At_RPB2    --QS-SRYTRRDHSISLRHSETGMVDQVLLTT-NADGLRFVKVRVRSVRIPQIGDKFSSR
At_RPC2    SALSDSQYRPAREYFKGPEGETQVVDRVALCS-DKKGQLCIKYIIRHTRRPELGDKFSSR
At_RPA2    ----------KTRHMKRKGTDPVIVDFVSVDMKSKKHPQRANIRFRHARNPIIGDKFSSR
Sc_RPB2    GQRT-AYHSKRDASTPLRSTENGIVDQVLVTT-NQDGLKFVKVRVRTTKIPQIGDKFASR
Ec_RpoB    -IFGEKASDVKDSSLRVPNGVSGTVIDVQV-------LKIVKVYLAVKRRIQPGDKMAGR
consensus     S         KD  SIKLK  TETGIVD  VLLTS N  DGLKFVKVRLR    R  PQIGDKFSSR
                                     113 aa deleted
```


Copyright ©2005 by Cell Press

Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

```
At_RPD2     HGQKGVLGYLEEQQNFPFT-IQGIVPDIVINPHAFPSRQTPGQLLEAALSKGIACP---I
At_RPB2     HGQKGTVGMTYTQEDMPWT-IEGVTPDIIVNPHAIPSRMTIGQLIECIMGK---------
At_RPC2     HGQKGVCGIIIQQEDFPFS-ELGICPDLIMNPHGFPSRMTVGKMIELLGSKAG-------
At_RPA2     HGQKGVCSQLWPDIDMPFNGVTGMRPDLIINPHAFPSRMTIAMLLESIAAKGGSLHGKFV
Sc_RPB2     HGQKGTIGITYRREDMPFT-AEGIVPDLIINPHAIPSRMTVAHLIECLLSK---------
Ec_RpoB     HGNKGVISKINPIEDMPYD-ENGTPVDIVLNPLGVPSRMNIGQILETHLGMAAKGIGDKI
consensus   HGQKGVIGIIY QEDMPFT I GI PDIIINPHAFPSRMTIGQLIE ILSKAG      I
```

```
                                         Conserved domain I
At_RPD2     QKEGSSAAYTKLTRHATPFSTPGVTEITEQLHRAGFSRWGNERVYNGRSGEMMRSMIFMG
At_RPB2     -----VAAHMGKEGDATPFTDVTVDNISKALHKCGYQMRGFERMYNGHTGRPLTAMIFLG
At_RPC2     --VSCGRFHYGSAFGERSGHADKVETISATLVEKGFSYSGKDLLYSGISGEPVEAYIFMG
At_RPA2     DATPFRDAVKKTNGEEESKSSLLVDDLGSMLKEKGFNHYGTETLYSGYLGVELKCEIFMG
Sc_RPB2     -----VAALSGNEGDASPFTDITVEGISKLLREHGYQSRGFEVMYNGHTGKKLMAQIFFG
Ec_RpoB     NAMLKQQQEVAKLREFIQ--|---------LLKLGDLPTSGQIRLYDGRTGEQFERPVTVG
consensus         AA   G   GDATPFS |I VD IS LLHE GFQ   G ERLYNG TGE L A IFMG
                           51 aa deleted
```

```
At_RPD2     PTFYQRLVHMSEDKVKFRNTGPVHPLTRQPVADRKRFGGIKFGEMERDCLIAHGASANLH
At_RPB2     PTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFLK
At_RPC2     PIYYQKLKHMVLDKMHARGSGPRVMMTRQPTEGKSKNGGLRVGEMERDCLIAYGASMLIY
At_RPA2     PVYYQRLRHMVSDKFQVRSTGQVDQLTHQPIKGRKRGGGIRFGEMERDSLLAHGASYLLH
Sc_RPB2     PTYYQRLRHMVDDKIHARARGPMQVLTRQPVEGRSRDGGLRFGEMERDCMIAHGAASFLK
Ec_RpoB     YMYMLKLNHLVDDKMHARSTGSYSLVTQQPLGGKAQFGGQRFGEMEVWALEAYGAAYTLQ
consensus   PTYYQRLKHMVDDKIHARGTGPV ILTRQPVEGRSR GGLRFGEMERDCLIAHGAS   L
```

```
At_RPD2     ERLFTLSDSSQMHICRKCKTYANVIERTPSSG------------RKIRGPYCRVCVSSDH
At_RPB2     ERLFDQSDAYRVHVCEVCG-LIAIANLKKNS------------------FECRGCKNKTD
At_RPC2     ERLMISSDPFEVQVCRACGLLGYYNYKLKKA-------------------VCTTCKNGDN
At_RPA2     DRLHTSSDHHIADVCSLCGSLLTSSVVNVQQKKLIQEIGKLPPGRTPKKVTCYSCKTSKG
Sc_RPB2     ERLMEASDAFRVHICGICGLMTVIAKLNHNQ------------------FECKGCDNKID
Ec_RpoB     EMLTVKSDDVNGRTKMYKNIVDGNHQMEP-------------------------------
consensus   ERL    SD F VHVC ICGLL   I  L  N                  CR CKN
```

```
At_RPD2     VVRVYVPYGAKLLCQELFSMGITLNFDTKLC-------
At_RPB2     IVQVYIPYACKLLFQELMSMAIAPRMLTKHLKSAKGRQ
At_RPC2     IATMKLPYACKLLFQELQSMNVVPRLKLTEA-------
At_RPA2     METVAMPYVFRYLAAELASMNIKMTLQLSDREGVTD--
Sc_RPB2     IYQIHIPYAAKLLFQELMAMNITPRLYTDRSRDF----
Ec_RpoB     ----GMPESFNVLLKEIRSLGINIELEDE---------
consensus   I   V IPYA KLLFQEL SMNI PRL T
```

The alignment was performed as described previously for the largest subunits. Positions with identical amino acids are indicated by green shading, while similar amino acids are indicated by yellow shading. The last line in the alignment indicates the consensus sequence. Conserved domains (Sweetser et al., 1987) are indicated with letters and bold lines above the alignments. The active site (metal B site; Cramer et al., 2001) is indicated with asterisks. Protein sequences examined are: At_RPD2 (Pol IV), At_RPB2 (Pol II), At_RPC2 (Pol III), At_RPA2 (Pol I), Sc_RPB2 (Pol II) and Ec_RpoB. Regions of the *E. coli* β subunit that do not align with the eukaryotic RNAP proteins were deleted, as indicated below the alignments.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Figure S3. Comparison of Conserved Domains A–H in RPD1a and DNA-Dependent RNA Polymerase Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 3. Domain Alignments for DNA-dependent RNA Polymerase Largest Subunits

```
DOMAIN  GENE      AMINO ACIDS SEQUENCE

A       At RPD1   37-96       EAPNQVTDSRLGLPNPDSVCRTCGSKDRKVCEGHFGVINFAYSIINPYFLKEVAALLNKI
        At RPB1   47-105      PKVGGLSDTRLGTIDRKVKCETCMAN-MAECPGHFGYLELAKPMYHVGFMKTVLSIMRCV
        At RPC1   56-114      PYENGLLDPRMGPPNKKSICTTCEGN-FQNCPGHYGYLKLDLPVYNVGYFNFILDILKCI
        At RPA1   60-118      PFPGGLYDLKLGPKDDKQACNSCGQL-KLACPGHCGHIELVFPIYHPLLFNLLFNFLQRA
        Sc RPB1   48-106      AKIGGLNDPRLGSIDRNLKCQTCQEG-MNECPGHFGHIDLAKPVFHVGFIAKIKKVCECV
        Ec RPOC   51-106      PERDGLFCARIFGPVKDYECLCGKYK-RLK---HRGVICEKCGVEVTQTKVRRERMGHIE
        consensus             P  GGL D RLG PDKK  C TC    R  CPGHFG IELA PVYHVGFI  I  IL CI

B       At RPD1   214-260     MFNSLGLTSFPVTPNGYRVTEIVHQFNGARLIFDERTRIYKKLVGFE
        At RPB1   242-288     RPDWMILEVLPIPPPPVRPSVMMDATSRSEDDLTHQLAMIIRHNENL
        At RPC1   249-295     RPENLIITCMLVPPLSIRPSVMIGGIQSNENDLTARLKQIILGNASL
        At RPA1   337-382     DYSMFFLESVLVPPTKFRPPTTGGD-SVMEHPQTVGLNKVIESNNIL
        Sc RPB1   230-276     RPEWMILTCLPVPPPPVRPSISFNESQRGEDDLTFKLADILKANISL
        Ec RPOC   233-279     KPEWMILTVLPVLPPDLRPLVPLDGGRFATSDLNDLYRRVINRNNRL
        consensus             RPEWMILT LPVPPP VRPSVMI G Q AE DLT RLR IIK N  L

C       At RPD1   301-356     PKLCGLR-FMKDVLLGKRSDHTFRTVVVGDPSLKLNEIGIPESIAKRLQVSEHLNQC
        At RPB1   339-395     SRLKAKEGRIRGNLMGKRVDFSARTVITPDPTINIDELGVPWSIALNLTYPETVTPY
        At RPC1   344-400     QRLKGKGGRFRANLSGKRVEFTGRTVISPDPNLKITEVGIPILMAQILTFPECVSRH
        At RPA1   428-484     QLLEKKEGLFRQKMMGKRVNHACRSVISPDPYIAVNDIGIPPCFALKLTYPERVTPW
        Sc RPB1   327-383     ARLKGKEGRIRGNLMGKRVDFSARTVISGDPNLELDQVGVPKSIAKTLTYPEVVTPY
        Ec RPOC   329-385     DMIKGKQGRFRQNLLGKRVDYSGRSVITVGPYLRLHQCGLPKKMALELFKPFIYGKL
        consensus              RLKGKEGRFRGNLMGKRVDFSARTVISPDP LKL EIGIP SIAL LTYPE VTPY

D       At RPD1   407-468     VLMNRPPSIHQHSLIAMTVRILPTTSVVSLNPICCLPFRGDFDGDCLHGYVPQSIQAKVELD
        At RPB1   451-511     VLFNRQPSLHKMSIMGHRIRIMP-YSTFRLNLSVTSPYNADFDGDEMNMHVPQSFETRAEVL
        At RPC1   460-520     VLFNRQPSLHRMSIMCHRARIMP-WRTLRFNESVCNPYNADFDGDEMNMHVPQTEEARTEAI
        At RPA1   562-623     VLVNRQPTLHKPSLMAHKVRVLKGEKTLRLHYANCSTYNADFDGDEMNVHFPQDEISRAEAY
        Sc RPB1   442-502     VLFNRQPSLHKMSMMAHRVKVIP-YSTFRLNLSVTSPYNADFDGDEMNLHVPQSEETRAELS
        Ec RPOC   421-481     VLLNRAPTLHRLGIQAFEPVLIE-GKAIQLHPLVCAAYNADFDGDQMAVHVPLTLEAQLEAR
        consensus             VLFNRQPSLHKMSIMAHRVRIIP YSTLRLN SVCSPYNADFDGDEMNMHVPQSEEARAEA

E       At RPD1   524-560     LPPPAIIKASPSSTEPQWTGMQLFGMLFPPGFD-YTYPLNNVVV
        At RPB1   569-607     VPAPAILKP-----RPLWTGKQVFNLIIPKQINLLRYSAWHADT
        At RPC1   549-587     LPTPTILKP-----IELWTGKQIFSVLLRPNASIRVYVTLNVKE
        At RPA1   703-741     TVTPAILKP-----VPLWTGKQVITAVLNQITKGHPPFTVEKAT
        Sc RPB1   560-593     IPTPAIIKP-----KPLWSGKQILSVAIPNGIHLQRF-----DE
        Ec RPOC   530-563     PKEAERLYR-----SGLASLHARVKVRITEYEKDANG-----EL
        consensus             LPTPAILKP      PLWTGKQIFGVLIP    L  Y     D

F       At RPD1   728-817     MSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAAWNDPNSPLRGAKGKDSTTTESYVPYGVIENSFLTGLNPLESFVHSVTSR
        At RPB1   760-835     MVTAGSKGSFINISQMTACVGQQNVEGKRIPFGFDGRTLPHFTKDDYGPESR--------------GFVENSYLRGLTPQEFFFHAMGGR
        At RPC1   757-832     MSQCGSKGSPINISQMVACVGQQTVNGHRAPDGFIDRSLPHFPRMSKSPAAK--------------GFVANSFYSGLTATEFFFHTMGGR
        At RPA1   941-1016    MTISGAKGSKVNFQQISSHLGQQDLEGKRVPRMVSGKTLPCFHPWDWSPRAG--------------GFISDRFLSGLRPQEYYFHCMAGR
        Sc RPB1   746-821     MVMAGSKGSFINIAQMSACVGQQSVEGKRIAFGFVDRTLPHFSKDDYSPESK--------------GFVENSYLRGLTPQEFFFHAMGGR
        Ec RPOC   725-780     MADSGARGSAAQIRQLAGMRGLMAKPDGS--------------------IIE--------------TPITANFREGLNVLQYFISTHGAR
        consensus             MS AGSKGS INI QMSACVGQQ VEGKRIP GF DRTLPHF K DYSP AK              GFIENSFLSGLTPQEFFFHTMGGR

        At RPD1   818-843     DSSFSGNADLP--GTLSRRLMFFMRDIY
        At RPB1   836-863     EGLIDTAVKTSETGYIQRRLVKAMEDIM
        At RPC1   833-860     EGLVDTAVKTASTGYMSRRLMKALEDLL
        At RPA1   1017-1044   EGLVDTAVKTSRSGYLQRCLMKNLESLK
        Sc RPB1   822-849     EGLIDTAVKTAETGYIQRRLVKALEDIM
        Ec RPOC   781-788     KGLADTALKTANSGYLTRRLVDVAQDLV
        consensus             EGLIDTAVKTA TGYLQRRLMKALEDIM

G       At RPD1   945-1006    LEIKNHLEKLSFSEIVSTSMIIFSPSSNTKVPLSPWVCHFHISEKVLKRKQLSAESVVSSLN
        At RPB1   1062-1121   GEIESRFLQSLVAPGEMIGCVAAQSIGEPATQMT--LNTFHYAGVSAKNVTLGVPRLREIIN
        At RPC1   967-1026    LYKASGVTDKQLEAGTAIGTIGAQSIGEPGTQMT--LKTFHFAGVASMNITQGVPRINEIIN
        At RPA1   1041-1100   KLVKSKFFASLAQPGEPVGVLAAQSVGEPSTQMT--LNTFHLAGRGEMNVTLGIPRLQEILM
        Sc RPB1   1047-1106   SNIEAQFLRSVVHPGEMVGVLAAQSIGEPATQMT--LNTFHFAGVASKKVTSGVPRLKEILN
        Ec RPOC   898-1146    CYGRDLARGHIINKGEAIGVIAAQSIGEPGTQLT--MRTFHIGGA---DITGGLPRVADLFE
        consensus               IKS F  SLV PGE IGVIAAQSIGEPATQMT  LNTFHFAGVA KNVTLGVPRL EILN
                                                                           192 aa deleted
H       At RPD1   1214-1268   PAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALAWGKVPGFGTGDQFEIIISP
        At RPB1   1410-1464   TGPLMRCSFEETVDILLDAAAYAETDCLRGVTENIMLGQLAPIGTGDCELYLNDE
        At RPC1   1261-1315   KSVLMQASFERTGDHLFSAAASGKVDNIEGVTECVIMGIPMKLGTGILKVLQRTD
        At RPA1   1515-1569   TSPFCRMTFETATKFIVQAATYGEKDTLETPSARICLGLPALSGTGCFDLMQRVE
        Sc RPB1   1394-1448   TGALMRCSFEETVEILFEAGASAELDDCRGVSENVILGQMAPIGTGAFDVMIDEE
        Ec RPOC   1317-1371   ESFISAASFQETTRVLTEAAVAGKRDELRGLKENVIVGRLIPAGTGYAYHQDRMR
        consensus             TSPLMRASFEETTDILLDAAA GERDDLRGVSENVILG LAPIGTG  DLMIR E
```

The alignment for each conserved domain, determined using ClustalX, was exported to BOXSHADE. Positions with identical amino acids are indicated by green shading; similar amino acids are indicated by yellow shading. The last line in the alignment indicates the consensus for all sequences. Proteins whose domains are aligned are: At_RPD1 (Pol IV), At_RPB1 (Pol II), At_RPC1 (Pol III), At_RPA1 (Pol I), Sc_RPB1 (Pol II), and Ec_RPOC (β' subunit)

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Figure S4. Comparison of Conserved Domains A–I in RPD2 and DNA-Dependent RNA Polymerase Second-Largest Subunits in *A. thaliana* (At), *S. cerevisiae* (Sc), and *E. coli* (Ec)

Supplemental Figure 4. Domain Alignments for RNAP Second-Largest Subunits

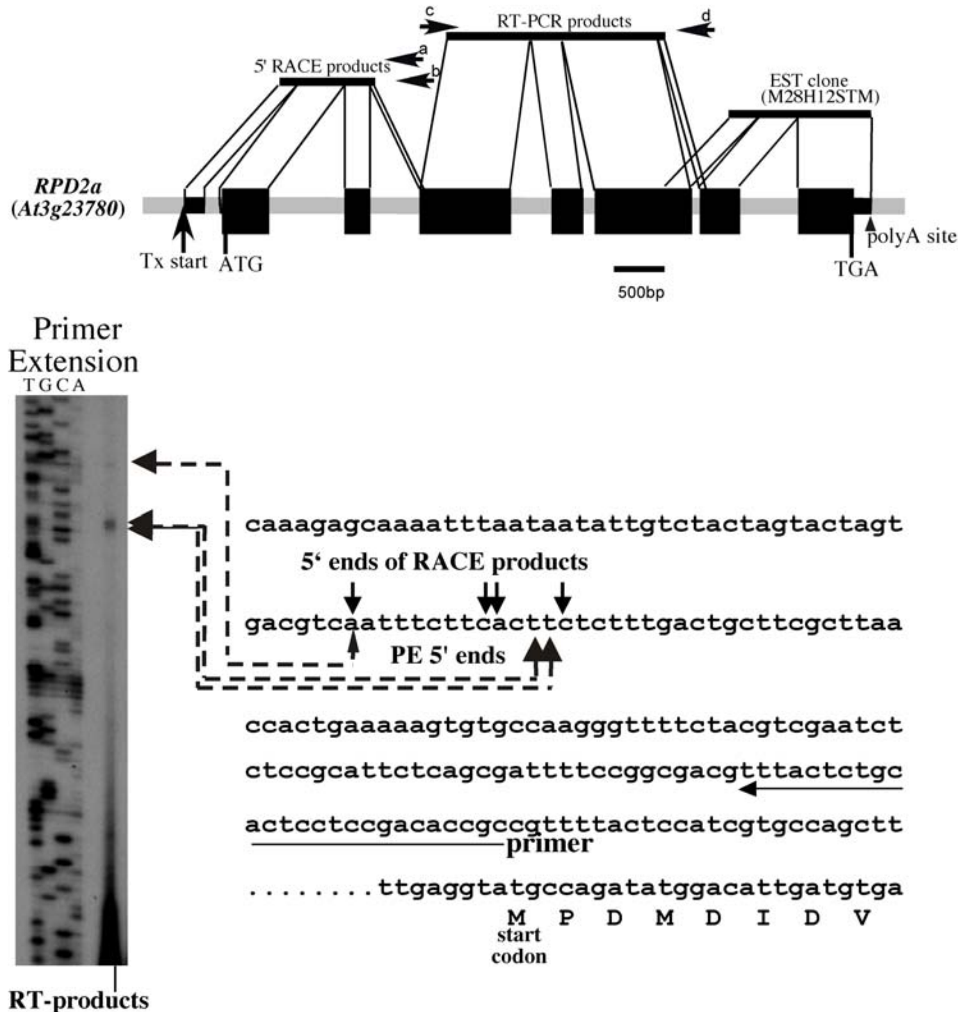
```
DOMAIN   GENE      AMINO ACIDS   SEQUENCE

A        At RPD2     119-142     WHARLQNMTYSARIKVNVQVEVFK
         At RPB2     113-136     KAARLRNLTYSAPLYVDVTKRVIK
         At RPC2     105-128     HMCRLADMTYAAPIFVNIEYVHGS
         At RPA2      38-61      FECRQAKISYTGTFMADVCFKYND
         Sc RPB2     115-138     QEARLRNLTYSSGLFVDVKKRTYE
         Ec RPOB      83-106     QECQIRGVTYSAPLRVKLRLVIYE
         consensus                e rlrnvtYsaplyvdv  riye

B        At RPD2     206-220     GYFVIKGAEKVFIAQ
         At RPB2     189-203     GYFIINGSEKVLIAQ
         At RPC2     176-190     GYFIIKGTEKVLLIQ
         At RPA2     105-119     GYFILNGIERVFRCV
         Sc RPB2     201-215     GYFIINGSEKVLIAQ
         Ec RPOB     134-148     GTFVINGTERVIVSQ
         consensus               GyFiinGtEkVliaq

C        At RPD2     385-400     GKRKCENRDSFRNKRI
         At RPB2     383-398     GRRPEDDRDHYGNKRL
         At RPC2     354-369     NKDAMDDKDYVGNKRL
         At RPA2     298-313     QTSLPDNPDSLQNQEI
         Sc RPB2     391-406     DRKDQDDRDHFGKKRL
         Ec RPOB     438-453     GKGEVDDIDHLGNRRI
         consensus               gkr  ddrDh gnkri

D        At RPD2     507-536     RYPHPSHWGRVCFLSTPDGENCGLVKNMSL
         At RPB2     503-532     RQLHNSQWGMMCPAETPEGQACGLVKNLAL
         At RPC2     480-509     RSLQPSQWGMLCPCDTPEGESCGLVKNLAL
         At RPA2     431-460     RKLLPESWGFLCPVHTPDGTPCGLLNHMTR
         Sc RPB2     512-541     RQLHNTHWGLVCPAETPEGQACGLVKNLSL
         Ec RPOB     548-577     RDVHPTHYGRVCPIETPEGPNIGLINSLSV
         consensus               R lhpshwGmvCpieTPeG  cGLvknlsl

E        At RPD2     696-714     LLGVSCAVVPFANHDHGRR
         At RPB2     711-729     ILGVCASIIPFPDHNQSPR
         At RPC2     675-693     ILGVVAGLIPYPHHNQSPR
         At RPA2     620-638     MISVVANLTPWSDHNQSPR
         Sc RPB2     748-766     ILGVAASIIPFPDHNQSPR
         Ec RPOB     660-678     VVSVGASLIPFLEHDDANR
         consensus               ilgV asliPfpdHnqspR

F        At RPD2     765-800     VLFNGQNAIVAVNVHLGYNQEDSIVMNKASLERGMF
         At RPB2     779-814     QLPAGINAIVAISCYSGYNQEDSVIMNQSSIDRGFF
         At RPC2     743-778     KLGAGQNATVAVMSFSGYDIEDAIVMNKSSLDRGFG
         At RPA2     688-723     ENPTGTNAIVAVLAHTGFDMEDAMILNKSSVERGMC
         Sc RPB2     816-851     ELPAGQNAIVAIACYSGYNQEDSMIMNQSSIDRGLF
         Ec RPOB     793-828     ELALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRF
         consensus               elpaGqNaiVAvm wsGynqEDsiimnkssvdrgmf

G        At RPD2     836-866     IGKVDSLEDDGFPFIGANMSTGDIVIGRCTE
         At RPB2     850-880     HGSYDKLDDDGLAPPGTRVSGEDVIIGKTTP
         At RPC2     811-841     AEKMQILDDDGLATPGEIIRPNDIYINKQVP
         At RPA2     755-785     KAEHFRIDADGLPSVGQKLYPDEPYCSIYDE
         Sc RPB2     887-917     HGTYDKLDDDGLIAPGVRVSGEDVIIGKTTP
         Ec RPOB     859-889     EAALSKLDESGIVYIGAEVTGGDILVGKVTP
         consensus               hg ldklddddGl  pG rvsgediligk tp

H        At RPD2     895-966     KNFAAVSLRQVRSPCLGDKFSSMHGQKGVLGYLEEQQNFPFT-IQGIVPDIVINPHAFPSRQTPGQLLEAALS
         At RPB2     922-993     LRFVKVRVRSVRIPQIGDKFSSRHGQKGTVGMTYTQEDMPWT-IEGVTPDIIVNPHAIPSRMTIGQLIECIMG
         At RPC2     886-957     QLCIKYIIRHTRRPELGDKFSSRHGQKGVCGIIIQQEDFPFS-ELGICPDLIMNPHGFPSRMTVGKMIELLGS
         At RPA2     816-888     PQRANIRFRHARNPIVGDKFSSRHGQKGVCSQLWPDIDMPFNGVTGMRPDLIINPHAFPSRMTIAMLLESIAA
         Sc RPB2     961-1032    LKFVKVRVRTTKIPQIGDKFASRHGQKGTIGITYRREDMPFT-AEGIVPDLIINPHAIPSRMTVAHLIECLLS
         Ec RPOB    1047-1118    LKIVKVYLAVKRRIQPGDKMAGRHGNKGVISKINPIEDMPYD-ENGTPVDIVLNPLGVPSRMNIGQILETHLG
         consensus               lkfvkvrlr  r pqlGDKfssrHGqKGvigmiy qedmPft i Gi pDiiiNPhafPSRmtigqllE ils

I     At RPD2    1003-1101
QLHRAGFSRWGNERVYNGRSGEMMRSMIFMGPTFYQRLVHMSEDKVKFRNTGPVHPLTRQPVADRKRFGGIKFGEMERDCLIAHGASANLHERLFTLSD
         At RPB2    1019-1117    ALHKCGYQMRGFERMYNGHTGRPLTAMIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFLKERLFDQSD
         At RPC2     988-1086    TLVEKGFSYSGKDLLYSGISGEPVEAYIFMGPIYYQKLKHMVLDKMHARGSGPRVMMTRQPTEGKSKNGGLRVGEMERDCLIAYGASMLIYERLMISSD
         At RPA2     928-1026    MLKEKGFNHYGTETLYSGYLGVELKCEIFMGPVYYQRLRHMVSDKFQVRSTGQVDQLTHQPIKGRKRGGGIRFGEMERDSLLAHGASYLLHDRLHTSSD
         Sc RPB2    1058-1156    LLREHGYQSRGFEVMYNGHTGKKLMAQIFFGPTYYQRLRHMVDDKIHARARGPMQVLTRQPVEGRSRDGGLRFGEMERDCMIAHGAASFLKERLMEASD
         Ec RPOB    1198-1296    LLKLGDLPTSGQIRLYDGRTGEQFERPVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTQQPLGGKAQFGGQRFGEMEVWALEAYGAAYTLQEMLTVKSD
         consensus               lLkekgfq  G erlYnGrtGe l a ifmGptyyqrLkHmvdDKmhaRgtGpv llTrQPlegrsr GGlrfGEMErdcliAhGA   lherL   SD
```

The alignments were conducted and displayed as described for Supplemental Figure 3. Proteins whose domains are aligned are: At_RPD2 (Pol IV), At_RPB2 (Pol II), At_RPC2 (Pol III), At_RPA2 (Pol I), Sc_RPB2 (Pol II), and Ec_RPOB (β′ subunit).

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

III. Determination of RPD2a Full-Length mRNA Sequence

Figure S5. Determination of the Full-Length mRNA Sequence for RPD2a by RT-PCR, 5′ RACE, and Primer Extension



The diagram shows the relative positions of the eight exons, depicted as black rectangles with coding regions expanded in size relative to the 5' and 3' untranslated regions. The transcription (Tx) start site, initiation codon (ATG), stop codon (TGA) and poly A addition sites are indicated. Also shown are the relative positions of a pre-existing partial cDNA (EST M28H12STM) and the clones obtained by RT-PCR and 5' RACE that were sequenced as part of this study. Shown at the lower left is an autoradiogram displaying primer extension products run adjacent to a sequencing ladder generated using the same primer. Minor and major start sites were detected by primer extension, corresponding closely to the 5' ends of sequenced 5' RACE products.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Supplemental Experimental Procedures

The 5′ portion of the RPD2a mRNA sequence was amplified by 5′ RACE (rapid amplification of cDNA ends) using Invitrogen's GeneRacer kit with nested-PCR primers 5′-CGGACCTGAAGGAGACTGTCCATG-3′ and 5′-TCCGAGAGGCGCACAATGAA-3′ (primers a and b, respectively in the diagram). The central region of the RPD2a mRNA sequence was amplified by reverse transcription followed by PCR (RT-PCR) using primers 5′-ATGCCAGATATGGACATTGATGTGAAGGAT-3′ and 5′-ATCAGCATAGCTTGGTGTCGAAGTTGAG -3′ (primers c and d, respectively in the figure). The resulting cDNA fragments were cloned using the TOPO TA Cloning Kit (Invitrogen) and sequenced using an ABI automated sequencer and big dye terminator technology. To verify the 5' ends determined by 5' RACE, primer extension was performed according to standard methods (Sambrook and Russell, 2001). A 30 nt antisense oligonucleotide (5′-AACGGCGGTGTCGGAGGAGTGCAGAGTAAA-3′) that was 5′ end-labeled using T4 polynucleotide kinase and [γ-^{32}P] ATP was used as the primer. The reverse transcription reaction was performed using ~1.0 ug Poly(A)$^{+}$ RNA and SuperScript RNase H^{-} reverse transcriptase (GIBCO BRL). Primer extension products were subjected to electrophoresis on a denaturing polyacrylamide sequencing gel alongside sequencing reactions generated using the same end-labeled primer. The resulting gel was vacuum dried onto filter paper and exposed to X-ray film.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

IV. Supporting Data for Cytological Observations

Table S3. Cytological Changes in *rpd2* Mutants

Chromocenters (CCs)

Genotype	Number of cells analyzed	Patterns observed	
		6-10 large, diffuse CC's	≤4 CCs
Wild-type	80	93%	7%
rpd2 double mutant	120	34%	66% (>20 small DAPI foci)

χ=68.56, p<0.001

NORs

Genotype	Number of cells analyzed	Number of FISH signals per nucleus				
		1	2	3	4	>4
Wild-type	60	0%	36%	25%	39%	0%
rpd2 double mutant	46	0%	19%	30%	28%	23%

χ=17.95, p<0.001

5S rRNA genes

Genotype	Number of cells analyzed	Patterns observed	
		Substantial colocalizion with centromeres	Substantial dispersal away from centromeres
Wild-type	65	69%	31%
rpd2 double mutant	72	42%	58%

χ=10.5, p=0.0012

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

Supplemental References

Bergsland, K.J., and Haselkorn, R. (1991). Evolutionary relationships among eubacteria, cyanobacteria, and chloroplasts: evidence from the rpoC1 gene of Anabaena sp. strain PCC 7120. J Bacteriol. *173*, 3446–3455.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. Science *292*, 1863–1876.

Jokerst, R.S., Weeks, J.R., Zehring, W.A., and Greenleaf, A.L. (1989). Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*. Mol. Gen. Genet. *215*, 266–275.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.P., Lottspeich, F., Garrett, R.A., and Zillig, W. (1989). Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc. Nat. Acad. Sci. USA *86*, 4569–4573.

Sambrook, J., and Russell, D.R. (2001). Molecular Cloning: A Laboratory Manuel, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schneider, G.J., and Hasekorn, R. (1988). RNA polymerase subunit homology among cyanobacteria, other eubacteria and archaebacteria. J. Bacteriol. *170*, 4136–4140.

Sweetser, D., Nonet, M., and Young, R.A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc. Nat. Acad. Sci. USA *84*, 1192–1196.

Wolfe, K.H., Gouy, M., Yang, Y.W., Sharp, P.M., and Li, W.H. (1989). Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Nat. Acad. Sci USA *86*, 6201–6205.

Cell Immediate Early Publication
Copyright ©2005 by Cell Press
Published Online February 10, 2005
DOI: 10.1016/S0092867405001510

CHAPTER THREE

SUBUNIT COMPOSITIONS OF THE RNA-SILENCING ENZYMES POL IV AND POL V REVEAL THEIR ORIGINS AS SPECIALIZED FORMS OF RNA POLYMERASE II

Published in *Molecular Cell*, vol. 33;2 192-203

My contributions to this work:

I designed and performed all experiments in the main text and supplemental materials including the purification of Pol IV and Pol V, the subsequent confirmation of associated subunits by co-immunoprecipitation, and the genetic analysis of Pol IV and Pol V subunit mutants. The western blot in Figure 2B to demonstrate the specificity of the NRPE1 antibody was performed by Jeremy R. Haag. Andrzej T. Wierzbicki created the *NRPD1-FLAG-biotin nrpd1-3* and *NRPE1-FLAG-biotin nrpe1-1*1 lines that were used to purify NRPD1 and NRPE1 complexes, respectively. The mass spectrometry analysis of the affinity purified complexes was performed by Carrie D. Nicora and Angela D. Norbeck under the leadership of Ljiljana Pasa-Tolic at the EMSL, which is sponsored by the DOE and is part of Pacific Northwest National Laboratories. Tom Guilfoyle and Gretchen Hagen provided the NRPB5 and NRPE5 antibodies used in Figure 4B-C. Jian-Kang Zhu provided evidence that the NRPE4/NRPD4 gene exists in Arabidopsis. This insight led me to discover the NRPD4/NRPE4 protein by BLAST searching and in our mass spectrometry dataset, where NRPD4/NRPE4 was annotated as a protein of unknown function.

I wrote the grant that was accepted by EMSL and the Department of Energy to fund the mass spectrometry analysis, with assistance by Craig Pikaard. I wrote the paper and edited the paper, with assistance by Craig Pikaard. I contributed significantly to the intellectual value of the paper, with assistance by Craig Pikaard.

Cell PRESS

Molecular Cell
Article

Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II

Thomas S. Ream,[1] Jeremy R. Haag,[1] Andrzej T. Wierzbicki,[1] Carrie D. Nicora,[2] Angela D. Norbeck,[2] Jian-Kang Zhu,[3] Gretchen Hagen,[4] Thomas J. Guilfoyle,[4] Ljiljana Paša-Tolić,[2] and Craig S. Pikaard[1,*]
[1]Biology Department, Washington University, St. Louis, MO 63130, USA
[2]Pacific Northwest National Laboratory, Richland, WA 99352, USA
[3]Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA
[4]Department of Biochemistry, University of Missouri, Columbia, MO 65211, USA
*Correspondence: pikaard@biology2.wustl.edu
DOI 10.1016/j.molcel.2008.12.015

SUMMARY

In addition to RNA polymerases I, II, and III, the essential RNA polymerases present in all eukaryotes, plants have two additional nuclear RNA polymerases, abbreviated as Pol IV and Pol V, that play nonredundant roles in siRNA-directed DNA methylation and gene silencing. We show that *Arabidopsis* Pol IV and Pol V are composed of subunits that are paralogous or identical to the 12 subunits of Pol II. Four subunits of Pol IV are distinct from their Pol II paralogs, six subunits of Pol V are distinct from their Pol II paralogs, and four subunits differ between Pol IV and Pol V. Importantly, the subunit differences occur in key positions relative to the template entry and RNA exit paths. Our findings support the hypothesis that Pol IV and Pol V are Pol II-like enzymes that evolved specialized roles in the production of noncoding transcripts for RNA silencing and genome defense.

INTRODUCTION

In bacteria and Archaea, a single multisubunit RNA polymerase transcribes genomic DNA into RNA. By contrast, eukaryotes have three essential nuclear DNA-dependent RNA polymerases that perform distinct functions. For instance, 45S ribosomal RNA (rRNA) genes are transcribed by RNA polymerase I (Pol I), mRNAs are transcribed by RNA polymerase II (Pol II), and tRNAs and 5S rRNA are transcribed by RNA polymerase III (Pol III) (Grummt, 2003; Schramm and Hernandez, 2002; Woychik and Hampsey, 2002).

Bacterial DNA-dependent RNA polymerase (RNAP) is composed of only four different proteins (β′, β, ω, α; with two molecules of α in the core enzyme), but archaeal RNAP and eukaryotic Pol I, II, and III are more complex (Cramer et al., 2001; Darst et al., 1998; Hirata et al., 2008). Archaea have a fundamental subunit number of 10, with the caveat that the two largest subunits are generally split into two genes (Werner, 2007). Pol I, II, and III have 12–17 subunits that include homologs of archaeal polymerase subunits, suggesting their functional diversification from an archaeal progenitor. The crystal structures of bacterial, archaeal, and eukaryotic Pol II are fundamentally similar (Cramer et al., 2001; Darst et al., 1998; Hirata et al., 2008). In each case, the largest and second-largest subunits, corresponding to the β′ and β subunits of *E. coli* RNAP, respectively, are the catalytic subunits that interact to form the DNA entry and exit channels, the active site, and the RNA exit channel.

Sequencing of the *Arabidopsis thaliana* genome revealed genes for the expected catalytic subunits of Pol I, II, and III but unexpectedly revealed two atypical largest subunit genes and two atypical second-largest subunit genes (reviewed in Pikaard et al., 2008). Moreover, five subunits of Pol I, II, and III that are typically encoded by single genes in yeast and mammals, namely *RPB5*, *RPB6*, *RPB8*, *RPB10*, and *RPB12* (named according to their discovery as Pol II subunits; aka *R*NA *P*olymerase *B*) (Cramer, 2002; Werner, 2007), are encoded by multigene families in *Arabidopsis*, as are the Pol II-specific subunits *RPB3*, *RPB4*, *RPB7*, and *RPB9*. The functional significance of the extensive subunit diversity in plants is unclear.

The genes encoding the atypical largest and second-largest polymerase subunits in *Arabidopsis* are not essential for viability (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005), unlike their Pol I, II, or III counterparts (Onodera et al., 2008). However, the atypical catalytic subunits are nuclear proteins (Onodera et al., 2005; Pontes et al., 2006) required for siRNA-directed DNA methylation and silencing of retrotransposons, endogenous repeats, and transgenes (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). The atypical catalytic subunit genes also play roles in the short-range or long-distance spread of RNA-silencing signals, responses to biotic and abiotic stresses, and the control of flowering time (Borsani et al., 2005; Brosnan et al., 2007; Dunoyer et al., 2007; Katiyar-Agarwal et al., 2007; Pontier et al., 2005; Smith et al., 2007). The atypical largest subunit genes are *NRPD1* and *NRPE1*. NRPD1 (formerly NRPD1a) is the largest subunit of Nuclear RNA polymerase IV (Pol IV; formerly Pol IVa) (Herr et al., 2005; Onodera et al., 2005), whereas NRPE1 (formerly NRPD1b) is the largest subunit of Pol V (formerly Pol IVb) (Kanno et al., 2005; Pontier

 ©2009 Elsevier Inc.

et al., 2005). The second-largest subunits of Pol IV and Pol V are encoded by the same gene, designated by the synonymous names *NRPD2a* (*NRPD2* for simplicity) or *NRPE2* (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). Pol IV and Pol V are functionally distinct, with Pol IV required for siRNA production and Pol V generating noncoding transcripts at target loci (Wierzbicki et al., 2008). Our current model is that siRNAs bind to Pol V nascent transcripts to bring the silencing machinery to the vicinity of the chromatin at target loci (Wierzbicki et al., 2008).

Aside from their largest and second-largest subunits, the subunit compositions of Pol IV and Pol V are unknown. Here, we show that Pol IV and Pol V have subunit compositions characteristic of Pol II but make differential use of RPB3, RPB4, RPB5, and RPB7 family variants in addition to having distinct catalytic subunits. Collectively, our results support the hypothesis that Pol IV and Pol V are RNA Pol II derivatives whose molecular niche is the production of noncoding transcripts for RNA-mediated silencing.

RESULTS

Identification of Pol IV, V, and II Subunits Using LC-MS/MS

To affinity purify Pol IV and Pol V from *Arabidopsis thaliana*, we engineered full-length *NRPD1 (NRPD1a)* and *NRPE1 (NRPD1b)* genomic clones, including their promoter regions and complete sets of introns and exons, adding a FLAG epitope tag to the protein's C terminus. The transgenes rescue the loss of RNA-directed DNA methylation in their respective null mutants (*nrpd1a-3* or *nrpd1b-11*), indicating that the recombinant proteins are functional (Pontes et al., 2006). NRPD1-FLAG and NRPE1-FLAG, and their respective associated subunits, were affinity purified on anti-FLAG resin, and tryptic peptides were identified by using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). For both Pol IV and Pol V, their two known catalytic subunits were detected, as expected. However, in each case, ten additional previously unknown subunits were identified, corresponding to the ten noncatalytic subunits of yeast RNA Pol II: RPB3, RPB4, RPB5, RPB6, RPB7, RPB8, RPB9, RPB10, RPB11, and RPB12 (Figure 1; see Table S1 and Figures S1 and S2, available online). The pairs of catalytic subunits specific to RNA Pol I, II, or III were not detected in Pol IV or Pol V samples, ruling out copurification of these polymerases as an explanation for the noncatalytic subunits detected in affinity-purified Pol IV or Pol V. Likewise, coimmunoprecipitation (coIP) data show that Pol IV and Pol V do not associate with each other or with Pol I, II, or III (Figure 2A).

For Pol V, peptide sequence data typically allowed unambiguous identification of subunits that are members of protein families (see Figure S1 for peptide coverage maps and Figures S4–S12 for family alignments). An exception was the RPB8 family, for which the sole peptide identified matched both variants, which are 96% identical. Two RPB3-related variants that are 88% identical are present in *Arabidopsis*, and both proteins are detected in Pol V, resulting in their designation as NRPE3a and NRPE3b (Figure 1, Figure 3A). The single RPB11 subunit encoded by the *Arabidopsis* genome was also detected; hence we refer to this protein as NRPE11 (Figure 1). Of six homologs of RPB5 in the genome, only one (NRPE5) is detected in Pol V (Figure 1, Figure S5). Two RPB9-like subunits were identified in Pol V (Figures 1 and 2D). These proteins, designated NRPE9a and NRPE9b, are 92% identical. There are four RPB7 homologs in *Arabidopsis*, only one of which is detected in Pol V, NRPE7. One of two RPB4-like subunits (NRPE4), one of two RPB10-like subunits (NRPE10), one of two RPB12-like subunits (NRPE12), and one of two RPB6-like subunits (NRPE6a) were also detected in Pol V (Figure 1).

Analysis of Pol IV's subunit composition revealed similarities and differences compared to Pol V (Figure 1, Figure S2). As with Pol V, peptides for the single RPB11-like subunit were identified. In the context of Pol IV, we refer to this protein as NRPD11; in the context of Pol V, we refer to this same protein as NRPE11. Similar nomenclature rules were adopted for other subunits shared by more than one polymerase (see Figure 1 for synonyms). NRPD4, NRPD6a, NRPD8b, and NRPD10 subunits were unambiguously identified (Figure 1). Similar to Pol V, both RPB3-like variants were detected in Pol IV, but one is predominant (NRPD3; see Figure 1). Interestingly, the RPB5-like subunit of Pol IV, NRPD5, is identical to the previously identified NRPB5 subunit of Pol II but differs from the NRPE5 subunit of Pol V (Figure 1) (Larkin et al., 1999). The major NRPD7 subunit detected in Pol IV is 62% identical to the Pol V NRPE7 subunit, but low-level peptide sequence coverage for the NRPE7 subunit was detected as well. The Pol IV NRPD9b subunit corresponds to NRPE9b detected in Pol V (Figures 1 and 2D).

The significant number of Pol II-like subunits in Pol IV and Pol V raised questions concerning the relative similarities of Pol II, Pol IV, and Pol V. Therefore, we affinity purified *Arabidopsis* Pol II by exploiting epitope-tagged NRPB2 (*NRPB2-FLAG*) expressed from a transgene that rescues the *nrpb2-1* null mutant (Onodera et al., 2008). LC/MS-MS revealed 12 subunits orthologous to their 12 yeast Pol II counterparts, with no contaminating subunits specific to Pol I, III, IV, or V (Figure 1, Figure S3). The same RPB10, RPB11, and RPB12 family subunits found in Pol IV and/or Pol V are present in Pol II (Figure 1). Sequenced peptide coverage for the RPB6, RPB8, and RPB9-like subunits in the Pol II dataset revealed that each of the two genes for these subunits encodes a subunit incorporated into Pol II (Figure S3), suggesting that the genes are redundant. A single RPB3-like subunit, NRPB3, is predominant in Pol II, consistent with a previous report (Ulmasov et al., 1996). However, peptides corresponding to the NRPE3b subunit were also detected at low frequency. The single RPB5 subunit identified in Pol II corresponds to the expected subunit based on a previous study (Larkin et al., 1999) and is identical to the NRPD5 subunit of Pol IV but distinct from the NRPE5 subunit of Pol V. Pol II also makes use of RPB4 and RPB7 variants that are distinct from the corresponding Pol IV and Pol V subunits. These NRPB4 and NRPB7 subunits correspond to subunits previously shown to associate with Pol II (Larkin and Guilfoyle, 1998).

Immunological Confirmation of Subunit Associations

To test subunit associations with all five nuclear RNA polymerases, we exploited *Arabidopsis* lines expressing FLAG-tagged

Function	Bacteria	Archaea	Sc Pol II	At Homologs	At Pol II	At Pol IV	At Pol V	Names/Synonyms
Catalytic	β'	RPOA' RPOA"	RPB1	At4g35800	59			NRPB1
				At1g63020		58		NRPD1
				At2g40030			74	NRPE1
	β	RPOB' RPOB"	RPB2	At4g21710	63			NRPB2
				At3g23780		18	37	NRPD2/NRPE2
Assembly	α	RPOD	RPB3	At2g15430	57	28	45	NRPB3/NRPD3/NRPE3a
				At2g15400	4	4	41	NRPE3b
	α	RPOL	RPB11	At3g52090	75	56	68	NRPB11/NRPD11/NRPE11
		RPON	RPB10	At1g11475	55	54	55	NRPB10/NRPD10/NRPE10
				At1g61700				NRPB10-like
		RPOP	RPB12	At5g41010	16	16	16	NRPB12/NRPD12/NRPE12
				At1g53690				NRPB12-like
Auxillary	ω	RPOK	RPB6	At5g51940	15	15	15	NRPB6a/NRPD6a/NRPE6a
				At2g04630	15	*	*	NRPB6b/NRPE6b
		RPOG	RPB8	At1g54250	30	*	*	NRPB8a/NRPE8a
				At3g59600	30	18	*	NRPB8b/NRPD8b/NRPE8b
		RPOH	RPB5	At3g22320	63	15		NRPB5/NRPD5
				At3g57080			39	NRPE5
				At5g57980				NRPB5-like
				At2g41340				NRPE5-like
				At3g54490				NRPE5-like
		RPOF	RPB4	At5g09920	61			NRPB4
				At4g15950		13	8	NRPD4/NRPE4
		RPOE	RPB7	At5g59180	51			NRPB7
				At4g14660		9	33	NRPE7
				At3g22900		52		NRPD7
				At4g14520				NRPB7-like
		TFS/RPOX	RPB9	At3g16980	22		22	NRPB9a/NRPE9a
				At4g16265	28	22	22	NRPB9b/NRPD9b/NRPE9b

Figure 1. Relationships of *Arabidopsis* Pol II, IV, and V Subunits to *E. coli*, Archaeal, and Yeast RNA Pol II Subunits

Numbers indicate percent protein coverage represented by peptides unique to that protein. "*" indicates that all peptides match both closely related proteins. Unshaded numbers represent alternate subunits detected at trace levels relative to the predominant subunit.

Pol I, II, and III second-largest subunits (NRPA2-FLAG, NRPB2-FLAG, or NRPC2-FLAG) or FLAG-tagged Pol IV and Pol V largest subunits (NRPD1-FLAG, NRPE1-FLAG), each expressed from trangenes that rescue corresponding null mutants (Onodera et al., 2008; Pontes et al., 2006). Plants expressing FLAG-tagged genomic clones of *NRPE6a*, *NRPE8b*, *NRPE10*, or *NRPE11* or an *NRPE5* cDNA were also engineered. Each recombinant protein could be immunoprecipitated from transgenic plants and detected by immunoblotting using anti-FLAG antibody (Figure 2A). Probing immunoblots with antibodies for NRPE1 and NRPE2 (Onodera et al., 2005) revealed that these Pol V catalytic subunits are present in NRPE1, NRPE6a, NRPE8b, NRPE10, NRPE11, and NRPE5 immunoprecipitates (Figure 2A; see also the anti-NRPE1 specificity control in Figure 2B), consistent with the detection of all of these subunits in Pol V (Figure 1). Controls show that NRPE2 and NRPE1 do not coimmunoprecipitate with Pol I, II, or III; that NRPE1 does not coimmunoprecipitate with Pol IV; and that NRPE2/NRPD2 is present in Pol IV and Pol V, as expected. The anti-NRPE1 antibody consistently reveals multiple NRPE1 isoforms (Figures 2A and 2B); whether these are degradation, posttranslational modification, or alternative splicing products is unclear.

To test whether NRPE5, NRPE6a, NRPE8b, NRPE10a, and NRPE11 subunits are shared by Pol I, II, and/or III, we used an anti-peptide antibody recognizing an invariant sequence in the Pol I, II, and III second-largest subunits (Onodera et al., 2005); this antibody fails to crossreact with NRPE2/NRPD2 due to a single amino acid substitution. In NRPE6a, NRPE8b, NRPE10, and NRPE11 immunoprecipitated fractions, Pol I, II, or III second-largest subunits are detected, consistent with the LC-MS/MS analysis of Pol II (Figures 1 and 2A). In yeast, RPB6, RPB8, and RPB10 are common to Pol I, II, and III, but RPB11 is Pol II specific. Second-largest subunits of Pol I, II, or III do not coimmunoprecipitate with FLAG-NRPE5, showing that NRPE5 is not a subunit of the essential polymerases (Figure 2A).

The LC-MS/MS data indicate that either of the two RPB8 homologs associate with Pol V. CoIP analysis confirms that NRPE8a or NRPE8b will coimmunoprecipitate with the Pol V catalytic subunits (Figures 2A and 2E). Although LC-MS/MS identified only one RPB6 variant (NRPE6a), its paralog (NRPE6b)

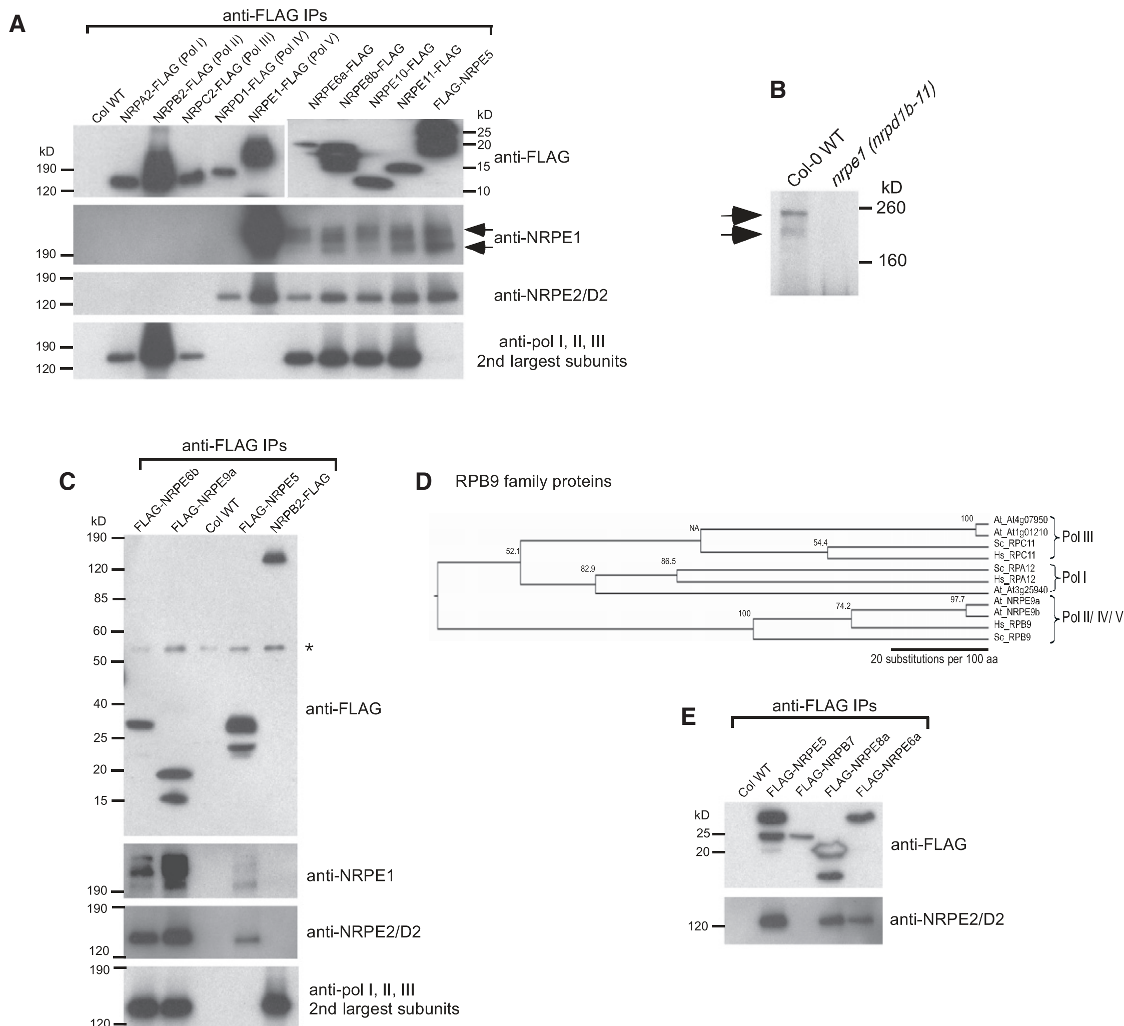


Figure 2. Verification of Pol V Subunit Associations

(A) Pol V includes subunits shared with other polymerases as well as a unique RPB5 family variant. Pol I, II, III, IV, and V were immunoprecipitated by virtue of FLAG-tagged catalytic subunits alongside NRPE6a, NRPE8b, NRPE10, NRPE11, and NRPE5 FLAG-tagged subunits. Duplicate immunoblots were probed with anti-FLAG, anti-NRPE1, anti-NRPE2/NRPD2 (abbreviated anti-NRPE2/D2), or an antibody recognizing the second-largest subunits of Pol I, II, or III. The two panels in the top row are from the same blot but focus on different size ranges.

(B) Control immunoblot showing that the multiple high-molecular-mass bands characteristic of NRPE1 are lost in an *nrpe1* null mutant (allele *nrpd1b-11*), indicating that the antibody is specific for NRPE1.

(C) NRPE6b and NRPE9a are subunits of Pol V as well as Pol I, II, or III. Immunoprecipitation and immunoblot detection was as in (A). NRPE5 and NRPB2 immunoprecipitations serve as controls for Pol V and Pol II, respectively. "*" denotes a nonspecific band detected by the anti-FLAG antibody.

(D) Phylogenetic tree based on a CLUSTALW alignment of *Arabidopsis* RPB9-like proteins with the RPB9 (Pol II), RPC11 (Pol III), and RPA12 (Pol I) subunit equivalents of yeast.

(E) NRPE8a and NRPE6a associate with Pol V. Immunoprecipitation and immunoblot detection was as in (A). NRPE5 and NRPB7 serve as controls for Pol V and Pol II, respectively.

can also associate with Pol V in vivo (Figure 2C). Both Pol II clade RPB9-like subunits (Figure 2D) were detected in Pol V by LC-MS/MS. CoIP analysis confirms that FLAG-NRPE9a associates with the Pol V NRPE1 and NRPE2 catalytic subunits in vivo (Figures 2C and 2D). NRPE6b and NRPE9a also coimmunoprecipitate the second-largest subunits of Pol I, II, or III (Figure 2C).

LC-MS/MS analysis of Pol V identified both potential RPB3 variants (Figure 3A). In confirmation of this result, HA-tagged NRPE3a and NRPE3b both coimmunoprecipitate the Pol V catalytic subunits (Figure 3B). NRPE3a, but not NRPE3b, also coimmunoprecipitates a subunit recognized by the antibody specific for Pol I, II, or III second subunits (Figure 3B); we deduce this to be the Pol II NRPB2 subunit because Pol I and Pol III use third-largest subunits distinct from RPB3. Moreover, the gene encoding NRPE3a was previously shown to encode a NRPB3 (see Figure 1) subunit present in purified Pol II (Ulmasov et al., 1996).

NRPE11, NRPE6a, NRPE8b, NRPE10, and NRPE9a all coimmunoprecipitate with the Pol IV and Pol II largest subunits

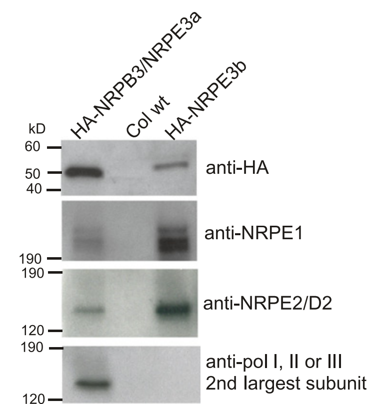


Figure 3. Pol V Utilizes a Distinct RPB3 Variant, NRPE3b, as well as an NRPE3a Variant Corresponding to the Pol II NRPB3 Subunit
(A) Alignment of the two *Arabidopsis* RPB3 family proteins with yeast RPB3.
(B) HA-tagged NRPE3a/NRPB3 and NRPE3b were immunoprecipitated and resulting immunoblots were probed using the indicated antibodies.

(Figures 1 and 4A). Upon immunoprecipitation of NRPE3b, no Pol II is detected in the immunoprecipitated fraction using an antibody recognizing the C-terminal domain (CTD) of the largest subunit. Likewise, Pol IV is detected in only trace amounts using the anti-NRPD1 antibody. We conclude that NRPE3b is used almost exclusively by Pol V (Figures 1 and 4A). In contrast, NRPB3, NRPD3, and NRPE3a are encoded by the same gene. Controls show that the NRPD1 subunit of Pol IV does not coimmunoprecipitate with Pol I, II, III, or V (Figure 4A). Likewise, the NRPB1 subunit of Pol II does not coimmunoprecipitate with Pol I, III, IV, or V (Figure 4A).

Using antibodies specific for NRPB5/NRPD5 or NRPE5 (Larkin et al., 1999), we tested their associations with FLAG-tagged Pol I, II, III, IV, or V (Figures 4B and 4C). Controls show that the NRPD2/NRPE2 subunit common to both Pol IV and Pol V is detected in NRPD1 and NRPE1 IPs, as expected, but not in Pol I, II, or III IPs (Figures 4B and 4C). NRPE5 was detected only in the NRPE1-FLAG immunoprecipitated fraction (Figure 4B), confirming that this subunit is unique to Pol V. By contrast, the NRPB5/NRPD5 subunit is detected in Pol I, II, III, and IV fractions, but not in Pol V (Figure 4C), in agreement with the LC-MS/MS data and previous studies showing that NRPB5/NRPD5 copurifies with Pol I, II, and III (Larkin et al., 1999) (Saez-Vasquez and Pikaard, 1997).

We affinity purified FLAG-tagged NRPE5 expressed in the *nrpe5* mutant background and identified the associated RNA polymerase subunits using LC-MS/MS. The results confirmed association of NRPE5 with all Pol V subunits except NRPE7 (Table S2, Figure S18), which most likely escaped detection in this experiment due to insufficient sample mass.

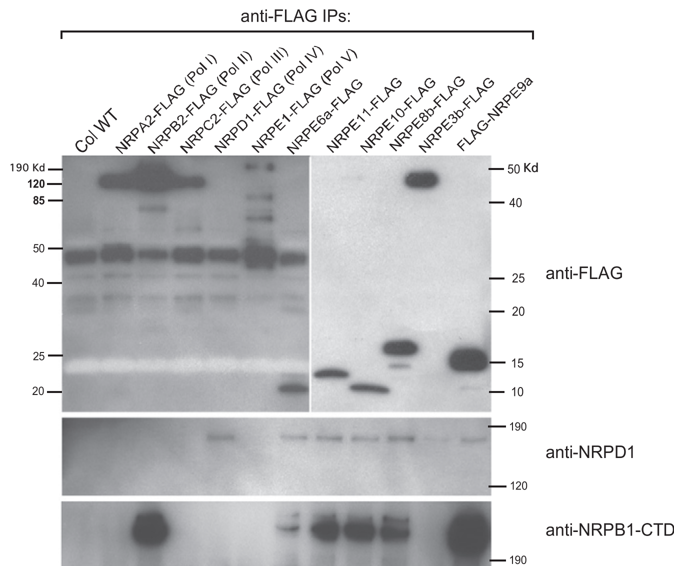
Collectively, the immunological tests of Figures 2–4 confirm the Pol V association of the NRPE1, NRPE2, NRPE3a, NRPE3b, NRPE5, NRPE6a, NRPE8, NRPE9a, NRPE10, and NRPE11 subunits detected by LC-MS/MS. Likewise, the immunological tests confirm the Pol IV associations of NRPD1, NRPD2, NRPD3, NRPD5, NRPD6a, NRPD8b, NRPD9a, NRPD10, and NRPD11. Pol IV and Pol V subunits that are shared with Pol II were also confirmed immunologically.

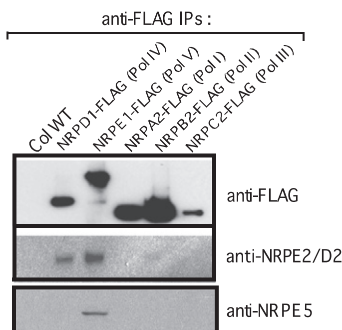
NRPE5 Is Required for DNA Methylation, siRNA Accumulation, and Gene Silencing at Pol V-Regulated Loci

Of the five full-length homologs of yeast *RPB5* in *Arabidopsis*, RT-PCR analysis shows that only *NRPB5/NRPD5* and *NRPE5* are constitutively expressed; other family members show organ-specific expression patterns (Figure 5A, Figures S5 and S13). Homozygous *nrpe5-1* mutants resulting from a T-DNA insertion (Figure 5B) are viable, as are Pol V *nrpe1* and *nrpe2* mutants. In contrast, homozygous *nrpd5-1/nrpb5-1* T-DNA insertion mutants were not recoverable due to female gametophyte lethality, as shown by reciprocal genetic crosses (Figures S14A and S14B). Female gametophyte lethality is a characteristic of Pol I, II, and III mutants, as demonstrated previously for *nrpa2*, *nrpb2*, *nrpc2*, and *nrpb12* (Onodera et al., 2008). A homozygous *nrpe11* T-DNA insertion mutant was also unrecoverable, consistent with this gene also encoding the Pol II subunit, NRPB11 (Figures S14A and S14B).

Like Pol IV and Pol V catalytic subunit mutants, *nrpe5-1* mutants lack obvious morphological phenotypes but flower later than wild-type plants under short-day conditions (Figure 5C), similar to mutants disrupting the 24 nt siRNA-directed DNA methylation pathway, including *RNA-DEPENDENT RNA POLYMERASE 2* (*RDR2*) and *DICER-LIKE 3* (*DCL3*) mutants (Chan et al., 2004; Liu et al., 2007; Pontier et al., 2005). Comparison of *nrpe5* and wild-type individuals suggests that the delay in flowering is stochastic, with some individuals showing substantial delays and others flowering at the same time as wild-type plants (Figure S15).

We tested *nrpe5-1* mutants for Pol V-dependent molecular phenotypes, including DNA hypermethylation at 5S rRNA gene clusters and at *AtSN1* and *AtSN2* retroelements. In *nrpd1* (*nrpd1a-3*), *nrpe1* (*nrpd1b-11*), and *nrpd2/nrpe2* mutants, loss





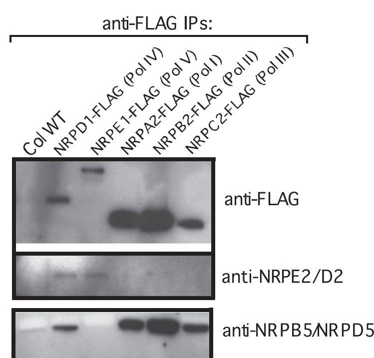


Figure 4. CoIP Tests of Pol V, IV, and II Subunit Associations

(A) Pol I, II, III, IV, and V were immunoprecipitated by virtue of FLAG-tagged catalytic subunits alongside immunoprecipitated NRPE6a, NRPE8b, NRPE9a, NRPE10, NRPE11, and NRPE3b FLAG-tagged subunits. Duplicate immunoblots were probed with anti-FLAG, anti-NRPD1 (Pol IV), or anti-NRPB1-CTD (Pol II). The two panels in the top row show different exposures of the same blot, focused on different size ranges.

(B) Pol I, II, III, IV, and V were immunoprecipitated using the indicated FLAG-tagged subunits and probed with anti-FLAG, anti-NRPE5, or anti-NRPE2/NRPD2.

(C) Immunoprecipitation and immunoblotting using the indicated antibodies were as in (B).

of methylation at 5S rDNA repeats results in increased digestion by the methylation-sensitive restriction endonucleases *Hpa*II and *Hae*III compared to wild-type plants (Figure 5D). In the *nrpe5* mutant, methylation at 5S rRNA genes is reduced compared to wild-type, but to a lesser extent than in *nrpe1* or *nrpd2/nrpe2* mutants (Figure 5D). Transformation of the *nrpe5-1* mutant with a *35S:FLAG-NRPE5* transgene restores methylation to wild-type levels, as shown in three independent transgenic lines (Figure 5D).

To test whether *nrpe5* affects DNA methylation at other Pol V-dependent loci, we examined the SINE retrotransposon families, *AtSN1* and *AtSN2* (Myouga et al., 2001). In wild-type plants, *AtSN1* and *AtSN2* elements are heavily methylated such that their DNA is not cut by *Hae*III and a PCR product can be obtained (Figures 5E and 5F). In *nrpe1* and *nrpe2/nrpd2* mutants, however, methylation is lost such that *Hae*III cuts and PCR amplification fails (Figures 5E and 5F). In *nrpe5-1*, decreased *AtSN1* and *AtSN2* methylation occurs, but not as severely as in *nrpe1* or *nrpe2/nrpd2* mutants. Nonetheless, the decreased methylation in *nrpe5-1* plants is rescued by a *35S:FLAG-NRPE5* transgene (Figures 5E and 5F).

RNA-directed DNA methylation silences *AtSN1* retroelements in wild-type plants such that loss of methylation correlates with increased *AtSN1* transcription (Hamilton et al., 2002; Herr et al., 2005; Kanno et al., 2005). *AtSN1* transcripts are barely detectable in wild-type plants but are abundant in *nrpe5* mutants, as in *nrpe1* or *nrpe2/nrpd2* mutants (Figure 5G). In the *nrpe5-1* genetic background, the *35S:FLAG-NRPE5* transgene restores *AtSN1* silencing (Figure 5G). Collectively, these results demonstrate that *NRPE5* is important for DNA methylation and silencing of *AtSN1* elements.

In the RNA-directed DNA methylation pathway, Pol IV is required for 24 nt siRNA production (Herr et al., 2005; Onodera et al., 2005) such that siRNAs are eliminated in *nrpd1* and *nrpd2* mutants (Figure 5H). In contrast, siRNAs in *nrpe1* mutants are reduced but not eliminated at 5S rRNA genes and *COPIA* elements (Figure 5H). Consistent with a Pol V mutant phenotype, siRNAs are reduced in *nrpe5* mutants relative to wild-type and are restored by the *35S:FLAG-NRPE5* transgene (Figure 5H). MicroRNA and *trans*-acting siRNA levels are unaffected in *nrpe5*, *nrpd1*, or *nrpe1* mutants, consistent with the lack of Pol IV or Pol V involvement in these pathways.

Crystallographic studies indicate that yeast RPB5 is composed of an N-terminal jaw domain and a C-terminal assembly domain separated by a short linker (Figures S5, S16, and S17A). These domains appear to be conserved in nearly all plant RPB5 homologs (Figure S16). A feature of *Arabidopsis* NRPE5, and its presumptive orthologs in other plants, is a short N-terminal extension compared to NRPB5 (Figure S16 and S17A). To test the functional significance of this N-terminal extension, we created a *35S:FLAG-ΔN-NRPE5* construct in which the extension was deleted (Figure S17A). This transgene fails to rescue *nrpe5-1* mutant phenotypes (Figures S17B–S17D). Surprisingly, immunoprecipitation of equal volumes of soluble extracts revealed that the FLAG-ΔN-NRPE5 protein

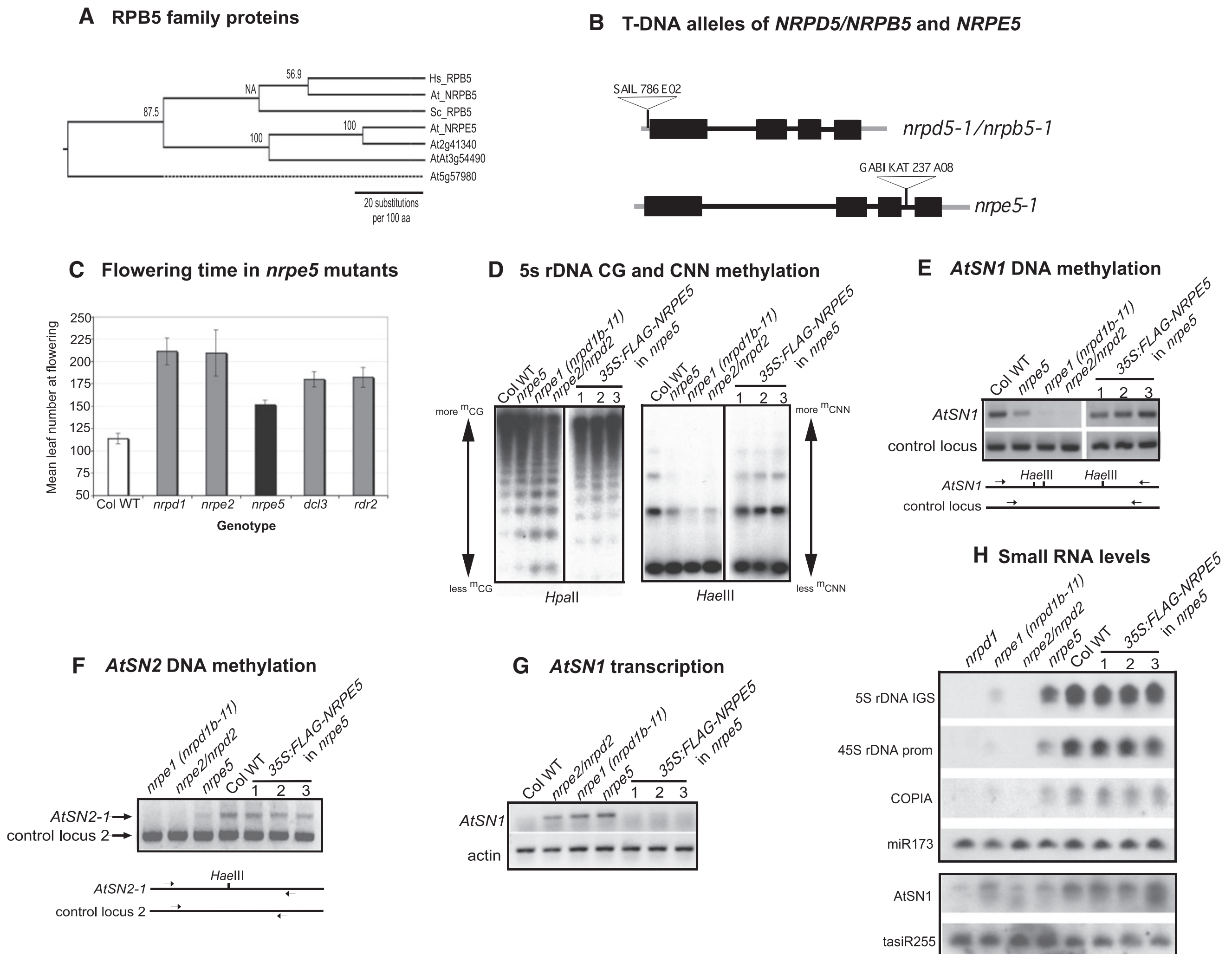


Figure 5. *nrpe5* Mutants Are Defective in RNA-Directed DNA Methylation and Retrotransposon Silencing

(A) Phylogenetic tree based on a CLUSTALW alignment of the five full-length RPB5-like proteins in *Arabidopsis* with the RPB5 subunits of yeast and human.

(B) Locations of T-DNA insertions in the *nrpb5-1/nrpd5-1* and *nrpe5-1* alleles. Black boxes represent exons, black bars represent introns, and gray bars represent 5′ and 3′UTRs.

(C) *nrpe5-1* homozygous mutant plants display a delay in flowering under short-day conditions (8 hr light, 16 hr dark). The mean (±SEM) number of rosette leaves when the floral bolt reached 10 cm is graphed. All mutants are significantly different from wild-type based on a Student's t test ($p < 0.05$).

(D) Methylation-sensitive Southern blot analyses of wild-type, *nrpe1*, *nrpe2/nrpd2*, and *nrpe5* mutants and three different *nrpe5*, *35S:FLAG-NRPE5* transgenic lines. Genomic DNA was digested with either *Hpa*II (left, reports on ^{me}CG) or *Hae*III (right, reports on ^{me}CNN) and probed for 5S rDNA repeats. Images for the *Hpa*II or *Hae*III digests are from the same exposures of the same Southern blots; the black vertical lines separate groups of lanes whose order was rearranged for clarity of presentation.

(E and F) PCR-based methylation assay of *AtSN1* and *AtSN2* family retroelements. Genomic DNA was digested with *Hae*III and subjected to PCR using *AtSN1*, *AtSN2-1*, or control primers that amplify sequences lacking *Hae*III sites (*At2g19920* in the case of [B], and an *AtSN2* family element lacking *Hae*III sites in the case of [C]). Diagrams show the relative positions of the primers flanking the *Hae*III sites.

(G) RT-PCR detection of *AtSN1* and actin transcripts.

(H) Small RNA blot analysis. Blots were probed for siRNAs corresponding to 45S or 5S rRNA genes, Copia or *AtSN1* transposons, and miRNA 173 or *trans*-acting siRNA 255.

is present at very low levels relative to full-length FLAG-NRPE5, despite similar transcript levels (Figure S17E). These data suggest that the N-terminal extension is important for the stability of the NRPE5 protein in vivo, possibly because the extended sequence facilitates Pol V-specific subunit interactions.

DISCUSSION

Origins of Pol V

Pol IV and Pol V are plant-specific enzymes that appear to have originated in an algal progenitor of land plants several hundred million years ago (Luo and Hall, 2007). Their specific involvement

Cell PRESS

in siRNA-mediated transcriptional gene silencing, which also occurs in other metazoans and fission yeast, has begged the question as to which polymerases accomplish the functions of Pol IV and Pol V in other eukaryotes. In fission yeast, Pol II transcripts traverse silenced loci, serving as binding sites for siRNAs and as templates for the sole RNA-dependent RNA polymerase, thereby generating precursors for further siRNA biogenesis (Buhler and Moazed, 2007; Buhler et al., 2006; Grewal and Elgin, 2007; Irvine et al., 2006). Several nonlethal mutations that disrupt siRNA-mediated silencing and/or siRNA accumulation in *S. pombe* have been mapped to the RPB1, RPB2, and RPB7 subunits of Pol II (Djupedal et al., 2005; Kato et al., 2005; Schramke et al., 2005). Our finding that Pol IV and V have Pol II-like subunit compositions fits the hypothesis that Pol IV and Pol V are derivatives of Pol II that evolved specialized roles in RNA silencing but no longer perform Pol II functions essential for viability, in contrast to fission yeast Pol II, which appears to accomplish all of these tasks. Presumably, the subunits of Pol IV/V that are not shared by Pol II, including NRPD1, NRPE1, NRPD2/NRPE2, NRPE3b, NRPD4/NRPE4, NRPE5, NRPD7, and NRPE7, account for Pol IV- or Pol V-specific activities. It is intriguing that most of these subunits occupy key positions with regard to the template channel and RNA exit paths (Figures 6A and 6B).

Previous analyses of Pol IV and Pol V catalytic subunits had pointed to a Pol II connection. In our initial study of Pol IV, we noted that the NRPD2/NRPE2 subunit is more closely related to the second-largest subunit of Pol II than to the corresponding subunits of Pol I or Pol III (Onodera et al., 2005). Moreover, five out of eight intron positions in the beginning of *NRPD1* and *NRPE1* match the intron positions in *NRPB1*, encoding the largest subunit of Pol II (Luo and Hall, 2007). Based on phylogenetic analyses, Luo and Hall proposed that Pol IV came into existence following a duplication of the *NRPB1* gene that generated the *NRPD1* gene. A subsequent duplication of *NRPD1* to generate *NRPE1* is proposed to have led to the evolution of Pol V after the emergence of land plants but prior to the divergence of angiosperms (flowering plants). Our finding that Pol IV utilizes the same RPB5-family subunit as Pol I, II, and III whereas Pol V uses a distinct variant (NRPE5) is consistent with the hypothesis that Pol V is more distantly related to Pol II than is Pol IV.

The fact that Pol IV and Pol V share numerous small subunits with Pol II, including NRPB3, NRPB6, NRPB8, NRPB9, NRPB10, NRPB11, and NRPB12 family subunits, can explain why alleles for these genes have not been identified in genetic screens; loss-of-function mutations in the subunits of essential polymerases cause female gametophyte lethality (Figure S14) (Onodera et al., 2008). Likewise, the use of more than one NRPE3, NRPE6, NRPE8, or NRPE9 variant by Pol IV or Pol V (Figures 6C and 1) can be expected to make identification of mutations in these genes problematic due to functional redundancies (Figure 6C).

Functions for Mystery Subunits

A number of observations in our study fill in gaps concerning the functions of RNA polymerase subunit families in *Arabidopsis*. For instance, Ulmasov et al. reported the existence of two *RPB3*-like genes in *Arabidopsis*, which they named *AtRPB36a* and *AtRPB36b* based on their predicted sizes of ~36 kD (Ulmasov et al., 1996). AtRPB36a was found in highly purified Pol II fractions (Ulmasov et al., 1996), but AtRPB36b was not, making the function of the latter variant unclear. Our study reveals that AtRPB36b is the NRPE3b subunit of Pol V. AtRPB36a (now NRPB3) and NRPB11 (formerly AtRPB13.6) in Pol II are the homologs and functional equivalents of the two α subunits (α and α') of *E. coli* RNA polymerase. Previous studies demonstrated that NRPB3 and NRPB11 copurify with Pol II in vivo and physically interact in yeast two-hybrid assays (Ulmasov et al., 1996). Interestingly, AtRPB36b/NRPE3b also interacted with NRPB11 in yeast two-hybrid assays (Ulmasov et al., 1996), which is likely to be meaningful, occurring in the context of Pol V in a manner equivalent to the interaction of NRPB3 and NRPB11 in Pol II. Interestingly, the AtRPB36a variant also associates with Pol V in vivo; therefore, this protein serves as the NRPB3 subunit of Pol II, the NRPD3 subunit of Pol IV, and one of two alternative Pol V NRPE3 subunits (NRPE3a). How these highly similar RPB3-like subunits are differentially assembled into Pol II, IV, or V is a question deserving further study.

Although peptide coverage for the NRPD4/NRPE4 subunit was low in our study, the Jian-Kang Zhu laboratory identified the *nrpd4/nrpe4* gene in a screen for defective RNA-directed DNA methylation and confirmed the Pol IV and Pol V association of the encoded protein (He, X.-J., Hsu, Y.-F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.-S., and Zhu, J.-K., unpublished data). In budding yeast, RPB4 forms a subcomplex with RPB7 that can be dissociated from the ten subunit Pol II core enzyme without abolishing Pol II catalytic activity in vitro (Cramer, 2004), although the subcomplex appears to be more stable in Pol II from plants (Larkin and Guilfoyle, 1998). In vivo, RPB7 is an essential protein in yeast, whereas RPB4 deletion mutants are temperature sensitive (McKune et al., 1993; Woychik and Young, 1989) and are impaired in transcription elongation and mRNA 3′ end processing (Runner et al., 2008; Verma-Gaur et al., 2008). It is intriguing that Pol II, IV, and V have unique RPB7-like subunits and that the NRPB4 subunit of Pol II is different from the NRPD4/NRPE4 subunits of Pol IV and Pol V. Given that the RPB4/RPB7 complex is thought to interact with the nascent RNA transcript (see Figure 6), these differences are likely to contribute to the unique functions of Pol II, IV, and V.

Previous studies had shown that one of the two consitutively expressed RPB5 family proteins is a subunit by Pol I, II, and III (Larkin et al., 1999; Saez-Vasquez and Pikaard, 1997). The function of the other variant, formerly designated AtRPB5b or AtRPB23.7, was unknown. Our study reveals that the latter protein is the NRPE5 subunit of Pol V. By contrast, the NRPD5 subunit of Pol IV is encoded by the same gene that encodes the Pol II NRPB5 subunit and the equivalent subunits of Pol I and III. As we have shown, *nrpe5-1* mutants display defects in DNA methylation, retroelement silencing, siRNA accumulation, and flowering time, similar to *nrpe1* mutants (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). However, *nrpe5-1* mutant phenotypes are typically less severe than *nrpe1* or *nrpe2/nrpd2* mutants. Because the T-DNA insertion is near the 3′ end of the gene, *nrpe5-1* may be a partially functional allele. It is also possible that other members of the multigene family are partially redundant with *NRPE5*, particularly

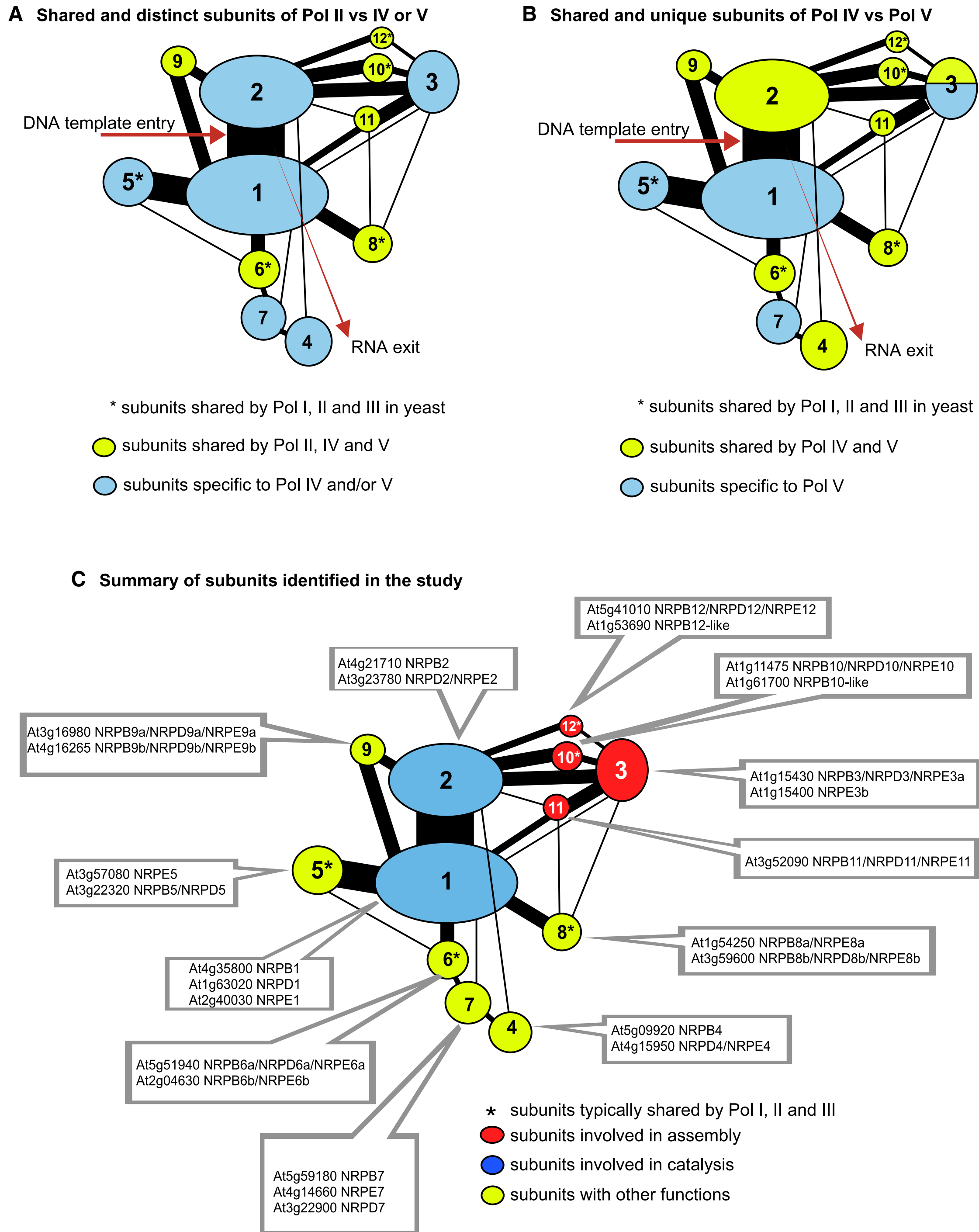


Figure 6. Comparison of RNA Polymerase Subunits in Pol II, IV, and V

(A) Subunits that are unique to Pol IV and/or Pol V compared to Pol II are shown in blue. Subunits common to Pol II, IV, and V are shown in green. The subunit interaction model is based on the yeast Pol II crystal structure (Armache et al., 2005; Cramer et al., 2001; Sampath et al., 2008). The thickness of lines connecting the subunits is proportional to the number of contacts.

(B) Subunits that are unique to Pol V are shown in blue. Subunits common to Pol IV and Pol V are shown in green. The half-blue, half-green shading of the third-largest subunit reflects the fact that Pol V uses the NRPE3b variant that is not used appreciably by Pol IV in addition to the NRPE3a/NRPD3 variant that predominates in Pol IV.

(C) Summary of the *Arabidopsis* genes that encode Pol II, IV, or V subunits.

Cell PRESS

At2g41340, which shares 70% identity with NRPE5, including the N-terminal extension that is missing in the NRPB5/NRPD5 subunit (Figure 5A and Figure S5). Consistent with this hypothesis, preliminary evidence suggests that a *nrpe5-1 At2g41340* double mutant has a more severe loss of DNA methylation phenotype than does *nrpe5-1* (data not shown). A third possibility is that NRPE5 may not be absolutely required for Pol V transcription. The failure to identify *nrpe5* alleles in genetic screens to date may stem from one or more of these reasons.

The fact that Pol V is unique in using the NRPE5 variant of the RPB5 family is likely to have functional significance. Crystal structures of yeast Pol II reveal that RPB5 interacts with RPB1 and RPB6 to form a mobile "shelf" module that stabilizes the template DNA as it enters the polymerase (Cramer et al., 2001; Gnatt et al., 2001). RPB5 also interacts with hepatitis B transcriptional activator protein X (HBx); the general transcription factor TFIIB; TIP120, a protein which facilitates recruitment of Pol II to the preinitiation complex (Cheong et al., 1995; Lin et al., 1997; Makino et al., 1999); and the yeast chromatin remodeling complex, RSC (Soutourina et al., 2006). Therefore, the differential use of the NRPD5 or NRPE5 subunits in the context of Pol IV or Pol V could mediate different template specificity, locus targeting, or transcriptional activation processes.

EXPERIMENTAL PROCEDURES

Plant Materials

A. thaliana nrpd1 (allele *nrpd1a-3*), *nrpe1* (allele *nrpd1b-11*), and *nrpd2/nrpe2* (*nrpd2a-2 nrpd2b-1*) have been described (Pontes et al., 2006). *nrpe11-1* (*nrpb11-1/nrpd11-1*) is from T-DNA line SALK_100563 (Alonso et al., 2003), *nrpd5-1/nrpb5-1* from T-DNA line SAIL_786_E02 (Sessions et al., 2002), and *nrpe5-1* from GABI-KAT T-DNA line 237A08 (Rosso et al., 2003). Primers for *nrpe11-1*, *nrpd5-1*, and *nrpe5-1* genotyping are listed in Table S3. Callus cultures were induced by germinating sterilized seeds on MS media containing Gamborg's vitamins (Sigma), 5% agargel (Sigma), 0.02 mg/L kinetin (Sigma), and 2 mg/L 2,4-dichlorophenoxyacetic acid (Sigma). Plates were incubated at 23°C. Callus frozen in liquid N_2 was stored at -80°C.

Affinity Purification of Pol IV, V, and II

Frozen callus (115–150 g) expressing FLAG-tagged *NRPE1* or *NRPD1* was ground in extraction buffer (300 mM NaCl, 20 mM Tris [pH 7.5], 5 mM $MgCl_2$, 5 mM DTT, 1 mM PMSF, and 1:100 plant protease inhibitor cocktail [Sigma]) at 4°C, filtered through two layers of Miracloth (Calbiochem), and centrifuged twice at 10,000 g, 15 min, 4°C. Pol II and NRPE5 were purified with the same protocol from 150 g of leaf tissue expressing FLAG-tagged *NRPB2* or *NRPE5*, respectively. Supernatants were incubated with anti-FLAG-M2 resin for 2–3 hr in a 15 ml tube using 30 μl of resin per 14 ml of extract. Resin was pelleted at 1000 rpm for 2 min and the supernatant incubated with fresh resin for 2–3 hr. Pooled resin was washed five times in 14 ml of extraction buffer containing 0.4% NP-40 (Sigma). Aliquots (125 μl) of resin were then mixed 2 min with 125 μl Ag/Ab Elution Buffer (Pierce) at 4°C. Resin was pelleted, and the eluted complex was pooled. Two ~500 μl batches of pooled complex were concentrated in YM-10 centricon columns (Millipore) at 4°C and desalted using Pierce 500 μl desalting columns. The final elute of ~70 μl containing ~10–50 μg of protein was subjected to LC-MS/MS.

Mass Spectrometry

Samples adjusted to 50% (v/v) 2,2,2-Trifluoroethanol (TFE) (Sigma) were sonicated 1 min at 0°C and then incubated 2 hr at 60°C with shaking at 300 rpm. Proteins were reduced with 2 mM DTT at 37°C for 1 hr, then diluted 5-fold with 50 mM ammonium bicarbonate. $CaCl_2$ (1 mM) and sequencing-grade modified porcine trypsin (Promega) was added at a 1:50 trypsin-to-protein mass ratio. After 3 hr at 37°C, samples were concentrated to ~30 μl and subjected to reversed-phase liquid chromatography (RPLC) coupled to an electrospray ionization source and LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific). Tandem mass spectra were searched against *A. thaliana* proteins using SEQUEST and filtering criteria, which provided a false discovery rate (FDR) <5%. See the Supplemental Data for details.

Cloning, Vectors, and Transgenic Lines

NRPD1 and *NRPE1* genomic clones (Pontes et al., 2006) were cloned into a Gateway-compatible vector (A.W. and C.S.P., unpublished data) that adds a C-terminal FLAG tag, 3C protease cleavage site, and biotin ligase recognition peptide. *NRPE5*, *NRPE6a*, *NRPE6b*, *NRPE8a*, *NRPE9a*, *NRPB7*, *NRPE3a*, and *NRPE3b* cDNAs were amplified by RT-PCR from poly-T primed cDNA cloned into pENTR-D-TOPO or pENTR-TEV-TOPO. cDNAs were recombined into pEarleyGate 201 (HA tag) or 202 (FLAG tag) (Earley et al., 2006). Genomic *NRPE8b*, *NRPE10*, *NRPE11*, and *NRPE6a* clones were similarly amplified by PCR and cloned into pEarleyGate 302 (FLAG tag). *NRPD1-FLAG*, *NRPE1-FLAG*, *NRPA2-FLAG*, *NRPB2-FLAG*, and *NRPC2-FLAG* transgenes were previously described (Onodera et al., 2008; Pontes et al., 2006).

Methylation Assays

5S rDNA Southern blot methylation assays and *AtSN1* PCR assays were performed using 250 ng–1 μg of DNA as in Onodera et al. (2005).

RT-PCR Analysis of *AtSN1*

For *AtSN1* transcripts, high-molecular-weight RNA was isolated from 300 mg of leaves using a miRVANA (Ambion) kit, and strand-specific RT-PCR was performed as described (Wierzbicki et al., 2008).

Small RNA Northern Blots

Inflorescence small RNA (7.5 μg) was analyzed by northern blot hybridization using *COPIA*, siR1003 (5S rRNA), 45S rRNA, miR173, and tasiR255 probes as described previously (Allen et al., 2005; Onodera et al., 2005; Pontes et al., 2006; Xie et al., 2004). Blots stripped twice with 50% formamide, 0.1× SSC, and 1% SDS at 65°C for 2 hr were reprobed to generate multiple figure panels.

Antibodies

Anti-NRPE2/NRPD2, anti-NRPB5/NRPD5, and anti-NRPE5 have been described (Larkin et al., 1999; Onodera et al., 2005). Anti-FLAG antibodies were from Sigma. Anti-NRPB1-CTD (8WG16) was purchased from Abcam. NRPE1 antibodies (Covance) recognize peptide N-CDKKNSETESDAAAWG-C. NRPD1 antibodies (Covance) recognize peptide N-CLKNGTLESGGF SENP-C. Anti-NRPA2/NRPB2/NRPC2 antibodies (US Biologicals) recognize N-CGDKFSSRHGQKG-C. Antibodies were affinity purified using immobilized peptides.

Immunoprecipitation and Immunoblotting

Leaves (2–4 g) were ground in extraction buffer (Baumberger and Baulcombe, 2005), filtered through Miracloth, and centrifuged at 10,000 g for 15 min. Supernatants were incubated 3–12 hr at 4°C with 30 μl of anti-FLAG-M2 resin (Sigma). Beads were washed three times in extraction buffer + 0.5% NP-40 (Sigma) and eluted with two bed volumes of 2× SDS sample buffer, and 5–20 μl was subjected to SDS-PAGE and transferred to Immobilon PVDF membranes (Millipore). Blots were incubated with antibodies in TBST + 5% (w/v) nonfat dried milk. Antibody dilutions were as follows: 1:250 (NRPE1), 1:500 (NRPD1), 1:2000 (NRPB1-CTD), 1:750 (NRPB5/NRPD5), 1:750 (NRPE5), 1:250 (NRPD2/NRPE2), 1:500 (anti-Pol I, II, and/or III) and 1:2000–1:10,000 (FLAG-HRP). The secondary antibody was anti-rabbit-HRP, diluted 1:5000–1:20,000; or anti-mouse-HRP, diluted 1:5000 (GE Healthcare, Sigma). Blots were washed four times for 4 min in TBST and visualized by chemiluminescence (GE Healthcare). Blots were stripped for 35 min in 25 mM glycine (pH 2.0), 1% SDS; re-equilibrated in TBST; and probed with additional antibodies.

Alignments

Sequences were aligned using ClustalW and highlighted using BOXSHADE. Construction of phylogenetic trees was performed using MegAlign. Trees are

based on ClustalW alignments of full-length proteins, and bootstrap values are based on 10,000 replicates. Dotted lines represent negative branch lengths.

SUPPLEMENTAL DATA

The Supplemental Data include Supplemental Experimental Procedures, three tables, and 18 figures and can be found with this article online at http://www.cell.com/molecular-cell/supplemental/S1097-2765(08)00858-7.

ACKNOWLEDGMENTS

T.S.R. and C.S.P. designed the study and wrote the paper. T.J.G. and G.H. contributed antibodies. J.R.H. generated Pol I, II, and III transgenic lines and NRPD1 and NRPE1 antibodies. A.W. made NRPD1- and NPRE1-FLAG-biotin lines. T.S.R. performed all experiments except LC-MS/MS analyses by C.D.N., A.N., and L.P.-T. at Pacific Northwest National Laboratory (PNNL). J.-K.Z. provided NRPD4/NRPE4 insights. We thank the Washington University greenhouse staff for plant care and Pikaard lab colleagues for discussions. T.S.R. and C.S.P. also thank Biology 4024 students who helped clone cDNAs: Silvano Ciani and Colin Clune (*At2g04630*), Andrew Pazandak and Kariline Bringe (*At1g54250* and *At3g16980*), Caitlin Ramsey and Colin Orr (*At5g59180*), Wan Shi and Soon Goo Lee (*At1g11475*), and Lily Momper and Charu Agrawal (*At5g51940*). Pikaard lab research is supported by National Institutes of Health (NIH) grant GM077590. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the NIH. Portions of this research were supported by the NIH National Center for Research Resources (RR18522), and the W.R. Wiley Environmental Molecular Science Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at PNNL. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RL01830.

Received: September 16, 2008
Revised: December 8, 2008
Accepted: December 10, 2008
Published online: December 24, 2008

REFERENCES

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell *121*, 207–221.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301*, 653–657.

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. (2005). Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. J. Biol. Chem. *280*, 7131–7134.

Baumberger, N., and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA *102*, 11928–11933.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in *Arabidopsis*. Cell *123*, 1279–1291.

Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M., and Carroll, B.J. (2007). Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. Proc. Natl. Acad. Sci. USA *104*, 14741–14746.

Buhler, M., and Moazed, D. (2007). Transcription and RNAi in heterochromatic gene silencing. Nat. Struct. Mol. Biol. *14*, 1041–1048.

Buhler, M., Verdel, A., and Moazed, D. (2006). Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. Cell *125*, 873–886.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Cheong, J.H., Yi, M., Lin, Y., and Murakami, S. (1995). Human RPB5, a subunit shared by eukaryotic nuclear RNA polymerases, binds human hepatitis B virus X protein and may play a role in X transactivation. EMBO J. *14*, 143–150.

Cramer, P. (2002). Multisubunit RNA polymerases. Curr. Opin. Struct. Biol. *12*, 89–97.

Cramer, P. (2004). RNA polymerase II structure: from core to functional complexes. Curr. Opin. Genet. Dev. *14*, 218–226.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Darst, S.A., Polyakov, A., Richter, C., and Zhang, G. (1998). Insights into Escherichia coli RNA polymerase structure from a combination of x-ray and electron crystallography. J. Struct. Biol. *124*, 115–122.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. (2005). RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. Genes Dev. *19*, 2301–2306.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. (2007). Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. Nat. Genet. *39*, 848–856.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. *45*, 616–629.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876–1882.

Grewal, S.I., and Elgin, S.C. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399–406.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. Genes Dev. *17*, 1691–1702.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. *21*, 4671–4679.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118–120.

Hirata, A., Klein, B.J., and Murakami, K.S. (2008). The X-ray crystal structure of RNA polymerase from Archaea. Nature *451*, 851–854.

Irvine, D.V., Zaratiegui, M., Tolia, N.H., Goto, D.B., Chitwood, D.H., Vaughn, M.W., Joshua-Tor, L., and Martienssen, R.A. (2006). Argonaute slicing is required for heterochromatic silencing and spreading. Science *313*, 1134–1137.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat. Genet. *37*, 761–765.

Katiyar-Agarwal, S., Gao, S., Vivian-Smith, A., and Jin, H. (2007). A novel class of bacteria-induced small RNAs in Arabidopsis. Genes Dev. *21*, 3123–3134.

Kato, H., Goto, D.B., Martienssen, R.A., Urano, T., Furukawa, K., and Murakami, Y. (2005). RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science *309*, 467–469.

Larkin, R.M., and Guilfoyle, T.J. (1998). Two small subunits in Arabidopsis RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. J. Biol. Chem. *273*, 5631–5637.

Larkin, R.M., Hagen, G., and Guilfoyle, T.J. (1999). Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. Gene *231*, 41–47.

Lin, Y., Nomura, T., Cheong, J., Dorjsuren, D., Iida, K., and Murakami, S. (1997). Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. J. Biol. Chem. *272*, 7132–7139.

Liu, F., Quesada, V., Crevillen, P., Baurle, I., Swiezewski, S., and Dean, C. (2007). The *Arabidopsis* RNA-binding protein FCA requires a lysine-specific demethylase 1 homolog to downregulate FLC. Mol. Cell *28*, 398–407.

Luo, J., and Hall, B.D. (2007). A multistep process gave rise to RNA polymerase IV of land plants. J. Mol. Evol. *64*, 101–112.

Makino, Y., Yogosawa, S., Kayukawa, K., Coin, F., Egly, J.M., Wang, Z., Roeder, R.G., Yamamoto, K., Muramatsu, M., and Tamura, T. (1999). TATA-binding protein-interacting protein 120, TIP120, stimulates three classes of eukaryotic transcription via a unique mechanism. Mol. Cell. Biol. *19*, 7951–7960.

McKune, K., Richards, K.L., Edwards, A.M., Young, R.A., and Woychik, N.A. (1993). RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. Yeast *9*, 295–299.

Myouga, F., Tsuchimoto, S., Noma, K., Ohtsubo, H., and Ohtsubo, E. (2001). Identification and structural analysis of SINE elements in the Arabidopsis thaliana genome. Genes Genet. Syst. *76*, 169–179.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613–622.

Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008). Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. Genetics *180*, 207–218.

Pikaard, C.S., Haag, J.R., Ream, T., and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. Trends Plant Sci. *13*, 390–397.

Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell *126*, 79–92.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev. *19*, 2030–2040.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. *53*, 247–259.

Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3′ processing factors. Mol. Cell. Biol. *28*, 1883–1891.

Saez-Vasquez, J., and Pikaard, C.S. (1997). Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. Proc. Natl. Acad. Sci. USA *94*, 11869–11874.

Sampath, V., Balakrishnan, B., Verma-Gaur, J., Onesti, S., and Sadhale, P.P. (2008). Unstructured N terminus of the RNA polymerase II subunit Rpb4 contributes to the interaction of Rpb4.Rpb7 subcomplex with the core RNA polymerase II of Saccharomyces cerevisiae. J. Biol. Chem. *283*, 3923–3931.

Schramke, V., Sheedy, D.M., Denli, A.M., Bonila, C., Ekwall, K., Hannon, G.J., and Allshire, R.C. (2005). RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription. Nature *435*, 1275–1279.

Schramm, L., and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. Genes Dev. *16*, 2593–2620.

Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., et al. (2002). A high-throughput Arabidopsis reverse genetics system. Plant Cell *14*, 2985–2994.

Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. (2007). An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. Plant Cell *19*, 1507–1521.

Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., et al. (2006). Rsc4 connects the chromatin remodeler RSC to RNA polymerases. Mol. Cell. Biol. *26*, 4920–4933.

Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. (1996). Association between 36- and 13.6-kDa alpha-like subunits of Arabidopsis thaliana RNA polymerase II. J. Biol. Chem. *271*, 5085–5094.

Verma-Gaur, J., Rao, S.N., Taya, T., and Sadhale, P. (2008). Genomewide recruitment analysis of Rpb4, a subunit of polymerase II in Saccharomyces cerevisiae, reveals its involvement in transcription elongation. Eukaryot. Cell *7*, 1009–1018.

Werner, F. (2007). Structure and function of archaeal RNA polymerases. Mol. Microbiol. *65*, 1395–1404.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635–648.

Woychik, N.A., and Hampsey, M. (2002). The RNA polymerase II machinery: structure illuminates function. Cell *108*, 453–463.

Woychik, N.A., and Young, R.A. (1989). RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. Mol. Cell. Biol. *9*, 2854–2859.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. *2*, E104. 10.1371/journal.pbio.0020104.

Molecular Cell, Volume *33*

Supplemental Data

Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II

Thomas S. Ream, Jeremy R. Haag, Andrzej Wierzbicki, Carrie D. Nicora, Angela Norbeck, Jian-Kang Zhu, Gretchen Hagen, Thomas J. Guilfoyle, Ljiljana Paša-Tolić, and Craig S. Pikaard

A. Supplemental Experimental Procedures

Liquid chromatography and mass spectrometry details. Two independent affinity-purified NRPE1 samples and one affinity purified NRPE5 sample were analyzed by LC-MS/MS in order to identify Pol V subunits. Affinity purified NRPD1 and NRPB2 samples were also analyzed to identify Pol IV and Pol II subunits, respectively. In each case, control samples derived from non-transgenic plants were subjected to the affinity purification procedure and analyzed by mass spectrometry.

All samples were prepared for analysis using the following procedure: a Coomassie protein assay (Pierce, Rockford, IL) was performed to determine the initial protein concentration of the sample. 2,2,2-trifluoroethanol (TFE) (Sigma, St. Louis, MO) was then added to the sample for a final concentration of 50% TFE. The sample was sonicated in an ice-water bath for 1 min. and incubated at 60ºC for 2 hours with gentle shaking at 300 rpm. The sample was then reduced with 2mM dithiothreitol (DTT) (Sigma, St. Louis, MO) with incubation at 37ºC for 1 hr with gentle shaking at 300rpm. Samples were then diluted 5-fold with 50mM ammonium bicarbonate for preparation for digestion. 1mM $CaCl_2$ and sequencing-grade modified porcine trypsin (Promega, Madison, WI) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio for 3 h at 37˚C. The sample was concentrated in a Speed Vac (ThermoSavant, Holbrook, NY) to

a volume of ~30µl and was then centrifuged at 14,000 rpm. The supernatant was removed and added to a sample vial for LC-MS/MS analysis.

Peptide samples were analyzed on a custom-built reversed-phase liquid chromatography (RPLC) system coupled via electrospray ionization (ESI) utilizing an ion funnel to a ThermoFisher Scientific LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific, San Jose, CA). Briefly, the capillary RPLC separation was performed under a constant pressure of 10,000 psi, using two ISCO (Lincoln, NE) Model 100 DM high-pressure syringe pumps and a column (60 cm × 75 µm i.d.) packed in-house (Pacific Northwest National Laboratory) with Phenomenex (Torrance, CA) Jupiter particles (C18 stationary phase, 5 µm particles, 300 Å pore size). Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B consisted of 100% acetonitrile. The RPLC system was equilibrated at 10,000 psi with 100% mobile phase A. A mobile phase selection valve was switched 50 min after injection to create a near-exponential gradient as mobile phase B displaced A in a 2.5 mL mixer. A split was used to provide an initial flow rate through the column of ~ 400 nL/min. The column was coupled to the mass spectrometer using an in-house manufactured ESI interface with homemade 20 µm i.d. chemically etched emitters. The heated capillary temperature and spray voltage were 200º C and 2.2 kV, respectively. Mass spectra were acquired for 80 min over the *m/z* range 400-2000 at a resolving power of 100K. A maximum of six data-dependent LTQ tandem mass spectra were recorded for the most intense peaks in each survey mass spectrum.

Tandem mass spectra were searched against an *Arabidopsis thaliana* protein file (The Institute for Genomic Research, TIGR 2008 http://www.tigr.org/plantProjects.shtml) containing 27,854 protein sequences after the removal of duplicates. Searching was performed using SEQUEST, allowing for a dynamic oxidation of methionine. In addition, peptide cleavage events were limited to fully tryptic sequences. For the spectra acquired in the Orbitrap, the monoisotopic masses were corrected prior to generation of the dta files used for searching using

the program DeconMSN, developed in house. Peptide sequences were considered confident if the scores passed Xcorr and delcn thresholds described by Washburn et al., which gave a False Discovery Rate (FDR) for all identified peptides of less than 5% and averaged 1.5% based on a reversed database search. Proteins with at least 2 filter passing peptides were considered confidently identified.

Generation of transgenic lines. Plants were transformed by *Agrobacterium tumefaciens* strain GV3101 harboring each transgene-bearing plasmid, using the floral dip method (Clough, S.J., and Bent, A.F. 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J *16*, 735-743). Transformants were selected by spraying with 0.05% Finale herbicide, containing 5.78% (w/v) glufosinate-ammonium (AgrEvo Environmental Health). Experiments demonstrating rescue of the *nrpe5-1* mutation by the *35S:FLAG-NRPE5* construct were performed for individual T1 transformants. Protein assays in the tagged RNA polymerase subunit lines were performed using 3- to 4-week-old pooled T2 progeny derived from single T1 plants.

Genotyping. One to three leaves were placed in a PCR tube and 125 µl of extraction buffer was added (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS). Tubes were heated using a thermocycler for 10 min. at 99˚C. Tubes were then subjected to centrifugation at 6000 x g for 10 min. The supernatant was transferred to a new PCR tube with 125 µl of isopropanol and mixed by inversion. After 15 min., the tubes were subjected to centrifugation at 6000 x g for 15 min. The supernatant was removed and the pellet was washed with 125 µl of 75% ethanol. The tubes were the spun for 5 min. at 6000 x g. The supernatant was removed and 75 µl of TE buffer was added to the pellets. The tubes were incubated in a thermocycler at 55˚C for 10 min. 2 µl of DNA was used in each 20 µl genotyping reaction with Go*Taq*Green polymerase according to the

manufacturer's instructions. Cycling conditions for genotyping *nrpb11-1/nrpd11-1/nrpe11-1*, *nrpd5-1/nrpb5-1* and *nrpe5-1* were: 94˚C 2 min. 30 sec., 36 cycles of 94˚C 30 sec., 55˚C 30 sec. and 72˚C 1 min. 15 sec. followed by a final extension of 72˚C for 7 min.

Flowering time assay. Mutants tested in the flowering time assay were all in the Columbia ecotype: *nrpd1a-3*, *dcl3-1*, *rdr2-1*. The *dcl3-1* and *rdr2-1* mutants were originally provided by Jim Carrington. Twelve to twenty plants of each genotype were grown under short-day (8 hrs. light, 16 hrs. dark) photoperiod conditions and their positions within the growth chamber were randomized every four to six days to minimize environmental influences. Flowering time was measured as the number of leaves produced in the basal rosette at the time the bolt height reached ten centimeters. P-values were derived from a two-tailed Student's-t-test of significance.

Protein alignments presented as supplemental material. Alignments were performed as described in the main methods. Sequences of RNA polymerase subunits were obtained by BLASTp searches using either *S. cerevisiae* or *A. thaliana* sequences.

RT-PCR. *NRPB5*-family first-strand cDNAs were generated using poly-T primers and PCR-amplified using gene-specific primers. *NRPE5* and *At2g41340* were amplified with the same primers and distinguished using *Spe*I (cleaves *NRPE5)* or *Hpa*II (cleaves *At2g41340*).

B. SUPPLEMENTAL DATA

Table S1. Genes whose known or predicted sequences were used in peptide coverage maps and/or protein alignments.

Supplemental Table 1					
Organism	**common name/class**	**Protein**	**Gene ID**	**accession no.**	**source**
Arabidopsis thaliana	thale cress	NRPD1	At1g63020	NP_176490	Genbank
		NRPE1	At2g40030	NP_181532	Genbank
		NRPA1	At3g57660	NP_191325	Genbank
		NRPB1	At4g35800	NP_195305	Genbank
		NRPC1	At5g60040	NP_200812	Genbank
		NRPD2a	At3g23780	NP_189020	Genbank
		NRPD2b	At3g18090	NP_188437	Genbank
		NRPA2	At1g29940	NP_564341	Genbank
		NRPB2	At4g21710	NP_193902	Genbank
		NRPC2	At5g45140	NP_199327	Genbank
		NRPB3a	At2g15430	NP_179145	Genbank
		NRPE3b	At2g15400	NP_179142	Genbank
		NRPB4	At5g09920	ABF58918	Genbank
		NRPD4/NRPE4	At4g15950	AAT71989	Genbank
		NRPB5	At3g22320	NP_188871	Genbank
		NRPE5	At3g57080	NP_191267	Genbank
		NRPB5-like	At5g57980	NP_200606	Genbank
		NRPE5-like	At2g41340	NP_181665	Genbank
		NRPE5-like	At3g54490	NP_191013	Genbank
		NRPB5-like	At3g16880	NP_188290	Genbank
		NRPB6a	At5g51940	NP_200007	Genbank
		NRPB6b	At2g04630	NP_178540	Genbank
		NRPB7	At5g59180	NP_200726	Genbank
		NRPE7	At4g14660	NP_193202	Genbank
		NRPD7	At3g22900	NP_566719	Genbank
		NRPB7-like	At4g14520	NP_849385	Genbank
		RPC25-like	At1g06790	NP_200726	Genbank
		RPA43-like	At1g75670	NP_974148	Genbank
		NRPB8a	At1g54250	NP_175827	Genbank
		NRPB8b	At3g59600	NP_191519	Genbank
		NRPB9a	At3g16980	NP_188323	Genbank
		NRPB9b	At4g16265	NP_567490	Genbank
		RPA12-like	At3g29540	ABD38906	Genbank
		RPC11-like	At4g07950	NP_192535	Genbank
		RPC11-like	At1g01210	NP_171629	Genbank
		NRPB10a	At1g11475	NP_849640	Genbank
		NRPB10-like	At1g61700	NP_176363	Genbank
		NRPB11	At3g52090	NP_190777	Genbank
		NRPB12a	At5g41010	NP_198917	Genbank
		NRPB12-like	At1g53690	NP_175773	Genbank
Homo sapiens	human	RPB5		BAA07406	Genbank
Drosophila melanogaster	fruit fly	RPB5		NP_610630	Genbank
Caenorhabditis elegans	nematode	RPB5		Q9N5K2	Genbank
Saccharomyces cerevisiae	yeast	RPB3		P16370	Genbank
		RPB4		NP_012395	Genbank
		RPB5		CAA85113	Genbank
		RPB6		CAA37382	Genbank
		RPB7		AAC60558	Genbank
		RPB8		CAA99443	Genbank
		RPB9		CAA96774	Genbank
		RPB10		CAA99425	Genbank
		RPB11		NP_014638	Genbank
		RPB12		AAB68994	Genbank
Brassica napus	rapeseed	RPB5		AAF81222	Genbank
Vitis vinifera	grape vine	RPB5a		CAO63075	Genbank
		RPB5b		CAO42914	Genbank
		RPB5c		CAO65489	Genbank
Oryza sativa	rice	RPB5a		NP_001065723	Genbank
		RPB5b		NP_001066119	Genbank
		RPB5c		EAY79909	Genbank
		RPB5d		EAZ13876	Genbank
		RPB5e		NP_001044564	Genbank
		RPB5f		CAD41325	Genbank
		RPB5g		EAZ31161	Genbank
Zea mays	maize	RPB5a		ACF87172	Genbank
		RPB5b		ACF81264	Genbank
		RPB5c		ACF85599	Genbank
Physcomitrella patens	moss	RPB5a		206246	JGI v1.1
		RPB5b		231299	JGI v1.1
		RPB5c		55574	JGI v1.1
		RPB5d		136486	JGI v1.1
Medicago truncuIata	legume	RPB5a		ABO78350	Genbank
		RPB5b		ABN07995	Genbank
		RPB5c		ABD28306	Genbank
Populus trichocarpa	black cottonwood	RPB5a		584052	JGI v1.0
		RPB5b		57931	JGI v1.0
		RPB5c		48513	JGI v1.0
Ostreococcus lucimarinus	green algae	RPB5		XP_001417617	Genbank
Chlamydomonas reinhardtii	green algae	RPB5		XP_001697601	Genbank

Table S2. Subunits of *Arabidopsis* Pol V identified by LC-MS/MS analysis of immunoprecipitated FLAG-NRPE5. Pol V subunit relationships to equivalent subunits of yeast Pol II, archaeal and bacterial RNAP are shown, as in Table 1 of the main text. Numbers denote the % of the protein covered by sequenced peptides that could only have come from the indicated protein; non-unique peptides matching related family members are excluded from the coverage calculation. Asterisks denote the fact that all sequenced peptides could be derived from either of two closely related variants.

function	Bacteria	Archaea	Sc Pol II	At homologs	NRPE5 IP	Names/synonyms
catalytic	ß'	RPOA' RPOA"	RPB1	At4g35800		NRPB1
				At1g63020		NRPD1
				At2g40030	22	NRPE1
	ß	RPOB' RPOB"	RPB2	At4g21710		NRPB2
				At3g23780	24	NRPD2/NRPE2
assembly	α	RPOD	RPB3	At2g15430	36	NRPB3/NRPD3/NRPE3a
				At2g15400	4	NRPE3b
	α	RPOL	RPB11	At3g52090	36	NRPB11/NRPD11/NRPE11
		RPON	RPB10	At1g11475	28	NRPB10/NRPD10/NRPE10
				At1g61700		NRPB10-like
		RPOP	RPB12	At5g41010	16	NRPB12/NRPD12/NRPE12
				At1g53690		NRPB12-like
auxillary	ω	RPOK	RPB6	At5g51940	*	NRPB6a/NRPD6a/NRPE6a
				At2g04630	*	NRPB6b/NRPD6b/NRPE6b
		RPOG	RPB8	At1g54250	*	NRPB8a/NRPD8a/NRPE8a
				At3g59600	*	NRPB8b/NRPD8b/NRPE8b
		RPOH	RPB5	At3g22320		NRPB5/NRPD5
				At3g57080	65	NRPE5
				At5g57980		NRPB5-like
				At2g41340		NRPE5-like
				At3g54490		NRPE5-like
		RPOF	RPB4	At5g09920		NRPB4
				At4g15950	8	NRPD4/NRPE4
		RPOE	RPB7	At5g59180		NRPB7
				At4g14660		NRPE7
				At3g22900		NRPD7
				At4g14520		NRPB7-like
		TFS/RPOX	RPB9	At3g16980	*	NRPB9a/NRPE9a
				At4g16265	*	NRPB9b/NRPD9b/NRPE9b

Table S3. List of primer sequences.

Primer	sequence	Used for:
cNRPE5-F	CACC ATG GAA GTG AAA GGG AAA GAG ACA G	cloning *NRPE5* cDNA
cNRPE5-R	TTA CCA CAC ACA TCG GAA GGC	cloning *NRPE5* cDNA
NRPA2-F	CACC GCC AAT GCT TTC GAG GAA CGG TTT	cloning *NRPA2* genomic fragment
NRPA2-R	ATC AGT TAC TCC TTC TCT ATC GCT TAA CTG AAG AGT C	cloning *NRPA2* genomic fragment
NRPB2-F	CACC TCA CTC TCC GTC TCT CTC TCT CTT	cloning *NRPB2* genomic fragment
NRPB2-R	CTG TCT GCC TTT AGC CGA TTT CAG G	cloning *NRPB2* genomic fragment
NRPC2-F	CAC CTG AAT ACA CCC TCC TTA GAG GCC A	cloning *NRPC2* genomic fragment
NRPC2-R	AGC CTC TGT GAG TTT CAG ACG C	cloning *NRPC2* genomic fragment
cNRPB3a-F	cacc ATGGACGGTGCCACATACCAAAG	cloning *NRPB3a* cDNA
cNRPB3a-R	TTA TCC TCC ACG CAT ATG GGC AC	cloning *NRPB3a* cDNA
cNRPE3b-F	cacc ATGGACGGTGTCACCTACCAAAG	cloning *NRPE3b* cDNA
cNRPE3b-R	TTA TCC TTC ACG CAT ATG GGC ACC	cloning *NRPE3b* cDNA
cNRPB6a-F	cacc ATGGCTGACGAAGATTACAACGACG	cloning *NRPB6a* cDNA
cNRPB6a-R	TTA ATC ACC ACC AAC TTG ACG TTT CC	cloning *NRPB6a* cDNA
cNRPB6b-F	cacc ATG GCT GAC GAC GAT TAC AAT GAA G	cloning *NRPB6b* cDNA
cNRPB6b-R	TTA ATC ACC ACC GAC TTG ACG TTT C	cloning *NRPB6b* cDNA
NRPB6a-F	cacc gcacaaaaactaaataatcacaacatc	cloning *NRPB6a* genomic fragment
NRPB6a-R	ATC ACC ACC AAC TTG ACG TTT C	cloning *NRPB6a* genomic fragment
cNRPB7-F	cacc ATG TTT TTC CAC ATA GTA TTG GAG CG	cloning *NRPB7* cDNA
cNRPB7-R	TTA TGC CGC TGC AGG GTC GT	cloning *NRPB7* cDNA
cNRPB8a-F	cacc ATGGCGAGCAATATCATCTTGTTCG	cloning *NRPB8a* cDNA
cNRPB8a-R	TTA CAG CTT CCT CAT GAG TAG GAA G	cloning *NRPB8a* cDNA
cNRPB8b-F	cacc ATGGCGAGCAATATTATCATGTTCG	cloning *NRPB8b* cDNA
cNRPB8b-R	TTA AAG CTT CCT CAT GAG TAG AAA GAG	cloning *NRPB8b* cDNA
cNRPE9a-F	cacc ATGAGTACTATGAAATTTTGCCGCG	cloning *NRPE9a* cDNA
cNRPE9a-R	TTA TTC TCT CCA GCG ATG ACC AC	cloning *NRPE9a* cDNA
NRPB10-F	cacc tgttctcgtaagcgtagagatcttc	cloning *NRPB10* genomic fragment
NRPB10-R	ACT GTT GTC TGA TTT CTC CAG AG	cloning *NRPB10* genomic fragment
NRPB11-F	cacc GTT GTG TCC GAA CAT ACC TCA C	cloning *NRPB11* genomic fragment
NRPB11-R	AAA CTG ATT CGA AAA CTT GGC C	cloning *NRPB11* genomic fragment
AtSN2-1 F	AGATAGTCACAATGTAAGGCATTCGTG	AtSN2/control methylation assay
AtSN2-1 R	TTGATCCTTTGTCAATGGAAGATTAC	AtSN2/control methylation assay
NRPB5a F	GAG AGG ATC TTG TTA CTC TTA AGG CTA	RT-PCR
NRPB5a R	CGA CCA GCC GTT TCA CTC GGA	RT-PCR
NRPB5c F	CTT GAA AAG AGA AGA GTT TGT TCA GAG G	RT-PCR
NRPB5c R	AAT GAA GTA GCA TCG CTT CGT C	RT-PCR
At3g54490 F	GAG GAG ACA ATG GCC GAA G	RT-PCR
At3g54490 R	CAT TGT TGG AAA TCT GAA TAT GAA GAG CA	RT-PCR
NRPE5 and At2g41340 F	TAC GAA GTC TCC GAC GAA GAT AT	RT-PCR
NRPE5 and At2g41340 R	CTC AAT GCT GAA CTT CTT GAG AAG TG	RT-PCR
At3g16880-F	GTT CTC TTT CTC TCT AGA AAC TTT TG	RT-PCR
At3g16880-R	CAC CAT GAA GAA ATA CAT AGA CCA GTT AAA ATC GGC A	RT-PCR
NRPE5 RNA F span	AAG GTC GAG ATA TTC CAG ATA ACG G	RT-PCR of *NRPE5* in *nrpe5-1*
NRPE5 RNA R span	GCG ATT CCG TGA GTT CGC CTC	RT-PCR of *NRPE5* in *nrpe5-1*
FLAG F	ATG GAC TAC AAA GAC GAT GAC GAC	RT-PCR of *FLAG-NRPE5* transgenes
NRPE5 cDNA R	CAG CCC AGT TAT GGT TTC TTG G	RT-PCR of *FLAG-NRPE5* transgenes
ΔN-NRPE5 F	CACC CTA TCG AGT GAA GAG AGT CAT AGA TAC	cloning *ΔN-NRPE5* cDNA
NRPE5-R	TTA CCA CAC ACA TCG GAA GGC	cloning *ΔN-NRPE5* cDNA
SAIL 786E02 LP	AGA GCA CAT GAA TCA GCG ACT	genotyping *nrpd5-1*
SAIL 786E02 RP	GGA GAG ATC GTC GTA GCA CTG	genotyping *nrpd5-1*
GABI KAT 237A08 LP	CTT CCC CTG CCC ATT TTT TTG CTA C	genotyping *nrpe5-1*
GABI KAT 237A08 RP	GTT TAA AGG GTC TGC TTC AAG AAG TG	genotyping *nrpe5-1*
SALK 100563 LP	GAGAGTATGGGCTGGTGATTG	genotyping *nrpe11-1*
SALK 100563 RP	AGAGCCTGTTGCTTTGAATTG	genotyping *nrpe11-1*
NRPE5 RNA 5' F	ATGGAAGTGAAAGGGAAAGAGACAG	RT-PCR of *NRPE5* in *nrpe5-1*
NRPE5 RNA 5' R	GTTCAATGGCTTTCAAGGCTTGATT	RT-PCR of *NRPE5* in *nrpe5-1*

Figure S1. Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V (NRPE1-FLAG). In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPE1 (At2g40030)

MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGAT
EPDKCEGHFGYIQLPVPIYHPAHVNELKQMLSLLCLKCLKIKKAKGTSGGLADRLLGVC
CEEASQISIKDRASDGASYLELKLPSRSRLQPGCWNFLERYGYRYGSDYTRPLLAREVKE
ILRRIPEESRKKLTAKGHIPQEGYILEYLPVPPNCLSVPEASDGFSTMSVDPSRIELKDVLK
KVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAKAARNIDMRYGVSKISDSSSS
KAWTEKMRTLFIRKGSGFSSRSVITGDAYRHVNEVGIPIEIAQRITFEERVSVHNRGYLQ
KLVDDKLCLSYTQGSTTYSLRDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHS
LQALRVYVHEDNTVKINPLMCSPLSADFDGDCVHLFYPQSLSAKAEVMELFSVEKQLLS
SHTGQLILQMGSDSLLSLRVMLERVFLDKATAQQLAMYGSLSLPPPALRKSSKSGPAWT
VFQILQLAFPERLSCKGDRFLVDGSDLLKFDFGVDAMGSIINEIVTSIFLEKGPKETLGFFD
SLQPLLMESLFAEGFSLSLEDLSMSRADMDVIHNLIIREISPMVSRLRLSYRDELQLENSIH
KVKEVAANFMLKSYSIRNLIDIKSNSAITKLVQQTGFLGLQLSDKKKFYTKTLVEDMAIF
CKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKNLMA
VLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKA
VLDSSPNSNSSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENAACTVRN
KLNKVSLKDTAVEFLVEYRKQPTISEIFGIDSCLHGHIHLNKTLLQDWNISMQDIHQKCE
DVINSLGQKKKKKATDDFKRTSLSVSECCSFRDPCGSKGSDMPCLTFSYNATDPDLERT
LDVLCNTVYPVLLEIVIKGDSRICSANIIWNSSDMTTWIRNRHASRRGEWVLDVTVEKSA
VKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRLSASVRMV
SKGVLKEHIILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKAAE
KCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTT
NADAFVSSPGFDVTEEEMAEWAESPERDSALGEPKFEDSADFQNLHDEGKPSGANWEK
SSSWDNGCSGGSEWGVSKSTGGEANPESNWEKTTNVEKEDAWSSWNTRKDAQESSKS
DSGGAWGIKTKDADADTTPNWETSPAPKDSIVPENNEPTSDVWGHKSVSDKSWDKKN
WGTESAPAAWGSTDAAVWGSSDKKNSETESDAAAWGSRDKNNSDVGSGAGVLGPWN
KKSSETESNGATWGSSDKTKSGAAAWNSWDKKNIETDSEPAAWGSQGKKNSETESGP
AAWGAWDKKKSETEPGPAGWGMGDKKNSETELGPAAMGNWDKKKSDTKSGPAAWG
STDAAAWGSSDKNNSETESDAAAWGSRNKKTSEIESGAGAWGSWGQPSPTAEDKDTN
EDDRNPWVSLKETKSREKDDKERSQWGNPAKKFPSSGGWSNGGGADWKGNRNHTPR
PPRSEDNLAPMFTATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPISDDDK
TFVLEKILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQDFSYRKSLNNY
LMKKYPDRAEEFIDKYFTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQT
QSQSPSQTRAQSPSQAQAQSPSQTQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSPSQ
TQTQSPSQTQAQAQSPSSQSPSQTQT

Notes:
1457/1976 amino acids are represented by sequenced peptides =74% coverage.
All peptides are specific to NRPE1 (NRPD1b), meaning that none are identical to any other protein, including NRPD1 (NRPD1a).

NRPE2/NRPD2 (At3g23780)
MPDMDIDVKDLEEFEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
WHARLQNMTYSARIKVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSI
PVMVKSILCKTSEKGKENCKKGDCAFDQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
WTVSFRSENKRNRFIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
HLYLFPGLQSLKKKARFLGYMVKCLLNSYAGKRKCENRDSFRNKRIELAGELLEREIRV
HLAHARRKMTRAMQKHLSGDGDLKPIEHYLDASVITNGLSRAFSTGAWSHPFRKMERV
SGVVANLGRANPLQTLIDLRRTRQQVLYTGKVGDARYPHPSHWGRVCFLSTPDGENCG
LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHKVLLNGDWVGLCAD
SESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTDAGRLLRPLLVVENLQKLKQEKPS
QYPFDHLLDHGILELIGIEEEEDCNTAWGIKQLLKEPKIYTHCELDLSFLLGVSCAVVPFA
NHDHGRRVLYQSQKHCQQAIGFSSTNPNIRCDTLSQQLFYPQKPLFKTLASECLKKEVLF
NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEKRKKMD
ELVQFGKTHSKIGKVDSLEDDGFPFIGANMSTGDIVIGRCTESGADHSIKLKHTERGIVQK
VVLSSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQKGVLGYLEEQQNFPFTIQGIVPDI
VINPHAFPSRQTPGQLLEAALSKGIACPIQKEGSSAAYTKLTRHATPFSTPGVTEITEQLH
RAGFSRWGNERVYNGRSGEMMRSMIFMGPTFYQRLVHMSEDKVKFRNTGPVHPLTRQ
PVADRKRFGGIKFGEMERDCLIAHGASANLHERLFTLSDSSQMHICRKCKTYANVIERTP
SSGRKIRGPYCRVCVSSDHVVRVYVPYGAKLLCQELFSMGITLNFDTKLC

Notes:
434/1172 amino acids represented in sequenced peptides =37% coverage.
155/1172= 13% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 24% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPE3a/NRPD3/NRPB3 (At2g15430)
MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEIEVNSS
VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCSVEFRLSSKCVTDQTLD
VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRRGQELKLRAIARKGIGKDHAKW
SPAATVTFMYEPDIIINEDMMDTLSDEEKIDLIESSPTKVFGMDPVTRQVVVVDPEAYTY
DEEVIKKAEAMGKPGLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLLKQKLDAVRLSD
DTVEADDQFGELGAHMRGG

Notes:
184/319 amino acids are represented by sequenced peptides =58% coverage
45% of the coverage corresponds to peptides that match only NRPE3a. The other 13% matches either NRPE3a or NRPE3b.

NRPE3b (At2g15400)
MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSVEFPLSAKCVTDQTLD
VTSRDLYSADPTVTPVDFTSNSSTSDSSEHKGIIIAKLRRGQELKLKALARKGIGKDHAK
WSPAATVTYMYEPDIIINEEMMNTLTDEEKIDLIESSPTKVFGIDPVTGQVVVVDPEAYT
YDEEVIKKAEAMGKPGLIEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLDAIRLSD
NTVEADDQFGELGAHMREG

Notes:
170/319 amino acids are represented by sequenced peptides = 53% coverage
131/319=41% coverage corresponds to peptides matching only NRPE3b, whereas the remaining 12% of the coverage matches either NRPE3b or NRPE3a.

NRPD4/NRPE4 (At4g15950)
MSEKGGKGLKSSLKSKDGGKDGSSTKLKKGRKIHFDQGTPPANYKILNVSSDQQPFQSS
AAKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDGIKGQLVGLSEDPSIKI
PVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAFIPS
LKTKKEVINQPLQDALEELSKLKKSE

17/205=8% coverage
All peptides are unique matches to NRPD4 only.

NRPB4 (At5g09920)
MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
HDDEAIEKMLNDLSLVKRFE

0/138=0% coverage
No peptides were identified that match this protein sequence.

NRPB5/NRPD5 (At3g22320)
MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKA
KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
FGLKRGQVVKIIRPSETAGRYVTYRYVV

0/205 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)
MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
RTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSLKPQHQVLNDEEKTTLLKKFSIEEK
QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRCVW

86/222 amino acids are represented by sequenced peptides = 39% coverage

All peptides identified correspond to peptides that match NRPE5 only and no other family member.

NRPB5 family member (At5g57980)
MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMNKVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVPVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMLTIEVFEEAELVTNITEHKLVNKYYVLDDQAKKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRSDATSLDYYTYRFAV

0/210 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (At2g41340)
MEGKGKEIVVGHSISKSSVECHKYYLARRTTMEMLRDRGYDVSDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKKISVVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLRPKHQVLNDKEKESLLKKFSIEEKQL
PRLSSKDPIVRYYGLETGQVMKVTYKDELSESHVTYRCVS

0/218 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (At3g54490)
MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKKILVVFMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTTFPFTVETFPIEDLLVNITKHIQQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVTYRCII

0/233 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)
MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence. This protein is truncated relative to the other NRPB5-like proteins and likely is a pseudogene.

NRPE6a/NRPB6a/NRPD6a (At5g51940)
MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDGSFEEWGVDELIVEDSWKRQVGGD

48/144 amino acids are represented by sequenced peptides = 33% coverage
22/144 = 15% coverage corresponds to peptides that are NRPE6a-specific, whereas the remaining 18% match either NRPE6a or NRPE6b.

NRPE6b/NRPB6b (At2g04630)
MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDMSYEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage
0/144=0% of the coverage corresponds to peptides unique to this member of the protein family; the sequenced peptide also matches an identical sequence of At5g51940.

NRPE7 (At4g14660)
MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYYVAVTTLDKI
GEGKIREHTGEVLFPVMFSGMTFKIFKGEIIHGVVHKVLKHGVFMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVVIGIKWMEVEREFQALASLEGDYLGPLSE

58/177 amino acids are represented by sequenced peptides = 33% coverage
All peptides match At4g14660 and only At4g14660.

NRPB7 (At5g59180)
MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQCVVFRPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTTSDGSVKIQKECEVRLKIIGTRVDATAIFCVGTIKDDFLGVINDPAAA

0/176 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that match this protein sequence.

NRPD7 (At3g22900)
MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVLDTEWREAEKDFMALASIDGDNLGPF

0/174 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPB7 family member (At4g14520)
MFSEVEMARDVAICAKHLNGQSPHQPILCRLLQDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFPVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLSEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE8a/NRPB8a/NRPD8a (At1g54250)
MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by the sequenced peptide = 9% coverage

0/146=0% of the coverage corresponds to peptides unique to this member of the protein family. This peptide also is an exact match to At3g59600.

NRPE8b/NRPD8b/NRPB8b (At3g59600)
MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAVGDKF
TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLQGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by the sequenced peptide = 9% coverage
0/146 = 0% of the coverage corresponds to peptides unique to this member of the protein family. This peptide is also an exact match to At1g54250.

NRPE9a/NRPD9a/NRPB9a (At3g16980)
MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEGMTLFFVCCNPNCGHRWRE

25/114 amino acids are represented by sequenced peptides = 22% coverage
25/114 = 22% coverage corresponds to peptides unique to this member of the protein family. Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from At4g16265.

NRPE9b/NRPD9b/NRPB9b (At4g16265)
MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
LSDVASDPTLPRTKAVRCAKCQHGEAVFFQATARGEEGMTLFFVCCNPNCSHRWRE

25/114 amino acids are represented by the sequenced peptide = 22% coverage
25/114 = 22% coverage corresponds to peptides unique to this member of the protein family. Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from At4g16265.

NRPE10/NRPD10/NRPB10 (At1g11475)
MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
LLNYNTLEKSDNS

50/71 amino acids are represented by sequenced peptides = 70% coverage
39/71= 55% coverage corresponds to peptides that only match this protein, whereas the remaining 15% match either At1g11475 or At1g61700.

NRPB10 family member (At1g61700)
MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
KLLNYNTMEKSDPN
11/71 amino acids are represented by the sequenced peptide = 15% coverage
0/71= 0% unique. The peptide identified for At1g61700 also matches At1g11475.

NRPE11/NRPD11/NRPB11 (At3g52090)
MNAPERYERFVVPEGTKKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
GYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLDKELDYLKNQFEAEVAKFSNQF

79/116 amino acids are represented by sequenced peptides = 68% coverage
All peptides identified match NRPE11 and only NRPE11.

NRPE12/NRPD12/NRPB12 (At5g41010)
MDPAPEPVTYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTRRVVQYEAR

8/51 amino acids are represented by the sequenced peptide = 16% coverage
The peptide is unique to this protein.

NRPB12 family member (At1g53690)
MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV
0/62 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

Figure S2. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol IV (NRPD1-FLAG). Highlighting is the same as in Fig. S1.

NRPD1 (At1g63020)
MEDDCEELQVPVGTLTSIGFSISNNNDRDKMSVLEVEAPNQVTDSRLGLPNPDSVCRTC
GSKDRKVCEGHFGVINFAYSIINPYFLKEVAALLNKICPGCKYIRKKQFQITEDQPERCRY
CTLNTGYPLMKFRVTTKEVFRRSGIVVEVNEESLMKLKKRGVLTLPPDYWSFLPQDSNI
DESCLKPTRRIITHAQVYALLLGIDQRLIKKDIPMFNSLGLTSFPVTPNGYRVTEIVHQFN
GARLIFDERTRIYKKLVGFEGNTLELSSRVMECMQYSRLFSETVSSSKDSANPYQKKSDT
PKLCGLRFMKDVLLGKRSDHTFRTVVVGDPSLKLNEIGIPESIAKRLQVSEHLNQCNKER
LVTSFVPTLLDNKEMHVRRGDRLVAIQVNDLQTGDKIFRSLMDGDTVLMNRPPSIHQHS
LIAMTVRILPTTSVVSLNPICCLPFRGDFDGDCLHGYVPQSIQAKVELDELVALDKQLINR
QNGRNLLSLGQDSLTAAYLVNVEKNCYLNRAQMQQLQMYCPFQLPPPAIIKASPSSTEP
QWTGMQLFGMLFPPGFDYTYPLNNVVVSNGELLSFSEGSAWLRDGEGNFIERLLKHDK
GKVLDIIYSAQEMLSQWLLMRGLSVSLADLYLSSDLQSRKNLTEEISYGLREAEQVCNK
QQLMVESWRDFLAVNGEDKEEDSVSDLARFCYERQKSATLSELAVSAFKDAYRDVQA
LAYRYGDQSNSFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAAWNDP
NSPLRGAKGKDSTTTESYVPYGVIENSFLTGLNPLESFVHSVTSRDSSFSGNADLPGTLSR
RLMFFMRDIYAAYDGTVRNSFGNQLVQFTYETDGPVEDITGEALGSLSACALSEAAYSA
LDQPISLLETSPLLNLKNVLECGSKKGQREQTMSLYLSEYLSKKKHGFEYGSLEIKNHLE
KLSFSEIVSTSMIIFSPSSNTKVPLSPWVCHFHISEKVLKRKQLSAESVVSSLNEQYKSRNR
ELKLDIVDLDIQNTNHCSSDDQAMKDDNVCITVTVVEASKHSVLELDAIRLVLIPFLLDS
PVKGDQGIKKVNILWTDRPKAPKRNGNHLAGELYLKVTMYGDRGKRNCWTALLETCL
PIMDMIDWGRSHPDNIRQCCSVYGIDAGRSIFVANLESAVSDTGKEILREHLLLVADSLS
VTGEFVALNAKGWSKQRQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALA
WGKVPGFGTGDQFEIIISPKVHGFTTPVDVYDLLSSTKTMRRTNSAPKSDKATVQPFGLL

HSAFLKDIKVLDGKGIPMSLLRTIFTWKNIELLSQSLKRILHSYEINELLNERDEGLVKMV
LQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFEDFSYHKCVLGATKIIAPKK
MNFYKSKYLKNGTLESGGFSENP

844/1453 amino acids are represented by sequenced peptides =58% coverage
All peptides are specific to NRPD1 (NRPD1a), meaning that none are identical to any other protein, including NRPE1 (NRPD1b).

NRPD2/NRPE2 (At3g23780)
MPDMDIDVKDLEEFEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
WHARLQNMTYSARIKVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSI
PVMVKSILCKTSEKGKENCKKGDCAFDQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
WTVSFRSENKRNRFIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
HLYLFPGLQSLKKKARFLGYMVKCLLNSYAGKRKCENRDSFRNKRIELAGELLEREIRV
HLAHARRKMTRAMQKHLSGDGDLKPIEHYLDASVITNGLSRAFSTGAWSHPFRKMERV
SGVVANLGRANPLQTLIDLRRTRQQVLYTGKVGDARYPHPSHWGRVCFLSTPDGENCG
LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHKVLLNGDWVGLCAD
SESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTDAGRLLRPLLVVENLQKLKQEKPS
QYPFDHLLDHGILELIGIEEEDCNTAWGIKQLLKEPKIYTHCELDLSFLLGVSCAVVPFA
NHDHGRRVLYQSQKHCQQAIGFSSTNPNIRCDTLSQQLFYPQKPLFKTLASECLKKEVLF
NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEKRKKMD
ELVQFGKTHSKIGKVDSLEDDGFPFIGANMSTGDIVIGRCTESGADHSIKLKHTERGIVQK
VVLSSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQKGVLGYLEEQQNFPFTIQGIVPDI
VINPHAFPSRQTPGQLLEAALSKGIACPIQKEGSSAAYTKLTRHATPFSTPGVTEITEQLH
RAGFSRWGNERVYNGRSGEMMRSMIFMGPTFYQRLVHMSEDKVKFRNTGPVHPLTRQ
PVADRKRFGGIKFGEMERDCLIAHGASANLHERLFTLSDSSQMHICRKCKTYANVIERTP
SSGRKIRGPYCRVCVSSDHVVRVYVPYGAKLLCQELFSMGITLNFDTKLC

211/1172=18% coverage
48/1172=4% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 14% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPD3/NRPE3a/NRPB3 (At2g15430)
MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEIEVNSS
VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCSVEFRLSSKCVTDQTLD
VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRRGQELKLRAIARKGIGKDHAKW
SPAATVTFMYEPDIIINEDMMDTLSDEEKIDLIESSPTKVFGMDPVTRQVVVVDPEAYTY
DEEVIKKAEAMGKPGLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLLKQKLDAVRLSD
DTVEADDQFGELGAHMRGG

101/319=32% coverage
90/319=28% coverage is accounted for by peptides unique to NRPD3. The remaining 4% of the peptides match NRPD3 as well as the NRPD3b variant.

NRPE3b (At2g15400)
MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSVEFPLSAKCVTDQTLD
VTSRDLYSADPTVTPVDFTSNSSTSDSSEHKGIIIAKLRRGQELKLKALARKGIGKDHAK
WSPAATVTYMYEPDIIINEEMMNTLTDEEKIDLIESSPTKVFGIDPVTGQVVVVDPEAYT
YDEEVIKKAEAMGKPGLIEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLDAIRLSD
NTVEADDQFGELGAHMREG

24/319=8% coverage
13/319=4% coverage is accounted for by peptides unique to NRPD3b. The remaining 4% of the peptides match NRPD3 as well as the NRPD3b variant.

NRPD4/NRPE4 (At4g15950)
MSEKGGKGLKSSLKSKDGGKDGSSTKLKKGRKIHFDQGTPPANYKILNVSSDQQPFQSS
AAKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDGIKGQLVGLSEDPSIKI
PVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAFIPS
LKTKKEVINQPLQDALEELSKLKKSE

26/205=13% coverage
All peptides are unique matches to NRPD4/NRPE4 only.

NRPB4 (At5g09920)
MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
HDDEAIEKMLNDLSLVKRFE

0/138=0% coverage
No peptides were identified that match this protein sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3) (At3g22320)
MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKA
KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
FGLKRGQVVKIIRPSETAGRYVTYRYVV

31/205=15% coverage
All peptides are unique matches to NRPB5/NRPD5 only.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)
MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
RTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSLKPQHQVLNDEEKTTLLKKFSIEEK
QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRCVW

0/222=0% coverage
No peptides were identified that match this protein sequence.

NRPB5-like family member (At5g57980)
MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMNKVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVPVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMLTIEVFEEAELVTNITEHKLVNKYYVLDDQAKKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRSDATSLDYYTYRFAV

0/210=0% coverage
No peptides were identified that match this protein sequence.

NRPE5-like family member (At2g41340)
MEGKGKEIVVGHSISKSSVECHKYYLARRTTMEMLRDRGYDVSDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKKISVVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLRPKHQVLNDKEKESLLKKFSIEEKQL
PRLSSKDPIVRYYGLETGQVMKVTYKDELSESHVTYRCVS

0/218=0% coverage
No peptides were identified that match this protein sequence.

NRPE5-like family member (At3g54490)
MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKKILVVFMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTTFPFTVETFPIEDLLVNITKHIQQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVTYRCII

0/233=0% coverage
No peptides were identified that match this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)
MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87=0% coverage
No peptides were identified that match this protein sequence.

NRPB6a/NRPD6a/NRPE6a (At5g51940)
MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDGSFEEWGVDELIVEDSWKRQVGGD

48/144=33% coverage
22/144=15% unique

NRPB6b/NRPE6b (At2g04630)
MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDMSYEEWGVDELIVEDSWKRQVGGD

26/144=18% coverage
0/144=0% unique
This peptide is not a unique match to NRPB6b—it matches either NRPB6a or NRPB6b.

NRPE7 (At4g14660)
MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYYVAVTTLDKI
GEGKIREHTGEVLFPVMFSGMTFKIFKGEIIHGVVHKVLKHGVFMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVVIGIKWMEVEREFQALASLEGDYLGPLSE

13/177=9% coverage
This peptide matches NRPE7 only. This protein might sometimes be used as an alternative NRPD7 subunit.

NRPD7 (At3g22900)
MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVLDTEWREAEKDFMALASIDGDNLGPF

90/174=52% coverage
These peptides match NRPD7 only.

NRPB7 family member (At4g14520)
MFSEVEMARDVAICAKHLNGQSPHQPILCRLLQDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFPVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLSEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200=0% coverage
No peptides were identified that match this protein sequence.

NRPB7 (At5g59180)
MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQCVVFRPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTTSDGSVKIQKECEVRLKIIGTRVDATAIFCVGTIKDDFLGVINDPAAA

0/176=0% coverage
No peptides were identified that match this protein sequence.

NRPD8a/NRPE8a/NRPB8a (At1g54250)
MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

0/146=0% coverage
No peptides were identified that match this protein sequence.

NRPD8b/NRPE8b/NRPB8b (At3g59600)
MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAVGDKF
TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLQGDPAHISHFELDQRLFLLMRKL

25/146=18% coverage
This peptide is a unique match to At3g59600.

NRPD9a/NRPE9a/NRPB9a (At3g16980)
MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEGMTLFFVCCNPNCGHRWRE

0/114=0% coverage
No peptides were identified that match this protein sequence.

NRPD9b/NRPE9b/NRPB9b (At4g16265)
MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
LSDVASDPTLPRTKAVRCAKCQHGEAVFFQATARGEEGMTLFFVCCNPNCSHRWRE

25/114=22% coverage
This peptide is a unique match to NRPD9.

NRPD10/NRPE10/NRPB10 (At1g11475)
MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
LLNYNTLEKSDNS

39/71=54% coverage
Both peptides are a unique match to NRPD10.

NRPB10 family member (At1g61700)
MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
KLLNYNTMEKSDPN

0/71=0% coverage
No peptides were identified that matched this protein sequence.

NRPD11/NRPB11/NRPE11 (At3g52090)
MNAPERYERFVVPEGTKKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
GYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLDKELDYLKNQFEAEVAKFSNQF

65/116=56% coverage
All peptides are a unique match to NRPD11.

NRPB12/NRPD12/NRPE12 (At5g41010)
MDPAPEPVTYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTRRVVQYEAR

7/51=16% coverage
This peptide is unique to At5g41010.

NRPB12 family member (At1g53690)
MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV

0/62=0% coverage
No peptides were identified that matched this protein sequence.

Figure S3. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS analysis of affinity purified Pol II (NRPB2-FLAG). Highlighting is the same as in Fig. S1.

NRPB1 (At4g35800)
MDTRFPFSPAEVSKVRVVQFGILSPDEIRQMSVIHVEHSETTEKGKPKVGGLSDTRLGTI
DRKVKCETCMANMAECPGHFGYLELAKPMYHVGFMKTVLSIMRCVCFNCSKILADEV
CRSLFRQAMKIKNPKNRLKKILDACKNKTKCDGGDDIDDVQSHSTDEPVKKSRGGCGA
QQPKLTIEGMKMIAEYKIQRKKNDEPDQLPEPAERKQTLGADRVLSVLKRISDADCQLL
GFNPKFARPDWMILEVLPIPPPPVRPSVMMDATSRSEDDLTHQLAMIIRHNENLKRQEK
NGAPAHIISEFTQLLQFHIATYFDNELPGQPRATQKSGRPIKSICSRLKAKEGRIRGNLMG
KRVDFSARTVITPDPTINIDELGVPWSIALNLTYPETVTPYNIERLKELVDYGPHPPPGKT
GAKYIIRDDGQRLDLRYLKKSSDQHLELGYKVERHLQDGDFVLFNRQPSLHKMSIMGH
RIRIMPYSTFRLNLSVTSPYNADFDGDEMNMHVPQSFETRAEVLELMMVPKCIVSPQAN
RPVMGIVQDTLLGCRKITKRDTFIEKDVFMNTLMWWEDFDGKVPAPAILKPRPLWTGK
QVFNLIIPKQINLLRYSAWHADTETGFITPGDTQVRIERGELLAGTLCKKTLGTSNGSLVH
VIWEEVGPDAARKFLGHTQWLVNYWLLQNGFTIGIGDTIADSSTMEKINETISNAKTAV
KDLIRQFQGKELDPEPGRTMRDTFENRVNQVLNKARDDAGSSAQKSLAETNNLKAMVT
AGSKGSFINISQMTACVGQQNVEGKRIPFGFDGRTLPHFTKDDYGPESRGFVENSYLRGL
TPQEFFFHAMGGREGLIDTAVKTSETGYIQRRLVKAMEDIMVKYDGTVRNSLGDVIQFL
YGEDGMDAVWIESQKLDSLKMKKSEFDRTFKYEIDDENWNPTYLSDEHLEDLKGIREL
RDVFDAEYSKLETDRFQLGTEIATNGDSTWPLPVNIKRHIWNAQKTFKIDLRKISDMHPV
EIVDAVDKLQERLLVVPGDDALSVEAQKNATLFFNILLRSTLASKRVLEEYKLSREAFE
WVIGEIESRFLQSLVAPGEMIGCVAAQSIGEPATQMTLNTFHYAGVSAKNVTLGVPRLR
EIINVAKRIKTPSLSVYLTPEASKSKEGAKTVQCALEYTTLRSVTQATEVWYDPDPMSTII
EEDFEFVRSYYEMPDEDVSPDKISPWLLRIELNREMMVDKKLSMADIAEKINLEFDDDL
TCIFNDDNAQKLILRIRIMNDEGPKGELQDESAEDDVFLKKIESNMLTEMALRGIPDINK
VFIKQVRKSRFDEEGGFKTSEEWMLDTEGVNLLAVMCHEDVDPKRTTSNHLIEIIEVLGI
EAVRRALLDELRVVISFDGSYVNYRHLAILCDTMTYRGHLMAITRHGINRNDTGPLMRC
SFEETVDILLDAAAYAETDCLRGVTENIMLGQLAPIGTGDCELYLNDEMLKNAIELQLPS
YMDGLEFGMTPARSPVSGTPYHEGMMSPNYLLSPNMRLSPMSDAQFSPYVGGMAFSPS
SSPGYSPSSPGYSPTSPGYSPTSPGYSPTSPGYSPTSPTYSPSSPGYSPTSPAYSPTSPSYSPT

SPSYSPTSPSYSPTSPSYSPTSPSYSPTSPSYSPTSPAYSPTSPAYSPTSPAYSPTSPSYSPTSP
SYSPTSPSYSPTSPSYSPTSPSYSPTSPAYSPTSPGYSPTSPSYSPTSPSYGPTSPSYNPQSAK
YSPSIAYSPSNARLSPASPYSPTSPNYSPTSPSYSPTSPSYSPSSPTYSPSSPYSSGASPDYSP
SAGYSPTLPGYSPSSTGQYTPHEGDKKDKTGKKDASKDDKGNP

1093/1840=59% coverage
All peptides are a unique match to NRPB1 and do not match the largest subunits of Pol I, III, IV or V.

NRPB2 (At4g21710)
MEYNEYEPEPQYVEDDDDEEITQEDAWAVISAYFEEKGLVRQQLDSFDEFIQNTMQEIV
DESADIEIRPESQHNPGHQSDFAETIYKISFGQIYLSKPMMTESDGETATLFPKAARLRNL
TYSAPLYVDVTKRVIKKGHDGEEVTETQDFTKVFIGKVPIMLRSSYCTLFQNSEKDLTEL
GECPYDQGGYFIINGSEKVLIAQEKMSTNHVYVFKKRQPNKYAYVGEVRSMAENQNRP
PSTMFVRMLARASAKGGSSGQYIRCTLPYIRTEIPIIIVFRALGFVADKDILEHICYDFADT
QMMELLRPSLEEAFVIQNQLVALDYIGKRGATVGVTKEKRIKYARDILQKEMLPHVGIG
EHCETKKAYYFGYIIHRLLLCALGRRPEDDRDHYGNKRLDLAGPLLGGLFRMLFRKLTR
DVRSYVQKCVDNGKEVNLQFAIKAKTITSGLKYSLATGNWGQANAAGTRAGVSQVLN
RLTYASTLSHLRRLNSPIGREGKLAKPRQLHNSQWGMMCPAETPEGQACGLVKNLALM
VYITVGSAAYPILEFLEEWGTENFEEISPSVIPQATKIFVNGMWVGVHRDPDMLVKTLRR
LRRRVDVNTEVGVVRDIRLKELRIYTDYGRCSRPLFIVDNQKLLIKKRDIYALQQRESAE
EDGWHHLVAKGFIEYIDTEEEETTMISMTISDLVQARLRPEEAYTENYTHCEIHPSLILGV
CASIIPFPDHNQSPRNTYQSAMGKQAMGIYVTNYQFRMDTLAYVLYYPQKPLVTTRAM
EHLHFRQLPAGINAIVAISCYSGYNQEDSVIMNQSSIDRGFFRSLFFRSYRDEEKKMGTLV
KEDFGRPDRGSTMGMRHGSYDKLDDDGLAPPGTRVSGEDVIIGKTTPISQDEAQGQSSR
YTRRDHSISLRHSETGMVDQVLLTTNADGLRFVKVRVRSVRIPQIGDKFSSRHGQKGTV
GMTYTQEDMPWTIEGVTPDIIVNPHAIPSRMTIGQLIECIMGKVAAHMGKEGDATPFTD
VTVDNISKALHKCGYQMRGFERMYNGHTGRPLTAMIFLGPTYYQRLKHMVDDKIHSR
GRGPVQILTRQPAEGRSRDGGLRFGEMERDCMIAHGAAHFLKERLFDQSDAYRVHVCE
VCGLIAIANLKKNSFECRGCKNKTDIVQVYIPYACKLLFQELMSMAIAPRMLTKHLKSA
KGRQ

750/1188=63% coverage
All peptides are a unique match to NRPB2 and do not match the second-largest subunits of Pol I, III, IV or V.

NRPB3/NRPD3/NRPE3a (At2g15430)
MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEIEVNSS
VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCSVEFRLSSKCVTDQTLD
VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRRGQELKLRAIARKGIGKDHAKW
SPAATVTFMYEPDIIINEDMMDTLSDEEKIDLIESSPTKVFGMDPVTRQVVVVDPEAYTY
DEEVIKKAEAMGKPGLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLLKQKLDAVRLSD
DTVEADDQFGELGAHMRGG

230/319=72% coverage
181/319=57% of the peptide coverage is unique to NRPB3a, whereas the other 15% matches either NRPB3a or NRPB3b.

NRPE3b (At2g15400)
MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSVEFPLSAKCVTDQTLD
VTSRDLYSADPTVTPVDFTSNSSTSDSSEHKGIIIAKLRRGQELKLKALARKGIGKDHAK
WSPAATVTYMYEPDIIINEEMMNTLTDEEKIDLIESSPTKVFGIDPVTGQVVVVDPEAYT
YDEEVIKKAEAMGKPGLIEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLDAIRLSD
NTVEADDQFGELGAHMREG

72/319=23% coverage
13/319=4% of the peptide coverage is unique to NRPB3b, whereas the other 19% matches either NRPB3a or NRPB3b. This variant may be used infrequently as an alternative NRPB3 subunit.

NRPB4 (At5g09920)
MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
HDDEAIEKMLNDLSLVKRFE

84/138=61% coverage
All of the peptides match NRPB4 and only NRPB4.

NRPD4/NRPE4 (At4g15950)
MSEKGGKGLKSSLKSKDGGKDGSSTKLKKGRKIHFDQGTPPANYKILNVSSDQQPFQSS
AAKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDGIKGQLVGLSEDPSIKI
PVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAFIPS
LKTKKEVINQPLQDALEELSKLKKSE

0/205=0% coverage
No peptides were found to match this sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3)(At3g22320)
MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKA
KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
FGLKRGQVVKIIRPSETAGRYVTYRYVV

129/205=63% coverage
All peptides match to NRPB5 and only NRPB5.

NRPE5 (formerly AtRPB5b, AtRPB23.7) (At3g57080)
MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
RTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV
LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSLKPQHQVLNDEEKTTLLKKFSIEEK
QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRCVW

0/222=0% coverage
No peptides were found to match this sequence.

NRPB5-like family member (At5g57980)
MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMNKVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVPVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMLTIEVFEEAELVTNITEHKLVNKYYVLDDQAKKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRSDATSLDYYTYRFAV

0/210=0% coverage
No peptides were found to match this sequence.

NRPE5-like family member (synonym AtRPB5d) (At2g41340)
MEGKGKEIVVGHSISKSSVECHKYYLARRTTMEMLRDRGYDVSDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKKISVVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLRPKHQVLNDKEKESLLKKFSIEEKQL
PRLSSKDPIVRYYGLETGQVMKVTYKDELSESHVTYRCVS

0/218=0% coverage
No peptides were found to match this sequence.

NRPE5-like family member (At3g54490)
MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKKILVVFMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTTFPFTVETFPIEDLLVNITKHIQQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVTYRCII

0/233=0% coverage
No peptides were found to match this sequence.

NRPB5 family member (likely pseudogene) (At3g16880)
MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87=0% coverage
No peptides were found to match this sequence.

NRPB6a/NRPD6a/NRPE6a (At5g51940)
MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDGSFEEWGVDELIVEDSWKRQVGGD

58/144=40% coverage
22/144=15% of the coverage is unique to At5g51940, whereas the other 25% matches either At5g51940 or At2g04630.

NRPB6b/NRPD6b/NRPE6b (At2g04630)
MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDMSYEEWGVDELIVEDSWKRQVGGD

58/144=40% coverage
22/144=15% of the coverage is unique to At2g04630, whereas the other 25% matches either At5g51940 or At2g04630.

NRPE7 (At4g14660)
MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYYVAVTTLDKI
GEGKIREHTGEVLFPVMFSGMTFKIFKGEIIHGVVHKVLKHGVFMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVVIGIKWMEVEREFQALASLEGDYLGPLSE

0/177=0% coverage
No peptides were found to match this sequence.

NRPD7 (At3g22900)
MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVLDTEWREAEKDFMALASIDGDNLGPF

0/174=0% coverage
No peptides were found to match this sequence.

NRPB7 (At5g59180)
MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQCVVFRPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTTSDGSVKIQKECEVRLKIIGTRVDATAIFCVGTIKDDFLGVINDPAAA

89/176=51% coverage
All peptide coverage is unique to NRPB7 only.

NRPB7 family member (At4g14520)
MFSEVEMARDVAICAKHLNGQSPHQPILCRLLQDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFPVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLSEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200=0% coverage
No peptides were found to match this sequence.

NRPB8a/NRPD8a/NRPE8a (At1g54250)
MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

96/146=66% coverage
44/146=30% of the coverage is unique to NRPB8a, whereas 33% matches either NRPB8a or NRPB8b.

NRPB8b/NRPD8b/NRPE8b (At3g59600)
MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAVGDKF
TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLQGDPAHISHFELDQRLFLLMRKL

96/146=66% coverage
40/146=30% of the coverage is unique to NRPB8b, whereas 33% matches either NRPB8a or NRPB8b.

NRPB9a/NRPD9a/NRPE9a (At3g16980)
MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEGMTLFFVCCNPNCGHRWRE

35/114=30% coverage
25/114=22% of the coverage is unique to NRPB9a, whereas the other 8% matches either NRPB9a or NRPB9b.

NRPB9b/NRPD9b/NRPE9b (At4g16265)
MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
LSDVASDPTLPRTKAVRCAKCQHGEAVFFQATARGEEGMTLFFVCCNPNCSHRWRE

42/114=37% coverage
32/114=28% of the coverage is unique to NRPB9b, whereas the other 9% matches either NRPB9a or NRPB9b.

NRPB10/NRPD10/NRPE10 (At1g11475)
MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
LLNYNTLEKSDNS

50/71=70% coverage
39/71=55% of the coverage matches only At1g11475, whereas the remaining 15% matches either At1g11475 or At1g61700.

NRPB10 family member (At1g61700)
MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
KLLNYNTMEKSDPN

11/71=15% coverage, matching either At1g11475 or At1g61700.
0/71=0% of the coverage is unique to At1g61700.

NRPB11/NRPD11/NRPE11 (At3g52090)
MNAPERYERFVVPEGTKKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
GYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLDKELDYLKNQFEAEVAKFSNQF

87/116=75% coverage
All peptide coverage matches NRPB11 only.

NRPB12/NRPD12/NRPE12 (At5g41010)
MDPAPEPVTYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTRRVVQYEAR

8/51=16% coverage
This peptide matches only At5g41010.

RPB12 family member (At1g53690)
MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV

0/62=0% coverage
No peptides were found to match this sequence.

Figures S4-S12

These figures show ClustalW alignments of Arabidopsis and yeast RPB4, 5, 6, 7, 8, 9, 10, 11 and 12 family proteins. Red highlighting denotes invariant residues, yellow denotes conserved residues and cyan denotes similar residues.

Figure S4. RPB4 family alignment

```
Sc_RPB4            MNVSTSTFQTRRRRLKKVEEEENAATLQLGQEFQLKQINHQGEEEELIALNLSEARLVIK
At_NRPB4           ---------------MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISE
At_NRPD4/NRPE4     MSEKGGKGLKSSLKSKDGGKDGSSTKLKKGRKIHFDQGTPPANYKILNVSSDQQPFQSSA

Sc_RPB4            EALVERRRAFKRSQKKHKKKHLKHENANDETTAVEDEDDDLDEDDVNADDDDFMHSETRE
At_NRPB4           DPMNQVSQVFEKS-----------------------------------------------
At_NRPD4/NRPE4     AKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDG-----------------

Sc_RPB4            KELESIDVLLEQTTGGNNKDLKNTMQYLTNFSRFRDQETVGAVIQLLKSTGLHPFEVAQL
At_NRPB4           ------------------------LQYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVL
At_NRPD4/NRPE4     -IKGQLVGLSEDPSIKIPVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVI

Sc_RPB4            GSLACDTADEAKTLIPSLNNK---ISDDELERILKELSNLETLY
At_NRPB4           GNLCPETVEEAVAMVPSLKTKGRAHDDEAIEKMLNDLSLVKRFE
At_NRPD4/NRPE4     ANASSESVDEVLAFIPSLKTK-KEVINQPLQDALEELSKLKKSE
```

Figure S5. RPB5 family alignment. For this alignment, the following codes are used:
At3g22320=NRPB5/NRPD5
At5g57980=NRPB5-like
At3g57080=NRPE5
At2g41340=NRPE5-like
At3g54490=NRPE5-like
At3g16880=likely pseudogene

```
Start of jaw domain
Sc_RPB5          ---------------------------MDQENERNISRLWRAFRTVKEMVKDRGYFITQE
Hs_RPB5          ---------------------------MDDEEE--TYRLWKIRKTIMQLCHDRGYLVTQD
At_NRPB5/NRPD5   -----------------------------MLTEEELKRLYRIQKTLMQMLRDRGYFIADS
At5g57980        ----------------------------MSDMDDEITRIFKVRRTVLQMLRDRGYTIEES
At_NRPE5         MEVKGKETASVL-----------CLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDE
At2g41340        MEGKGKEIVVGH-----------SISK----SSVECHKYYLARRTTMEMLRDRGYDVSDE
At3g54490        MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNEA
At3g16880        ------------------------------------------------------------

Sc_RPB5          EVELPLEDFKAKYCD--SMGRPQRKMMSFQANPTEESISKFPDMGSLWVEFCDEPSVGVK
Hs_RPB5          ELDQTLEEFKAQFGDKPSEGRPRRTDLTVLVAHNDD------PTDQMFVFFPEEPKVGIK
At_NRPB5/NRPD5   ELTMTKQQFIRKHGDN----MKREDLVTLKAKRNDN-------SDQLYIFFPDEAKVGVK
At3g57980        DLNLKREEFVQRFCKT--MNKVNKEALFVSANKGPN------PADKIYVFYPEGPKVGVP
At_NRPE5         DINLSLHDFRTVYGER-----PDVDRLRISALHRSD------STKKVKIVFFGTSMVKVN
At2g41340        DINLSLQQFRALYGEH-----PDVDLLRISAKHRFD------SSKKISVVFCGTGIVKVN
At3g54490        ELSLTLSEFRSVFGEK-----PELERLRICVPLRSD------PKKKILVVFMGTEPITVK
At3g16880        ------------------------------------------------------------

                                                     end of jaw domain  ─┤
Sc_RPB5          TMK-TFVIHIQEKNFQTGIFVYQNNITPSAMK----LVPSIPPATIETFNEAALVVNITH
Hs_RPB5          TIK-VYCQRMQEENITRALIVVQQGMTPSAKQS---LVDMAPKYILEQFLEQELLINITE
At_NRPB5/NRPD5d  TMK-MYTNRMKSENVFRAILVVQQNLTPFAR---TCISEISSKFHLEVFQEAEMLVNIKE
At5g57980        VIKKEVAIKMRDDKVHRGIVVVPMAITAPARMA---VSELNKMLTIEVFEEAELVTNITE
At_NRPE5         AIRSVVADILSQETITGLILVLQNHVTNQALKA-----IELFSFKVEIFQITDLLVNITK
At2g41340        AMRVIAADVLSRENITGLILVLQSHITNQALKA-----VELFSFKVELFEITDLLVNVSK
At3g54490        SVRALHIQISNNVGLHAMILVLQSKMNHFAQKA-----LTTFPFTVETFPIEDLLVNITK
At3g16880        -MK-KYIDQLKSANVFRAILVVQD-IKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKE

                 start of assembly domain
Sc_RPB5          HELVPKHIRLSSDEKRELLKRYRLKESQLPRIQRADPVALYLGLKRGEVVKIIRKSETSG
Hs_RPB5          HELVPEHVVMTKEEVSELLARYKLRENQLPRIQAGDPVARYFGIRRGQVVKIIRPSETAG
At_NRPB5/NRPD5   HVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARYFGLKRGQVVKIIRPSETAG
At5g57980        HKLVNKYYVLDDQAKKKLLNTYTVQDTQLPRILVTDPLARYYGLKRGQVVKIRRSDATSL
At_NRPE5         HSLKPQHQVLNDEEKTTLLKKFSIEEKQLPRISKKDAIVRYYGLEKGQVVKVNYRGELTE
At2g41340        HVLRPKHQVLNDKEKESLLKKFSIEEKQLPRLSSKDPIVRYYGLETGQVMKVTYKDELSE
At3g54490        HIQQPKIEILNKEEKEQLLRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVG
At3g16880        HVFVPEHQALTTEEKQKFLER---KRTSFQGFT---------------------------

Sc_RPB5          RYASYRICM
Hs_RPB5          RYITYRLVQ
At_NRPB5/NRPD5   RYVTYRYVV
At5g57980        DYYTYRFAV
At_NRPE5         SHVAFRCVW
At2g41340        SHVTYRCVS
At3g54490        DFVTYRCII
At3g16880        ---------
```

Figure S6. RPB6 family alignment

```
NRPB6a_At5g51940 MAD--EDYNDVDDLGYEDEPAEP-EIEEGVEEDVEMK--ENDDVNGEPIEA-----EDKV
NRPB6b_At2g04630 MAD--DDYNEVDDLGYEDEPAEP-EIEEGVEEDADIK--ENDDVNVDPLET-----EDKV
Sc_RPB6          MSDYEEAFNDGNEN-FEDFDVEHFSDEETYEEKPQFKDGETTDANGKTIVTGGNGPEDFQ

NRPB6a_At5g51940 ETEPVQR-----------PRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLE
NRPB6b_At2g04630 ETEPVQR-----------PRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLE
Sc_RPB6          QHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQISMNAPVFVDLEGETDPLR

NRPB6a_At5g51940 IAMKELRQRKIPFTIRRYLPDGSFEEWGVDELIVEDSWKRQVGGD
NRPB6b_At2g04630 IAMKELRQRKIPFTIRRYLPDMSYEEWGVDELIVEDSWKRQVGGD
Sc_RPB6          IAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL---------
```

Figure S7. RPB7 family alignment

```
Sc_RPA43           MSQVKRANENRETARFIKKHKKQVTNPIDEKNGTSNCIVRVPIALYVSLAPMYLENPLQG
At1g75670          ----------------------------------MEGLKLSEAELMIFIHPSQSRN-VFQ
Sc_RPB7            ------------------------------------MFFIKDLSLNITLHPSFFGP---R
NRPB7_At5g59180    ------------------------------------MFFHIVLERNMQLHPRFFGR---N
Sc_RPC25           ------------------------------------MFILSKIADLVRIPPDQFHR----
At1g06790          ------------------------------------MFYLSELEHSLRVPPHLLNL----
NRPE7_At4g14660    ------------------------------------MFLKVQLPWNVMIPAENMDAKGLM
NRPD7_At3g22900    ------------------------------------MFIKVKLPWDVTIPAEDMDT-GLM
At4g14520          ------------------------------------MFSEVEMARDVAICAKHLNG--QS

Sc_RPA43           VMKQHLNPLVMKYNNKVGGVVLGYEGLKILDADPLSKEDTSEKLIKITPDTPFGFTWCHV
At1g75670          GICRELSSLLFQYNETFDGVLLAYDATVKSKQAKILTG--------LHPYFG---VRVNT
Sc_RPB7            MKQYLKTKLLEEVEG-SCTGKFGYI-LCVLDYDNIDIQRG-----RILPTDGSAEFNVKY
NRPB7_At5g59180    LKENLVSKLMKDVEG-TCSGRHGFV-VAITGID--TIGKG-----LIRDGTGFVTFPVKY
Sc_RPC25           DTISAITHQLNNKFANKIIPNVGLC-ITIYDLLTVEEGQ-------LKPGDGSSYINVTF
At1g06790          PLEDAIKSVLQNVFLDKVLADLGLC-VSIYDIKSVEGGF-------VLPGDGAATYKVGL
NRPE7_At4g14660    LKRAILVELLEAFASKKATKELGYY-VAVTTLDKIGEGK-------IREHTGEVLFPVMF
NRPD7_At3g22900    LQRAIVIRLLEAFSKEKATKDLGYL-ITPTILENIGEGK-------IKEQTGEIQFPVVF
At4g14520          PHQPILCRLLQDLIHEKACREHGFY-LGITALKSIGNNKNNNIDNENNHQAKILTFPVSF

Sc_RPA43           NLYVWQPQVGDVLEGYIFIQSASHIGLLIHDAFNASIKKNNIPVDWTFVHNDVEEDADVI
At1g75670          RLLLFDPKPKSFVEG--KIVKISPESIHVIVLG-----FSAAVITDVDIREEFKYRVR--
Sc_RPB7            RAVVFKPFKGEVVDG--TVVSCSQHGFEVQVG------PMKVFVTKHLMPQDLTFNAGS-
NRPB7_At5g59180    QCVVFRPFKGEILEA--VVTLVNKMGFFAEAG------PVQIFVSKHLIPDDMEFQAG--
Sc_RPC25           RAVVFKPFLGEIVTG--WISKCTAEGIKVSLLG----IFDDIFIPQNMLFEGCYYTPE--
At1g06790          RIVVFRPFVGEVIAA--KFKESDANGLRLTLG-----FFDDIYVPAPLMPKPNRCEPDPY
NRPE7_At4g14660    SGMTFKIFKGEIIHG--VVHKVLKHGVFMRCG------PIENVYLSYTKMPDYKYIPG--
NRPD7_At3g22900    NGICFKMFKGEIVHG--VVHKVHKTGVFLKSG------PYEIIYLSHMKMPGYEFIPG--
At4g14520          TCRTFLPARGDILQG--TVKKVLWNGAFIRSG------PLRYAYLSLLKMPHYHYVHSPL

Sc_RPA43           NTDENNGNNNNEDNKDSNGGSNSLGKFSFGNRSLGHWVDSNGEPIDGKLRFTVRNVHTTG
At1g75670          ---DGEGSFVSRSHKR------------------------HALKLGTMLRLQVQSFDEEV
Sc_RPB7            ---NPPSYQSSEDVIT--------------------------IKSR--IRVKIEGCISQV
NRPB7_At5g59180    ---DMPNYTTSDGSVK--------------------------IQKECEVRLKIIGTRVDA
Sc_RPC25           ---ESAWIWPMDEETK------------------------LYFDVNEKIRFRIEREVFVD
At1g06790          NRKQMIWVWEYGEPKED-----------------------YIVDDACQIKFRVESISYPS
NRPE7_At4g14660    ---ENPIFM-NEKTSR--------------------------IQVETTVRVVVIGIKWME
NRPD7_At3g22900    ---ENPFFM-NQYMSR--------------------------IQIGARVRFVVLDTEWRE
At4g14520          SEDEKPHFQ-KDDLSK--------------------------IAVGVVVRFQVLAVRFKE

Sc_RPA43           RVVSVDGTLISDADEEGNGYNSSRSQAESLPIVSNKKIVFDDEVSIENKESHKELDLPEV
At1g75670          MHIAG------------------------------------------SLLPENTGCVKWL
Sc_RPB7            SSI---------------------------------------------------HAIGSI
NRPB7_At5g59180    TAI---------------------------------------------------FCVGTI
Sc_RPC25           VKPKSP-----------------KERELEERAQLENEIEGKNEETPQNEKPPAYALLGSC
At1g06790          VP---------------------------------------TERAEDAKPFAPMVVTGNM
NRPE7_At4g14660    VER-------------------------------------------------EFQALASL
NRPD7_At3g22900    AEK-------------------------------------------------DFMALASI
At4g14520          RPHK---------------------------------------------RRNDYYVLATL

Sc_RPA43           KEDNGSEIVYEENTSESNDGESSDSD
At1g75670          EKKSEEALPTDRDHKRRKLA------
Sc_RPB7            KED-YLGAI-----------------
NRPB7_At5g59180    KDD-FLGVINDPAAA-----------
Sc_RPC25           QTD-GMGLVSWWE-------------
At1g06790          DDD-GLGPVSWWDSYEQVDQEE----
NRPE7_At4g14660    EGD-YLGPLSEE--------------
NRPD7_At3g22900    DGD-NLGPF-----------------
At4g14520          EGNGSFGPISLTGSDEPYM-------
```

Figure S8. RPB8 family alignment

```
NRPB8a_At1g54250 MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMF-MHLDVNTEVYPLAVGDKFT
NRPB8b_At3g59600 MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMF-MHLDVNTEVYPLAVGDKFT
Sc_RPB8          MSN--TLFDDIFQVSEVDPG--RYNKVCRIEAASTTQDQCKLTLDINVELFPVAAQDSLT

NRPB8a_At1g54250 LALAPTLNLDGTPDTG-----YFTP--GAKKTLADKYEYIMHGKLYKISERDGKTPKAEL
NRPB8b_At3g59600 LAMAPTLNLDGTPDTG-----YFTP--GAKKTLADKYEYIMHGKLYKISERDGKTPKAEL
Sc_RPB8          VTIASSLNLEDTPANDSSATRSWRPPQAGDRSLADDYDYVMYGTAYKFEEVS--KDLIAV

NRPB8a_At1g54250 YVSFGGLLMLLKGDPAHISHFELDQRLFLLMRKL
NRPB8b_At3g59600 YVSFGGLLMLLQGDPAHISHFELDQRLFLLMRKL
Sc_RPB8          YYSFGGLLMRLEGNYRNLNNLKQEN-AYLLIRR-
```

Figure S9. RPB9 family alignment

```
Sc_RPA12         --MSVVGSLIFCLDCGDLLENPNAVLG--SNVECSQCKAIYPKSQFSNLKVVTTTADDAF
At3g25940        MEKSRESEFLFCNLCGTMLVLKST-----KYAECPHCKTTRNAKDIIDKEIAYTVSAEDI
Sc_RPB9-         -----MTTFRFCRDCNNMLYPREDKENNRLLFECRTCSYVEEAGS-PLVYRHELITNIGE
NRPE9a_At3g16980 -----MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADN-SCVYRNEVHHSVSE
NRPE9b_At4g16265 -----MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADN-NCVYRNEVHHSVSE
Sc_RPC11         -------MLSFCPSCNNMLLITSGDS-GVYTLACRSCPYEFPIEG-IEIYDRKKLPRKEV
At4g07950        --------MEFCPTCGNLLRYEGG---GSSRFFCSTCPYVANIERRVEIKKKQLLVKKSI
At1g01210        --------MEFCPTCGNLLRYEGG---GNSRFFCSTCPYVAYIQRQVEIKKKQLLVKKSI

Sc_RPA12         PSSLRAKKSVVKTSLKKNELKDGATIKEKCPQCGNEEMNYHTLQLRSADEGATVFYTCTS
At3g25940        RRELGISLFGEKTQAEAELPKI----KKACEKCQHPELVYTTRQTRSADEGQTTYYTCPN
Sc_RPB9-         TAGVVQDIGSDPTLPRSDRE---------CPKCHSRENVFFQSQQRRKDTSMVLFFVCLS
NRPE9a_At3g16980 RTQILTDVASDPTLPRTKAVR--------CSKCQHREAVFFQATARGEE-GMTLFFVCCN
NRPE9b_At4g16265 QTQILSDVASDPTLPRTKAVR--------CAKCQHGEAVFFQATARGEE-GMTLFFVCCN
Sc_RPC11         DDVLG-GGWDNVDQTKTQCPN--------YDTCGGESAYFFQLQIRSADEPMTTFYKCVN
At4g07950        EPVVTKDDIPTAAETEAPCP-----------RCGHDKAYFKSMQIRSADEPESRFYRCLK
At1g01210        EAVVTKDDIPTAAETEAPCP-----------RCGHDKAYFKSMQIRSADEPESRFYRCLK

Sc_RPA12         --CGYKFRTNN--------
At3g25940        --CAHRFTEG---------
Sc_RPB9-         --CSHIFTSDQKNKRTQFS
NRPE9a_At3g16980 PNCG---------------
NRPE9b_At4g16265 PNCSHRWREHRWRE-----
Sc_RPC11         --CGHRWKEN---------
At4g07950        --CEFTWREE---------
At1g01210        --CEFTWREE---------
```

Figure S10. RPB10 family alignment

```
NRPB10/NRPD10/NRPE10 MIIPVRCFTCGKVIGNKWDQYLDLLQLD-YTEGDALDALQLVRYCCRRMLMTHVDLIEKL
At1g61700            MIVPVRCFTCGKVIGNKWDTYLELLQAD-YAEGDALDALGLVRYCCRRMLMTHVDLIEKL
Sc_RPB10             MIVPVRCFSCGKVVGDKWESYLNLLQEDELDEGTALSRLGLKRYCCRRMILTHVDLIEKF

NRPB10/NRPD10/NRPE10 LNYNTLEKSDNS
At1g61700            LNYNTMEKSDPN
Sc_RPB10             LRYNPLEKRD--
```

Figure S11. RPB11 alignment

```
NRPB11_At3g52090 MNAPERYERFVVPEGTKKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFAG
Sc_RPB11         MNAPDRFELFLLGEGESKLKIDPDTKAPNAVVITFEKEDHTLGNLIRAELLNDRKVLFAA

NRPB11_At3g52090 YQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLDKELDYLKNQFEAEVAKFS----NQF
Sc_RPB11         YKVEHPFFARFKLRIQTTEGYDPKDALKNACNSIINKLGALKTNFETEWNLQTLAADDAF
```

Figure S12. RPB12 family alignment

```
NRPB12/NRPD12/NRPE12 -----------MDP--------APEP-VTYVCGDCGQENTLKSGDVIQCRECGYRILYKK
At1g53690            -----------MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKK
Sc_RPB12             MSREGFQIPTNLDAAAAGTSQARTAT-LKYICAECSSKLSLSRTDAVRCKDCGHRILLKA

NRPB12/NRPD12/NRPE12  RTRRVVQYEAR-
At1g53690             RILDKKETRIGV
Sc_RPB12              RTKRLVQFEAR-
```

Figure S13. Expression patterns of the RPB5 family. RT-PCR detection of mRNAs corresponding to the six *Arabidopsis* genes homologous to yeast RPB5. Actin served as a control to show that similar amounts of RNA were isolated from the tissues tested.

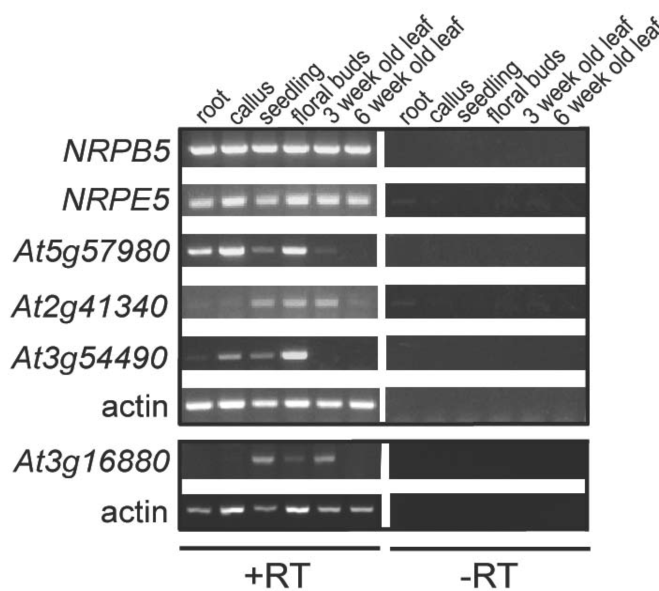
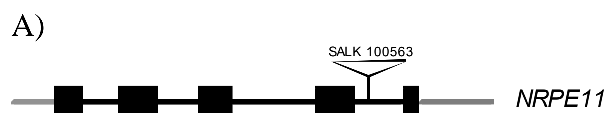


Figure S14. Analysis of *nrpd5-1* and *nrpe11-1* T-DNA insertion mutants. A. Gene structure of *NRPE11* and location of the T-DNA insertion. B. Genotyping results for offspring from a selfed *nrpd5*/+ and *nrpe11*/+ heterozygotes (top) and genotyping results of F1 offspring of reciprocal crosses between *nrpd5*/+ heterozygotes and wild-type plants *nrpe11*/+ heterozygotes and wild-type plants (bottom). *nrpd5-1* homozygotes are not recovered due to female gametophyte lethality, as shown by reciprocal crosses, whereas *nrpe11-1/nrpd11-1/nrpb11-1* homozygous mutants (abbreviated as *nrpe11-1* below) appear to be embryo lethal since the T-DNA is passed through both the male and female gametophyte. C. RT-PCR of transcript levels in Col wt vs. *nrpe5-1* mutants using primers that span the T-DNA insertion or are upstream of the T-DNA insertion. Actin served as a control to show that similar amounts of RNA were loaded in each genotype.

A)



B)

Genetic analysis of RNA polymerase subunits.

Progeny of:	+/+	+/-	-/-	total
nrpd5-1/+	80	21	0	101
nrpe11-1/+	33	63	0	96

F1 progeny of female x male:	+/+	+/-	total
nrpd5-1/+ x +/+	78	0	78
+/+ x *nrpd5-1/+*	23	10	33
nrpe11-1/+ x +/+	13	4	17
+/+ x *nrpe11-1/+*	25	25	50

C)

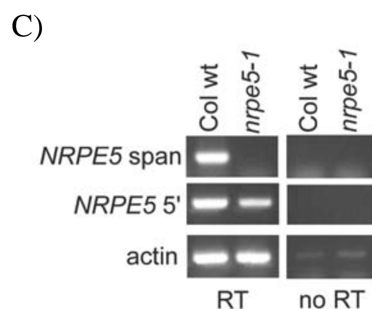


Figure S15.

Flowering time of individual plants from wild-type (ecotype Col-0) and *nrpe5-1* populations.

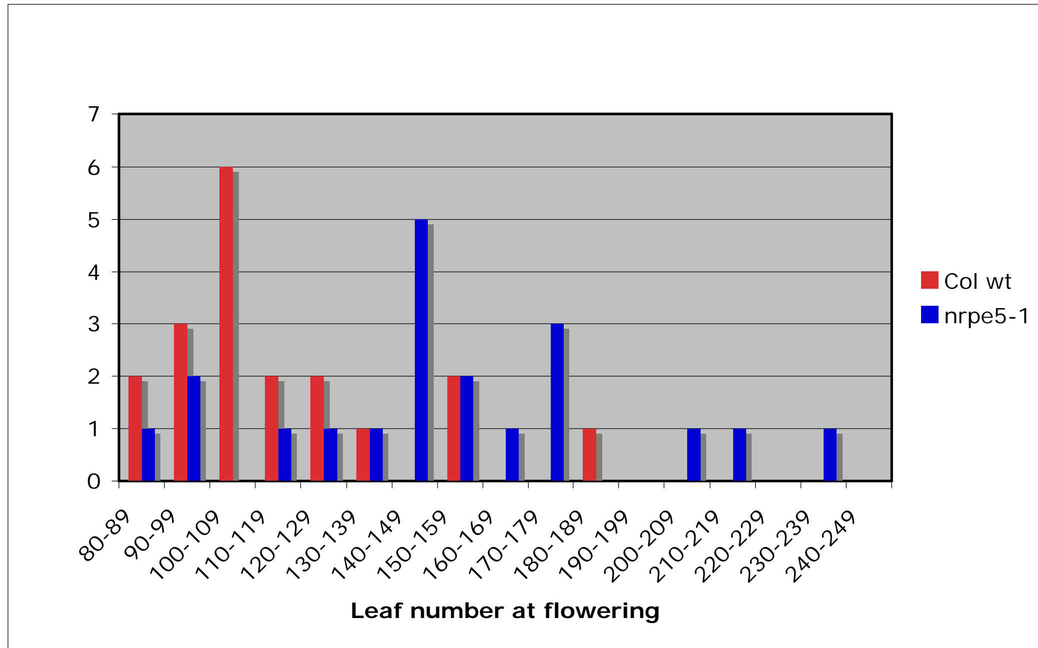


Figure S16. Alignment of RPB5 family variants in diverse plants with non-plant RPB5s. Red: absolutely conserved residues; yellow: consensus residues; cyan: similar residues. Locations of the jaw and assembly domains are indicated by arrows. Hs= *Homo sapiens*; Dm = *Drosophila melanogaster*; Ce = *Caenorhabditis elegans*; Sc = *Saccharomyces cerevisiae*

```
                             →Start of Jaw domain
C_reinhardtii_XP_001697601   ----------------------------------------------MDN-
O_lucimarinus_XP_001417617   ----------------------------------------------MSND
Hs_RPB5_BAA07406             ----------------------------------------------MDDE
Dm_RPB5_NP_610630            ----------------------------------------------MDDE
Ce_RPB5_Q9N5K2               ---------------------------------------------MADDE
Populus_trichocarpa_584052   ---------------------------------------------MTLTE
Vitis_vinifera_CAO65489      ---------------------------------------------MSASE
NRPB5/NRPD5_At3g22320        ----------------------------------------------MLTE
Medicago_truncatula_ABO78350 ---------------------------------------------MVFSE
Oryza_sativa_EAZ13876        -----------------------------------------MSAGLVTEE
Oryza_sativa_NP_001044564    -----------------------------------------MSAGLVTEE
Zea_mays_ACF85599            -----------------------------------------MSAGLVTDE
Oryza_sativa_CAD41325        ---------------------------------------------MAS-E
Oryza_sativa_EAZ31161        ---------------------------------------------MAS-E
Zea_mays_ACF81264            ---------------------------------------------MASPD
Physcomitrella_patens_206246 -------------------------------------------MSGQSLD
Physcomitrella_patens_55574  -------------------------------------------MSGQSLD
Physcomitrella_patens_231299 -------------------------------------------MAEHVLD
Physcomitrella_patens_136486 -------------------------------------------MAEHVLD
Sc_RPB5_CAA85113             --------------------------------------------MDQENE
NRPB5-like_At5g57980         ---------------------------------------------MSDMD
NRPE5-like_At2g41340         --------------------MEGKG-------KEIVVGHSISK----SS-
Brassica_napus_AAF81222      --------------------MEGKG-------KELAVGSGLSKSLDESR-
NRPE5_At3g57080_NP_191267    --------------------MEVKG-------KETASVLCLSKYVDLSS-
Populus_trichocarpa_57931    --------------------MES-------------LGRCLSSFVDEGS-
Vitis_vinifera_CAO63075      --------------------MDGGGWFDGDLNGDFEVKRCLSSFVDEGR-
Medicago_truncatula_ABN07995 -MATENGGGQNGTTETAITTMEIENGDITTQPQLQEQPQCLFTKKDNGS-
Populus_trichocarpa_48513    ----------------MAATTETFNGNGASFHGVLDRDRCLTDFVDEGS-
Vitis_vinifera_CAO42914      ----------------MESQAGSH-GNGS----------CITADMEQGS-
NRPE5-like_At3g54490         ----------------MEETMAEEGCCENVESTFDDGTNCISKTEDTGG-
Medicago_truncatula_ABD28306 -------------------MAMIENGNET-------RSECLVRICNEESN
Oryza_sativa_NP_001065723    --------------------MAAEMEVDDV--DVHEVPECIASMIDRG-S
Oryza_sativa_NP_001066119    --------------------MAAEMEVDDV--DVHEVPECIASMIDRG-S
Oryza_sativa_EAY79909        --------------------MAAEMEADDV--DVHEVPECIASMIDRG-S
Zea_mays_ACF87172            MESAESTAAAAARASNGAARAVVEDDDEDD--DVPEVAACISTMLDRGGS
consensus                                                                   E
```

(cont'd below)

```
C_reinhardtii_XP_001697601   --LTRLWRVRRTCLQMLNDRGYLVSQEEIGTTKDQFRDRFGENP-----R
O_lucimarinus_XP_001417617   -KRTRLFRVRKTIHKMLAARGYLVSAKELERDIDSFTEDFGEEP-----K
Hs_RPB5_BAA07406             EETYRLWKIRKTIMQLCHDRGYLVTQDELDQTLEEFKAQFGDKPSEGRPR
Dm_RPB5_NP_610630            AETYKLWRIRKTIMQLSHDRGYLVTQDELDQTLEQFKEMFGDKPSEKRPA
Ce_RPB5_Q9N5K2               LETYRLWRIRKTVLQMVHDRGYLVAQDELDQPLETFKVQYGDRPSEKKPA
Populus_trichocarpa_584052   EEIKRLLRIRKTVMQMLKDRGYFVGDFEIKMTREQFESKYGNNM-----K
Vitis_vinifera_CAO65489      EEISRLFRIRKTVMQMLKDRGYFVGDFEINMTKHQFVSKFGENM-----K
NRPB5/NRPD5_At3g22320        EELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNM-----K
Medicago_truncatula_ABO78350 EEITRLYRIRKTVMQMLKDRNYLVGDFELNMSKHDFKDKYGENM-----K
Oryza_sativa_EAZ13876        VMVGRLVRIRRTVMQMLRDRGYLVVEHELAMGRRDFLRKYGESF-----H
Oryza_sativa_NP_001044564    VMVGRLVRIRRTVMQMLRDRGYLVVEHELAMGRRDFLRKYGESF-----H
Zea_mays_ACF85599            ATVGRLYRIRRTVMQMLRDRGYLVVDHELATSRRDFLRKFGESF-----H
Oryza_sativa_CAD41325        EETSRLFRIRRTVMQMLRDRGYLVTELDIDLPRGDFVARFGDPV-----D
Oryza_sativa_EAZ31161        EETSRLFRIRRTVMQMLRDRGYLVTELDIDLPRGDFVARFGDPV-----D
Zea_mays_ACF81264            DEISRLFRIRRTVYEMLRDRGYGVRDEQIKLERHKFIERYGNPV-----R
Physcomitrella_patens_206246 EQSARLYRIRKTVMEMLRDRDYVVADYELTLSKEQFREKYGDEP-----K
Physcomitrella_patens_55574  EQCARLYRIRKTVMEMLRDRDYVVAEFELNSTKEEFREKYGDEP-----K
Physcomitrella_patens_231299 RQSTHLYQVRKKVLEMMRDLDYVVADNELTLTNEQFCEKYREDP-----K
Physcomitrella_patens_136486 RQSTHLYQVRKKVLEMMRDLDYVVADNELTLTNEQFCEKYREDP-----K
Sc_RPB5_CAA85113             RNISRLWRAFRTVKEMVKDRGYFITQEEVELPLEDFKAKYCDSMG-----
NRPB5-like_At5g57980         DEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMN--KVN
NRPE5-like_At2g41340         VECHKYYLARRTTMEMLRDRGYDVSDEDINLSLQQFRALYGEHP-----D
Brassica_napus_AAF81222      VDSHSYYLARRTTMEMLRDRGYDISNEDINLTLQEFRALYGDRP-----N
NRPE5_At3g57080_NP_191267    EESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDFRTVYGERP-----D
Populus_trichocarpa_57931    TESHRYYLSRRTVLEMLKDRGYSVPSSEIDISLQDFRGVYGQNP-----D
Vitis_vinifera_CAO63075      IESHRYYLARRTLLEMLRDRGYSIPALDIDISLQDFRSFYSQKP-----D
Medicago_truncatula_ABN07995 IESHRYYLSRRTVLEMLKDRGYSIPSDEIQLSLDDFRQIHGQSP-----D
Populus_trichocarpa_48513    AESYRYYISRRTVLEMLKDRGYDVLDSELNRSLTEFRSVFGNSP-----D
Vitis_vinifera_CAO42914      IESYRYYLSRRTLFQMLSDRGYNVPHSELTRSLSDFRASFGHNP-----D
NRPE5-like_At3g54490         IESKRFYLARTTAFEMLRDRGYEVNEAELSLTLSEFRSVFGEKP-----E
Medicago_truncatula_ABD28306 IETIRYFECRKTLMDMLHDRGYNVSESDLTLSLSEFRSRFGEFP-----K
Oryza_sativa_NP_001065723    VESHRLFLARRTAMEMLRDRGYSVPEAEIARTLPEFRAWWAEKP-----G
Oryza_sativa_NP_001066119    VESHRLFLARRTAMEMLRDRGYSVPEAEIARTLPEFRAWWAEKP-----G
Oryza_sativa_EAY79909        VESHRLFLARRTAMEMLRDRGYSVPEAEIARTPPEFRAWWAEKP-----G
Zea_mays_ACF87172            VESHRLFLARRTALEMLRDRGYAVPEEELARTLPEFRAWWEYRP-----E
consensus                     ES RLYRIRRTVMEMLRDRGY V E EL LTL DFR KYGE P

C_reinhardtii_XP_001697601   KDDLTILVPRQDDPTEQIFVFFP---------EEQKVGVKTIK-LLAERM
O_lucimarinus_XP_001417617   RESLTILAPKRDDPSENIFVFFP---------DEEKVGVKTIK-DLAKRM
Hs_RPB5_BAA07406             RTDLTVLVAHNDDPTDQMFVFFP---------EEPKVGIKTIK-VYCQRM
Dm_RPB5_NP_610630            RSDLIVLVAHNDDPTDQMFVFFP---------EEPKIGIKTIK-TYCTRM
Ce_RPB5_Q9N5K2               RSDLTILVAHNDDPADQMFVFFP---------EDAKIGIKTIK-AICQQM
Populus_trichocarpa_584052   REDLVINKTKRNDSSDQIYVFFP---------EEAKVGVKTMK-TYTNRM
Vitis_vinifera_CAO65489      REDLVINKAKRTDSSDQIYVFFP---------EEQKVGVKTMK-TYTNRM
NRPB5/NRPD5_At3g22320        REDLVTLKAKRNDNSDQLYIFFP---------DEAKVGVKTMK-MYTNRM
Medicago_truncatula_ABO78350 REDLVINKTKKDKPSDQIYVFFP---------EEAKVGVKTMK-TYTNRM
Oryza_sativa_EAZ13876        REDLLINKYKKNDPSDQIYVFFP---------NDDKVGMKHIK-KYVEMM
Oryza_sativa_NP_001044564    REDLLINKYKKNDPSDQIYVFFP---------NDDKVGMKHIK-KYVEMM
Zea_mays_ACF85599            REDLLINKYKKNDPSDQIYVFFP---------NDDKVGMKHIK-KYVEMM
Oryza_sativa_CAD41325        RDHLVFSRHKKDNGADQIYVFFP---------KDAKPGVKTIR-SYVERM
Oryza_sativa_EAZ31161        RDHLVFSRHKKDNGADQIYVFFP---------KDAKPGVKTIR-SYVERM
Zea_mays_ACF81264            RDELTFNATKLNGPSDQIYVFFP---------NEAKPGVKTIR-NYVEKM
Physcomitrella_patens_206246 REDLVIQKPRRSNNAEHIFVFFP---------EEAKVGVKTIK-TYVDRM
Physcomitrella_patens_55574  REDLVIQKPKRSNNAEHIFVFFP---------EEAKVGVKTIK-TYVDRM
Physcomitrella_patens_231299 QEDLMILKPKSSNNAEHGPKTGG----------KGRVGLKTIK-TCKKRM
Physcomitrella_patens_136486 QEDLMILKPKSSNNAEHVMVFHEF-----FSPFPTLVGLKTIK-TCKKRM
Sc_RPB5_CAA85113             RPQRKMMSFQANPTEESISKFPDMGSLWVEFCDEPSVGVKTMK-TFVIHI
NRPB5-like_At5g57980         KEALFVSANKGPNPADKIYVFYP---------EGPKVGVPVIKKEVAIKM
NRPE5-like_At2g41340         VDLLRISAKHRFDSSKKISVVFC---------GTGIVKVNAMRVIAADVL
Brassica_napus_AAF81222      VDRLRISAQHCSDSSKKIAVVFC---------GSGIVKVSAIRDIAADVL
NRPE5_At3g57080_NP_191267    VDRLRISALHRSDSTKKVKIVFF---------GTSMVKVNAIRSVVADIL
Populus_trichocarpa_57931    IELLKFSATHKSDPSKRMLVIFC---------GLGVVKVGMIRLITVQIT
Vitis_vinifera_CAO63075      PDRLRISAALRSDPSKKILVIFC---------GPDVVKVNAIRSIATQIV
Medicago_truncatula_ABN07995 VDRLRLTATHATNPSKRILVVFS---------GPGIVKVNGVRDIAGQIV
Populus_trichocarpa_48513    LDSLRFSVSLRSIPHKKTLVMFL---------GTDEIKTANIRTVYGQIL
Vitis_vinifera_CAO42914      PSRLRICLPLISSPSKKILVVFC---------GTDEIRKAVIRVIF-QQI
NRPE5-like_At3g54490         LERLRICVPLRSDPKKKILVVFM---------GTEPITVKSVRALHIQIS
Medicago_truncatula_ABD28306 PHTLGVSVSLRSNPSIKVQVVFP---------GTDDIRKSNLIVIQSQIV
Oryza_sativa_NP_001065723    IERLAFTTTLVSDPSKKVQLVFC---------PPEPVKIATIREIYLQTK
Oryza_sativa_NP_001066119    IERLAFTTTLVSDPSKKVQLVFC---------PPEPVKIATIREIYLQTK
Oryza_sativa_EAY79909        IERLAFTTTLVSDPSKKVQLVFC---------PPEPVKIATIREIYLQTK
Zea_mays_ACF87172            LERLAFSTTLTSDPSSKVKVVFC---------PPGPVKIAAIRLIYTEVK
consensus                    RE L I    RSDPSD IYVFFP          E KVGVKTIK  Y   M
```

```
                                                            end Jaw domain→
C_reinhardtii_XP_001697601    KDEKVNRAIMVTPSKFTPFAKSALEDMR-PKYHIEHFLESELLVNITEHV
O_lucimarinus_XP_001417617    KDENVFRAIIVVQASLTPFAKQSLLECQTQKFYIEQFQETELLVNIIDHV
Hs_RPB5_BAA07406              QEENITRALIVVQQGMTPSAKQSLVDMA-PKYILEQFLEQELLINITEHE
Dm_RPB5_NP_610630             QEENIHRAIVVVQGGMTPSAKQSLVDMA-PKYILEQFLESELLINITEHE
Ce_RPB5_Q9N5K2                QEQNISRAIIVVQTGMTPSAKQSIGDMA-PKYMLEHFLEAELMVNITEHE
Populus_trichocarpa_584052    KSENVFRAILVVQQNLTPFARTCINEIS-TKFHLEVFQEAELLVNIKEHV
Vitis_vinifera_CAO65489       KSENVFRAILVVQQNLTPFARTCINEIS-TKFHLEVFQEAELLVNIKEHV
NRPB5/NRPD5_At3g22320         KSENVFRAILVVQQNLTPFARTCISEIS-SKFHLEVFQEAEMLVNIKEHV
Medicago_truncatula_ABO78350  NSENVYRAILVCQTSLTPFAKTCVSEIA-SKFHLEVFQEAELLVNIKEHV
Oryza_sativa_EAZ13876         KAENVSRAVLVLQQNLTPFARSFLQELE-PKIHLEIFQEAELLINIKEHV
Oryza_sativa_NP_001044564     KAENVSRAVLVLQQNLTPFARSFLQELE-PKIHLEIFQEAELLINIKEHV
Zea_mays_ACF85599             THENVSRAVLVLQQNLTPFAKSFLIELE-PKIHLEIFQEAEMLINIKEHV
Oryza_sativa_CAD41325         KQESVFNGILVVQQALSAFARSAVQEVS-QKFHLEVFQEAELLVNIKDHT
Oryza_sativa_EAZ31161         KQESVFNGILVVQQALSAFARSAVQEVS-QKFHLEVFQEAELLVNIKDHT
Zea_mays_ACF81264             KNENVFAGILVVQQALSAFARSAVQEVS-QKYHLEVFQEAELLVNIKDHV
Physcomitrella_patens_206246  KTENVHRAILVVQQNLTPFARQCVSEMA-SKYHLEVFQEAELLVNIKEHV
Physcomitrella_patens_55574   KTENVHRAILVVQQNLTPFARQCVSEMS-SKYHVEVFQEAELLVNIKDHV
Physcomitrella_patens_231299  KRENVPRAVFVVQQHITPLSKQYISRKA-QKYHLEVFLEPEFLVNITECY
Physcomitrella_patens_136486  KRENVPRAVFVVQQHITPLSKQYISRKA-QKYHLEVFLEPEFLVNITECY
Sc_RPB5_CAA85113              QEKNFQTGIFVYQNNITPSAMKLVPSIP--PATIETFNEAALVVNITHHE
NRPB5-like_At5g57980          RDDKVHRGIVVVPMAITAPARMAVSELN-KMLTIEVFEEAELVTNITEHK
NRPE5-like_At2g41340          SRENITGLILVLQSHITNQALKAV-ELF--SFKVELFEITDLLVNVSKHV
Brassica_napus_AAF81222       GRENLTGLILVLQSDITNQALKAV-ELF--SFKVELFQLTELLVNITKHV
NRPE5_At3g57080_NP_191267     SQETITGLILVLQNHVTNQALKAI-ELF--SFKVEIFQITDLLVNITKHS
Populus_trichocarpa_57931     DRDSLTGLILVLQNNITNQAMKAL-DLF--KFKIEIFQITDLLVNITKHI
Vitis_vinifera_CAO63075       NKDSLSKLILVLQNHITSQALKAV-DLF--SFQVEKFQITDLLVNITKHV
Medicago_truncatula_ABN07995  NRESLTGLILIVQNQITSQALKAV-NLL--SFKVEIFQITDLLVNATKHV
Populus_trichocarpa_48513     NKESLHGLILILQSKMNHFAKKEL-EKF--PFKVEVFQITDLLVNITKHV
Vitis_vinifera_CAO42914       NREGLHRLILVLQSKMNSHARKVV-DEY--PIKVEFFQITELLINITKHV
NRPE5-like_At3g54490          NNVGLHAMILVLQSKMNHFAQKAL-TTF--PFTVETFPIEDLLVNITKHI
Medicago_truncatula_ABD28306  DKERLSRLILVMQSKMTSYARKEL-ENC--PFKVEIIQLNDLLVNVTKHV
Oryza_sativa_NP_001065723     E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Oryza_sativa_NP_001066119     E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Oryza_sativa_EAY79909         E-ENLSRLVLILQSKILSRAREAIKEIF--KFKVDIFQATDLLVNITKHV
Zea_mays_ACF87172             D-ENLSRLILILQGKIMSTTRESIKEIF--RFKVDTFQITELLVNITKHV
consensus                       ENV RAILVVQQ IT  AR  V EL   KF LEVFQE ELLVNITEHV
                              Start of Assembly domain
C_reinhardtii_XP_001697601    LVPEHRILSPDEKRTLLDRYKIKETQ-------------LPRIQASDAVA
O_lucimarinus_XP_001417617    LVPEHILLSDDQKRTLLDRYKVKDTQ-------------LPRIQMHDPIA
Hs_RPB5_BAA07406              LVPEHVVMTKEEVSELLARYKLRENQ-------------LPRIQAGDPVA
Dm_RPB5_NP_610630             LVPEHVVMTVEEKQELLSRYKLKENM-------------LMRIQAGDPVA
Ce_RPB5_Q9N5K2                LVPEHVVMTAEEKAELLARYKLKDSQ-------------LPRIQQCDPVA
Populus_trichocarpa_584052    LVPEHQVLSNEEKKTLLERYTVKETQ-------------LPRIQITDPIA
Vitis_vinifera_CAO65489       LVPEHQVLTSEEKKTLLERYTVKETQ-------------LPRIQVSDPIA
NRPB5/NRPD5_At3g22320         LVPEHQVLTTEEKKTLLERYTVKETQ-------------LPRIQVTDPIA
Medicago_truncatula_ABO78350  LVPEHQILNDTEKKTLLERYTVKETQ-------------LPRIQVTDPVA
Oryza_sativa_EAZ13876         LVPEHQVLNNGEKKTLLERYTLKETQVYIHDHMLGEIIFLRRSHVNDPMA
Oryza_sativa_NP_001044564     LVPEHQVLNNEEKKTLLERYTLKETQ-------------LPRIQITDPIA
Zea_mays_ACF85599             LVPEHQVLTNEEKKTLLERYTLKETQ-------------LPRIQITDPIA
Oryza_sativa_CAD41325         LVPEHELLTPEQKKTLLERYTVKETQ-------------LPRIQITDPIA
Oryza_sativa_EAZ31161         LVPEHELLTPEQKKTLLERYTVKETQILSLTQLV-KCVNLPRIQITDPIA
Zea_mays_ACF81264             LVPEHVLLTPEDKKTLLERYTVKETQ-------------LPRIQITDPIA
Physcomitrella_patens_206246  LVPLHEVLTPDEKKTLLERYTVKET-------------QLPRMQENDPVA
Physcomitrella_patens_55574   LVPQHEVLNAEEKITLLQRYTVKET-------------QLPRMQENDPVA
Physcomitrella_patens_231299  LVPLHEILTPEEKNTLLERYTEGNPVML---------VLLPWMQHNDPVA
Physcomitrella_patens_136486  LVPLHEILTPEEKNTLLERYTEGNP--------------LPWMQHNDPVA
Sc_RPB5_CAA85113              LVPKHIRLSSDEKRELLKRYRLKESQ-------------LPRIQRADPVA
NRPB5-like_At5g57980          LVNKYYVLDDQAKKKLLNTYTVQDTQ-------------LPRILVTDPLA
NRPE5-like_At2g41340          LRPKHQVLNDKEKESLLKKFSIEEKQ-------------LPRLSSKDPIV
Brassica_napus_AAF81222       LRPKHHVLNEQEKESLFKKFSIQEQQ-------------LPKLLKKDPTA
NRPE5_At3g57080_NP_191267     LKPQHQVLNDEEKTTLLKKFSIEEKQ-------------LPRISKKDAIV
Populus_trichocarpa_57931     LKPKHQVLSEQAKQRLLKKYSIEEKQ-------------LPRLLKKDAIS
Vitis_vinifera_CAO63075       LKPKHRVLTDQEKNKLLKKYSLNEKQ-------------LPRMLQQDAIA
Medicago_truncatula_ABN07995  LKPKHQVLTDKQKKNLLKKYDIQEKQ-------------LPRMLQTDAIA
Populus_trichocarpa_48513     LQPQMDILTAEQKQQVMNKYKLEDKQ-------------LPRMLESDAIV
Vitis_vinifera_CAO42914       SVPKHEILSAQEKRKLVNKYKLEDKQ-------------FPIMQKDDAIA
NRPE5-like_At3g54490          QQPKIEILNKEEKEQLLRKHALEDKQ-------------LPYLQEKDSFV
Medicago_truncatula_ABD28306  LQPKYEVLTANEKQKLLNKYKVEEKQ-------------LPHMLRTDAIA
Oryza_sativa_NP_001065723     LKPKHEVLSADQKAKLLKEYNVEDSQ-------------LPRMLETDAVA
Oryza_sativa_NP_001066119     LKPKHEVLSADQKAKLLKEYNVEDSQ-------------LPRMLETDAVA
Oryza_sativa_EAY79909         LKPKHEVLSADQKAKLLKEYNVEDSQ-------------LPRMLETDAVA
Zea_mays_ACF87172             LKPKHEVLTAEGKAKLLKEYNVVDSQ-------------LPRMLENDAVA
consensus                     LVP H VLT EEK TLL RYTVKETQ             LPRIQ  DPIA
```

```
C_reinhardtii_XP_001697601   RYLG--LQRGQVVRIVRP-SETAGRYVTYRFCPPLWR
O_lucimarinus_XP_001417617   RYYG--MRRGQVVRIIRP-SETAGRYVTYRLCV----
Hs_RPB5_BAA07406             RYFG--IRRGQVVKIIRP-SETAGRYITYRLVQ----
Dm_RPB5_NP_610630            RYFG--LKRGQVVKIIRS-SETAGRYISYRLVC----
Ce_RPB5_Q9N5K2               RYFG--LRRGQVVKIIRP-SETAGRYITYRLVV----
Populus_trichocarpa_584052   RYYG--LKRGQVVKIIRP-SETAGRYVTYRYVI----
Vitis_vinifera_CAO65489      RYFG--LKRGQVVKIIRP-SETAGRYITYRYVV----
NRPB5/NRPD5_At3g22320        RYFG--LKRGQVVKIIRP-SETAGRYVTYRYVV----
Medicago_truncatula_ABO78350 RYYG--LKRGQVVKIIRP-SETAGRYVTYRFVV----
Oryza_sativa_EAZ13876        VIVGNLNYLSHIQLAIAPNMSTYGKYCMEAGLVP---
Oryza_sativa_NP_001044564    RYYG--LRRGQVVKIIRP-SETAGRYVTYRYVV----
Zea_mays_ACF85599            RYYG--LRRGQVVKIIRP-SETAGRYVTYRYVV----
Oryza_sativa_CAD41325        RYYG--MKRGQVVKIIRA-SETAGRYVTYRYVV----
Oryza_sativa_EAZ31161        RYYG--MKRGQVVKIIRA-SETAGRYVTYRYVV----
Zea_mays_ACF81264            RYYG--MKRGQVVKITRA-SETAGRYITYRYVV----
Physcomitrella_patens_206246 RYYG--LKRGQVVKIIRP-SETAGRYVTYRFVV----
Physcomitrella_patens_55574  RYYG--LKRGQVVKIIRP-SETAGRYVTYRFVV----
Physcomitrella_patens_231299 RYYG--INPGQVVKIIQS-SETAGRYVTYRLFV----
Physcomitrella_patens_136486 RYYG--INPGQVVKIIQS-SETAGRYVTYRLFV----
Sc_RPB5_CAA85113             LYLG--LKRGEVVKIIRK-SETSGRYASYRICM----
NRPB5-like_At5g57980         RYYG--LKRGQVVKIRRS-DATSLDYYTYRFAV----
NRPE5-like_At2g41340         RYYG--LETGQVMKVTYKDELSES-HVTYRCVS----
Brassica_napus_AAF81222      KYYG--LEKGQVVEVTYKGEGSESDHVSYRCAW----
NRPE5_At3g57080_NP_191267    RYYG--LEKGQVVKVNYRGELTES-HVAFRCVW----
Populus_trichocarpa_57931    RYYG--LERGQVVKVTYDGDITGS-HVTYRCVW----
Vitis_vinifera_CAO63075      RYYG--LEKGQVVKVIYNGEITGS-HVTYRCVW----
Medicago_truncatula_ABN07995 RYYG--LQRGQVVKVTYTGEITQM-HVTYRCVW----
Populus_trichocarpa_48513    QYYG--LQKGQMVKITYSGEIVDH-LVTYRCVT----
Vitis_vinifera_CAO42914      RYYG--LEKGQVVKITYKGGMTDS-LVTYRCVS----
NRPE5-like_At3g54490         RYYG--LKKKQVVKITYSKEPVGD-FVTYRCII----
Medicago_truncatula_ABD28306 SYYG--LEKGQVVKISHSGEMFNS-LVMYRCVV----
Oryza_sativa_NP_001065723    RYYG--FDKGTVVKVIYDGELTGK-RVAYRCVF----
Oryza_sativa_NP_001066119    RYYG--FDKGTVVKVTYDGELTGK-RVAYRCVF----
Oryza_sativa_EAY79909        RYYG--FDKGTVVKVIYDGELTGK-RVAYRCVF----
Zea_mays_ACF87172            RYYG--LGKGTVVKVIYDSELTGN-HVTYRCIT----
consensus                    RYYG  LKRGQVVKIIR  SETAGRYVTYR VV
```

Figure S17. The N-terminal extension of NRPE5 is required for the protein's stability and function.

A. Diagram highlighting the jaw and assembly domains and the short N-terminal extension present in NRPE5 but absent in NRPB5/NRPD5. Underlined amino acids were deleted in the *35S:FLAG-ΔN-NRPE5* transgene.

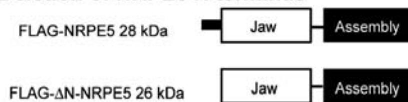
B. *AtSN1* retrotransposon expression in Pol V mutants, wild-type, and *35S:FLAG-ΔN-NRPE5 nrpe5* lines assayed by strand-specific RT-PCR.

C. *AtSN1* methylation in *35S:FLAG-NRPE5 nrpe5*, *35S:FLAG-ΔN-NRPE5 nrpe5* lines and Pol V mutants compared to wild-type.

D. Methylation-sensitive Southern blot analysis of 5S rRNA genes in Pol V mutants, wild-type, and *35S:FLAG-ΔN-NRPE5 nrpe5* lines.

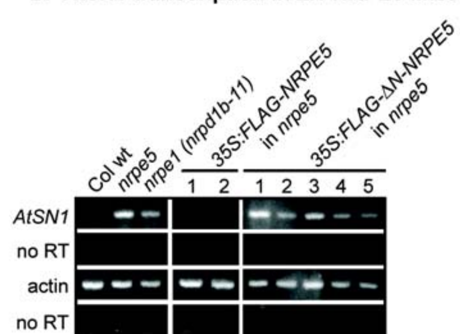
E. RT-PCR and immunoblot analysis of mRNA and protein levels in T2 generation plants of *35S:FLAG-NRPE5 nrpe5* and *35S:FLAG-ΔN-NRPE5 nrpe5* lines. The upper panels show RT-PCR reactions, including actin and no reverse transcriptase (no RT) controls. In the bottom panel, equal amounts of tissue homogenate were subjected to anti-FLAG IP and immunoblot detection of the tagged proteins.

A. Deletion of NRPE5 N terminus

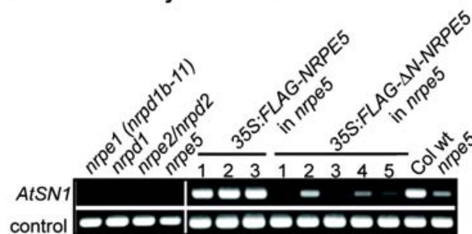


MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDE
DINLSLHDFRTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAI
RSVVADILSQETITGLILVLQNHVTNQALKAIELFSFKVEIFQITDLLVNI
TKHSLKPQHQVLNDEEKTTLLKKFSIEEKQLPRISKKDAIVRYYGLEKGQ
VVKVNYRGELTESHVAFRCVW

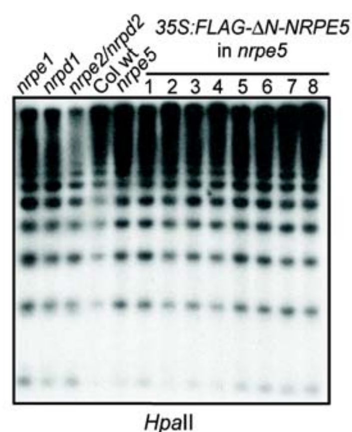
B. *AtSN1* transcription in ΔN-NRPE5 lines



C. *AtSN1* methylation in ΔN-NRPE5 lines



D. 5s rDNA methylation in ΔN-NRPE5 lines



E. RNA and protein levels in ΔN-NRPE5 lines

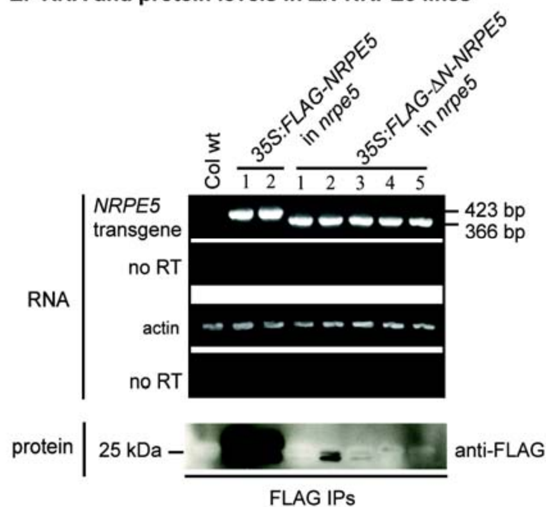


Figure S18. Peptide coverage maps of RNA polymerase subunits detected by LC-MS/MS in affinity purified FLAG-NRPE5 samples. In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPE1 (At2g40030)

MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGAT
EPDKCEGHFGYIQLPVPIYHPAHVNELKQMLSLLCLKCLKIKKAKGTSGGLADRLLGVC
CEEASQISIKDRASDGASYLELKLPSRSRLQPGCWNFLERYGYRYGSDYTRPLLAREVKE
ILRRIPEESRKKLTAKGHIPQEGYILEYLPVPPNCLSVPEASDGFSTMSVDPSRIELKDVLK
KVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAKAARNIDMRYGVSKISDSSSS
KAWTEKMRTLFIRKGSGFSSRSVITGDAYRHVNEVGIPIEIAQRITFEERVSVHNRGYLQ
KLVDDKLCLSYTQGSTTYSLRDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHS
LQALRVYVHEDNTVKINPLMCSPLSADFDGDCVHLFYPQSLSAKAEVMELFSVEKQLLS
SHTGQLILQMGSDSLLSLRVMLERVFLDKATAQQLAMYGSLSLPPPALRKSSKSGPAWT
VFQILQLAFPERLSCKGDRFLVDGSDLLKFDFGVDAMGSIINEIVTSIFLEKGPKETLGFFD
SLQPLLMESLFAEGFSLSLEDLSMSRADMDVIHNLIIREISPMVSRLRLSYRDELQLENSIH
KVKEVAANFMLKSYSIRNLIDIKSNSAITKLVQQTGFLGLQLSDKKKFYTKTLVEDMAIF
CKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKNLMA
VLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKA
VLDSSPNSNSSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENAACTVRN
KLNKVSLKDTAVEFLVEYRKQPTISEIFGIDSCLHGHIHLNKTLLQDWNISMQDIHQKCE
DVINSLGQKKKKKATDDFKRTSLSVSECCSFRDPCGSKGSDMPCLTFSYNATDPDLERT
LDVLCNTVYPVLLEIVIKGDSRICSANIIWNSSDMTTWIRNRHASRRGEWVLDVTVEKSA
VKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRLSASVRMV
SKGVLKEHIILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKAAE
KCHTDSLSTVVGSCSWGKRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTT
NADAFVSSPGFDVTEEEMAEWAESPERDSALGEPKFEDSADFQNLHDEGKPSGANWEK
SSSWDNGCSGGSEWGVSKSTGGEANPESNWEKTTNVEKEDAWSSWNTRKDAQESSKS
DSGGAWGIKTKDADADTTPNWETSPAPKDSIVPENNEPTSDVWGHKSVSDKSWDKKN
WGTESAPAAWGSTDAAVWGSSDKKNSETESDAAAWGSRDKNNSDVGSGAGVLGPWN
KKSSETESNGATWGSSDKTKSGAAAWNSWDKKNIETDSEPAAWGSQGKKNSETESGP
AAWGAWDKKKSETEPGPAGWGMGDKKNSETELGPAAMGNWDKKKSDTKSGPAAWG
STDAAAWGSSDKNNSETESDAAAWGSRNKKTSEIESGAGAWGSWGQPSPTAEDKDTN
EDDRNPWVSLKETKSREKDDKERSQWGNPAKKFPSSGGWSNGGGADWKGNRNHTPR
PPRSEDNLAPMFTATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPISDDDK
TFVLEKILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQDFSYRKSLNNY
LMKKYPDRAEEFIDKYFTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQT

QSQSPSQTRAQSPSQAQAQSPSQTQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSPSQ
TQTQSPSQTQAQAQSPSSQSPSQTQT

Notes:
427/1976 amino acids are represented by sequenced peptides =22% coverage.
All peptides are specific to NRPE1 (NRPD1b), meaning that none are identical to any other protein, including NRPD1 (NRPD1a).

NRPE2/NRPD2 (At3g23780)
MPDMDIDVKDLEEFEATTGEINLSELGEGFLQSFCKKAATSFFDKYGLISHQLNSYNYFI
EHGLQNVFQSFGEMLVEPSFDVVKKKDNDWRYATVKFGEVTVEKPTFFSDDKELEFLP
WHARLQNMTYSARIKVNVQVEVFKNTVVKSDKFKTGQDNYVEKKILDVKKQDILIGSI
PVMVKSILCKTSEKGKENCKKGDCAFDQGGYFVIKGAEKVFIAQEQMCTKRLWISNSP
WTVSFRSENKRNRFIVRLSENEKAEDYKRREKVLTVYFLSTEIPVWLLFFALGVSSDKEA
MDLIAFDGDDASITNSLIASIHVADAVCEAFRCGNNALTYVEQQIKSTKFPPAESVDECL
HLYLFPGLQSLKKKARFLGYMVKCLLNSYAGKRKCENRDSFRNKRIELAGELLEREIRV
HLAHARRKMTRAMQKHLSGDGDLKPIEHYLDASVITNGLSRAFSTGAWSHPFRKMERV
SGVVANLGRANPLQTLIDLRRTRQQVLYTGKVGDARYPHPSHWGRVCFLSTPDGENCG
LVKNMSLLGLVSTQSLESVVEKLFACGMEELMDDTCTPLFGKHKVLLNGDWVGLCAD
SESFVAELKSRRRQSELPREMEIKRDKDDNEVRIFTDAGRLLRPLLVVENLQKLKQEKPS
QYPFDHLLDHGILELIGIEEEDCNTAWGIKQLLKEPKIYTHCELDLSFLLGVSCAVVPFA
NHDHGRRVLYQSQKHCQQAIGFSSTNPNIRCDTLSQQLFYPQKPLFKTLASECLKKEVLF
NGQNAIVAVNVHLGYNQEDSIVMNKASLERGMFRSEQIRSYKAEVDAKDSEKRKKMD
ELVQFGKTHSKIGKVDSLEDDGFPFIGANMSTGDIVIGRCTESGADHSIKLKHTERGIVQK
VVLSSNDEGKNFAAVSLRQVRSPCLGDKFSSMHGQKGVLGYLEEQNFPFTIQGIVPDI
VINPHAFPSRQTPGQLLEAALSKGIACPIQKEGSSAAYTKLTRHATPFSTPGVTEITEQLH
RAGFSRWGNERVYNGRSGEMMRSMIFMGPTFYQRLVHMSEDKVKFRNTGPVHPLTRQ
PVADRKRFGGIKFGEMERDCLIAHGASANLHERLFTLSDSSQMHICRKCKTYANVIERTP
SSGRKIRGPYCRVCVSSDHVVRVYVPYGAKLLCQELFSMGITLNFDTKLC

Notes:
281/1172 amino acids represented in sequenced peptides =24% coverage.
72/1172= 6% coverage is accounted for by peptides unique to NRPE2/NRPD2a. The remaining 18% of the peptides match NRPE2/NRPD2a as well as the NRPD2b pseudogene. However, the latter gene is non-functional, and no peptides that would uniquely identify NRPD2b were detected.

NRPE3a/NRPD3/NRPB3 (At2g15430)
MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEIEVNSS
VLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCSVEFRLSSKCVTDQTLD
VTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRRGQELKLRAIARKGIGKDHAKW
SPAATVTFMYEPDIIINEDMMDTLSDEEKIDLIESSPTKVFGMDPVTRQVVVVDPEAYTY
DEEVIKKAEAMGKPGLIEISPKDDSFIFTVESTGAVKASQLVLNAIDLLKQKLDAVRLSD
DTVEADDQFGELGAHMRGG

Notes:
155/319 amino acids are represented by sequenced peptides =48% coverage

115/319=36% unique coverage. 36% of the coverage corresponds to peptides that match only NRPE3a. The other 12% matches either NRPE3a or NRPE3b.

NRPE3b (At2g15400)
MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVKIEVNS
SVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSVEFPLSAKCVTDQTLD
VTSRDLYSADPTVTPVDFTSNSSTSDSSEHKGIIIAKLRRGQELKLKALARKGIGKDHAK
WSPAATVTYMYEPDIIINEEMMNTLTDEEKIDLIESSPTKVFGIDPVTGQVVVVDPEAYT
YDEEVIKKAEAMGKPGLIEIHPKHDSFVFTVESTGALKASQLVLNAIDILKQKLDAIRLSD
NTVEADDQFGELGAHMREG

Notes:
53/319 amino acids are represented by sequenced peptides = 16% coverage
13/319=4% coverage corresponds to peptides matching only NRPE3b, whereas the remaining 12% of the coverage matches either NRPE3b or NRPE3a.

NRPE4/NRPD4 (At4g15950)
MSEKGGKGLKSSLKSKDGGKDGSSTKLKKGRKIHFDQGTPPANYKILNVSSDQQPFQSS
AAKCGKSDKPTKSSKNSLHSFELKDLPENAECMMDCEAFQILDGIKGQLVGLSEDPSIKI
PVSYDRALAYVESCVHYTNPQSVRKVLEPLKTYGISDGEMCVIANASSESVDEVLAFIPS
LKTKKEVINQPLQDALEELSKLKKSE

17/205 amino acids are represented by sequenced peptides=8% coverage. All peptides sequenced match only At4g15950 and no other RPB4-like protein.

NRPB4 (At5g09920)

MSGEEEENAAELKIGDEFLKAKCLMNCEVSLILEHKFEQLQQISEDPMNQVSQVFEKSL
QYVKRFSRYKNPDAVRQVREILSRHQLTEFELCVLGNLCPETVEEAVAMVPSLKTKGRA
HDDEAIEKMLNDLSLVKRFE

0/138 amino acids are represented by sequenced peptides=0% coverage. No peptides were identified that matched this protein sequence.

NRPB5/NRPD5 (formerly AtRPB5a, AtRPB24.3) (At3g22320)
MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKA
KRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEIS
SKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARY
FGLKRGQVVKIIRPSETAGRYVTYRYVV

0/205 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5 (formerly AtRPB5b, AtRPB23.7)(At3g57080)
MEVKGKETASVLCLSKYVDLSSEESHRYYLARRNGLQMLRDRGYEVSDEDINLSLHDF
RTVYGERPDVDRLRISALHRSDSTKKVKIVFFGTSMVKVNAIRSVVADILSQETITGLILV

LQNHVTNQALKAIELFSFKVEIFQITDLLVNITKHSLKPQHQVLNDEEKTTLLKKFSIEEK
QLPRISKKDAIVRYYGLEKGQVVKVNYRGELTESHVAFRCVW

145/222 amino acids are represented by sequenced peptides = 65% coverage
All peptides identified correspond to peptides that match NRPE5 only and no other family member.

NRPB5-like family member (synonym AtRPB5c) (At5g57980)
MSDMDDEITRIFKVRRTVLQMLRDRGYTIEESDLNLKREEFVQRFCKTMNKVNKEALF
VSANKGPNPADKIYVFYPEGPKVGVPVIKKEVAIKMRDDKVHRGIVVVPMAITAPARM
AVSELNKMLTIEVFEEAELVTNITEHKLVNKYYVLDDQAKKKLLNTYTVQDTQLPRILV
TDPLARYYGLKRGQVVKIRRSDATSLDYYTYRFAV

0/210 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (synonym AtRPB5d) (At2g41340)
MEGKGKEIVVGHSISKSSVECHKYYLARRTTMEMLRDRGYDVSDEDINLSLQQFRALY
GEHPDVDLLRISAKHRFDSSKKISVVFCGTGIVKVNAMRVIAADVLSRENITGLILVLQS
HITNQALKAVELFSFKVELFEITDLLVNVSKHVLRPKHQVLNDKEKESLLKKFSIEEKQL
PRLSSKDPIVRYYGLETGQVMKVTYKDELSESHVTYRCVS

0/218 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE5-like family member (At3g54490)
MEETMAEEGCCENVESTFDDGTNCISKTEDTGGIESKRFYLARTTAFEMLRDRGYEVNE
AELSLTLSEFRSVFGEKPELERLRICVPLRSDPKKKILVVFMGTEPITVKSVRALHIQISNN
VGLHAMILVLQSKMNHFAQKALTTFPFTVETFPIEDLLVNITKHIQQPKIEILNKEEKEQL
LRKHALEDKQLPYLQEKDSFVRYYGLKKKQVVKITYSKEPVGDFVTYRCII

0/233 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPB5 family member (likely pseudogene) (At3g16880)
MKKYIDQLKSANVFRAILVVQDIKAFSRQALVFLGAVYPIFHIEVFQEKELIVNVKEHVF
VPEHQALTTEEKQKFLERKRTSFQGFT

0/87 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence. This protein is truncated relative to the other NRPB5-like proteins and likely is a pseudogene.

NRPE6a/NRPD6a/NRPB6a (At5g51940)
MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDGSFEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage

0/144 = 0% coverage corresponds to peptides that are NRPE6a-specific, the sequenced peptide also matches At2g04630.

NRPE6b/NRPD6b/NRPB6b (At2g04630)
MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRP
RKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYL
PDMSYEEWGVDELIVEDSWKRQVGGD

26/144 amino acids are represented by sequenced peptides = 18% coverage
0/144=0% of the coverage corresponds to peptides unique to this member of the protein family; the sequenced peptide also matches an identical sequence of At5g51940.

NRPE7 (At4g14660)
MFLKVQLPWNVMIPAENMDAKGLMLKRAILVELLEAFASKKATKELGYYVAVTTLDKI
GEGKIREHTGEVLFPVMFSGMTFKIFKGEIIHGVVHKVLKHGVFMRCGPIENVYLSYTK
MPDYKYIPGENPIFMNEKTSRIQVETTVRVVVIGIKWMEVEREFQALASLEGDYLGPLSE

0/177 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that match this protein sequence.

NRPB7 (At5g59180)
MFFHIVLERNMQLHPRFFGRNLKENLVSKLMKDVEGTCSGRHGFVVAITGIDTIGKGLIR
DGTGFVTFPVKYQCVVFRPFKGEILEAVVTLVNKMGFFAEAGPVQIFVSKHLIPDDMEF
QAGDMPNYTTSDGSVKIQKECEVRLKIIGTRVDATAIFCVGTIKDDFLGVINDPAAA

0/176 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that match this protein sequence.

NRPD7 (At3g22900)
MFIKVKLPWDVTIPAEDMDTGLMLQRAIVIRLLEAFSKEKATKDLGYLITPTILENIGEGK
IKEQTGEIQFPVVFNGICFKMFKGEIVHGVVHKVHKTGVFLKSGPYEIIYLSHMKMPGYE
FIPGENPFFMNQYMSRIQIGARVRFVVLDTEWREAEKDFMALASIDGDNLGPF

0/174 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPB7 family member (At4g14520)
MFSEVEMARDVAICAKHLNGQSPHQPILCRLLQDLIHEKACREHGFYLGITALKSIGNNK
NNNIDNENNHQAKILTFPVSFTCRTFLPARGDILQGTVKKVLWNGAFIRSGPLRYAYLSL
LKMPHYHYVHSPLSEDEKPHFQKDDLSKIAVGVVVRFQVLAVRFKERPHKRRNDYYVL
ATLEGNGSFGPISLTGSDEPYM

0/200 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

NRPE8a/NRPD8a/NRPB8a (At1g54250)
MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAVGDKF
TLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLKGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by sequenced peptides = 9% coverage
0/146=0% of the coverage corresponds to peptides unique to this member of the protein family. This peptide also is an exact match to At3g59600.

NRPE8b/NRPB8b/NRPD8b (At3g59600)
MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAVGDKF
TLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTPKAELYVSFG
GLLMLLQGDPAHISHFELDQRLFLLMRKL

13/146 amino acids are represented by sequenced peptides = 9% coverage
0/146 = 0% of the coverage corresponds to peptides unique to this member of the protein family. This peptide is also an exact match to At1g54250.

NRPE9a/NRPD9a/NRPB9a (At3g16980)
MSTMKFCRECNNILYPKEDKEQKILLYACRNCDHQEVADNSCVYRNEVHHSVSERTQIL
TDVASDPTLPRTKAVRCSKCQHREAVFFQATARGEEGMTLFFVCCNPNCGHRWRE

10/114 amino acids are represented by sequenced peptides = 9% coverage
0/114 = 0% coverage corresponds to peptides unique to this member of the protein family. Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from At4g16265.

NRPE9b/NRPD9b/NRPB9b (At4g16265)
MSTMKFCRECNNILYPKEDKEQSILLYACRNCDHQEAADNNCVYRNEVHHSVSEQTQI
LSDVASDPTLPRTKAVRCAKCQHGEAVFFQATARGEEGMTLFFVCCNPNCSHRWRE

10/114 amino acids are represented by sequenced peptides = 9% coverage
0/114 = 0% coverage corresponds to peptides unique to this member of the protein family. Two amino acid differences in the identified peptide (underlined) discriminates At3g16980 from At4g16265.

NRPE10/NRPB10/NRPD10 (At1g11475)
MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHVDLIEK
LLNYNTLEKSDNS

20/71 amino acids are represented by sequenced peptides = 28% coverage
20/71= 28% coverage corresponds to peptides that only match this protein and not At1g61700.

NRPB10 family member (At1g61700)
MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTHVDLIE
KLLNYNTMEKSDPN
11/71 amino acids are represented by sequenced peptides = 15% coverage
0/71= 0% unique. The peptide identified for At1g61700 also matches At1g11475.

NRPE11/NRPB11/NRPD11 (At3g52090)
MNAPERYERFVVPEGTKKVSYDRDTKIINAASFTVEREDHTIGNIVRMQLHRDENVLFA
GYQLPHPLKYKIIVRIHTTSQSSPMQAYNQAINDLDKELDYLKNQFEAEVAKFSNQF

42/116 amino acids are represented by sequenced peptides = 36% coverage
All peptides identified match NRPE11 and only NRPE11.

NRPE12/NRPB12/NRPD12 (At5g41010)
MDPAPEPVTYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTRRVVQYEAR

8/51 amino acids are represented by the sequenced peptide = 16% coverage
The peptide is a unique match to this protein.

NRPB12 family member (At1g53690)
MDLQQSETDDKQPEQLVIYVCGDCGQENILKRGDVFQCRDCGFRILYKKRILDKKETRI
GV
0/62 amino acids are represented by sequenced peptides = 0% coverage
No peptides were identified that matched this protein sequence.

CHAPTER FOUR

SUBUNIT COMPOSITIONS OF *ARABIDOPSIS* RNA POLYMERASES I AND III REVEAL INSIGHTS INTO THE EVOLUTION, FUNCTIONAL DIVERSIFICATION AND REDUNDANCY OF SUBUNITS AMONG ALL FIVE DNA-DEPENDENT RNA POLYMERASES

A manuscript in preparation

My contributions to this work:

In this work, I designed and performed all of the experiments, wrote the manuscript and contributed significantly to the intellectual and physical aspects of the work. Jeremy R. Haag assisted with cloning the *NRPA3*, *NRPC3* and *NRPA11* genomic fragments. Carrie D. Nicora and Angela D. Norbeck performed mass spectrometry analyses under the supervision of Ljiljana Pasa-Tolic at Pacific Northwest National Laboratories. Craig S. Pikaard helped design and interpret experiments and helped edit the manuscript.

Subunit compositions of *Arabidopsis* RNA polymerases I and III reveal insights into the evolution, functional diversification and redundancy of subunits among all five DNA-dependent RNA polymerases

Thomas S. Ream[1], Carrie D. Nicora[2], Angela D. Norbeck[2], Jeremy R. Haag[1], Ljiljana Paša-Tolić[2] and Craig S. Pikaard[1*]

[1] Biology Department, Washington University, St. Louis, MO, 63130
[2] Pacific Northwest National Laboratory, Richland, WA, 99352

* Corresponding author: pikaard@biology2.wustl.edu,
phone 314-935-7569, fax 314-935-4432

Running title: Subunit compositions of *Arabidopsis* DNA-dependent RNA polymerases

Abstract

Eukaryotes contain three essential DNA-dependent RNA polymerases that transcribe the genome into RNAs, namely Pol I, II and III, as well as two non-essential, plant-specific RNA polymerases, Pol IV and Pol V. Subunit compositions of Pol II, Pol IV and Pol V were recently determined. By contrast, Pol I and Pol III have been purified from plants but their complete subunit compositions are unknown. In *Arabidopsis*, numerous RNA polymerase subunits that are encoded by single genes in yeast or humans are present in multi-gene families. Some of this diversity among Pol II-related subunit families is explained by the emergence of Pol IV and V. However, multiple putative homologs of the yeast Pol III-specific subunits Rpc11, Rpc17, Rpc31 and Rpc53 are also present in *Arabidopsis*, as are two homologs of the yeast Pol I and Pol III subunit, Rpac40. We purified Pol I and Pol III from *Arabidopsis* and determined their subunit compositions using mass spectrometry and co-immunoprecipitation. Our results indicate that the Pol I and Pol III subunit compositions are conserved between yeast and plants. However, Pol I and Pol III utilize distinct, non-redundant subunits that are homologs of the single Rpac40 in yeast. *Arabidopsis* Pol III uniquely utilizes two alternative Rpc53-related variants. The identification of Pol I, II, III, IV and V subunits completed with this study reveals examples of functional diversification and cases of subunit redundancy among RNA polymerases in plants.

(233 words)

Introduction

DNA-dependent RNA polymerases are multi-subunit enzymes that serve as the catalytic engines that synthesize RNA from DNA. Eukaryotes contain three essential polymerases, namely RNA polymerases I, II and III. Pol I transcribes the precursors of the ribosomal RNA (rRNA) genes (Grummt 2003). Pol II transcribes messenger RNAs (mRNAs), and non-coding RNAs including small nuclear RNAs (snRNAs) and microRNAs (miRNAs) (Woychik and Hampsey 2002). Pol III transcribes 5S rRNA gene precursors, transfer RNAs (tRNAs) and some snRNAs such as U6 and 7SL RNAs (Schramm and Hernandez 2002).

Eukaryotic RNA polymerase II is comprised of twelve subunits, yeast Pol I has fourteen subunits and yeast and human Pol III contain seventeen subunits (Cramer et al. 2001; Huang and Maraia 2001; Hu et al. 2002; Jasiak et al. 2006; Fernandez-Tornero et al. 2007; Kuhn et al. 2007). Each of these polymerases contains subunits that are homologs of the five subunit *E. coli* RNA polymerase, composed of ß', ß, ω, and two copies of α. The ß' and ß subunits interact to form the catalytic center, whereas the α homodimer and ω subunit promote polymerase assembly. The largest and second-largest catalytic subunits of Pol I, II and III are the eukaryotic homologs of the ß' and ß prokaryotic subunits, respectively.

In yeast, the three RNA polymerases are referred to as Rpa, Rpb or Rpc, which stands for RNA polymerase I (A), II (B) or III (C). Subunits of Pol II are named in descending order by size using numbers 1-12 and are given the letter designation of the polymerase with which they associate. In other yeast nomenclatures, subunits are named based on their apparent molecular weight, especially for Pol I- and Pol III-specific

subunits. In cases where subunits are shared by more than one polymerase, they are either referred to in the context of Pol II or given multiple letter designations. For example, the *E. coli* ω-like subunit in yeast is referred to as Rpb6 or ABC23, signifying this subunit is the sixth largest subunit of Pol I, II and III with an apparent molecular weight of 23 kDa. In referring to yeast subunits in this paper, we use the Rpb1-Rpb12 nomenclature when referring to Pol II subunits, or subunits shared by Pol II and other polymerases. For Pol I or III subunits, we use the nomenclature based on apparent molecular weight.

An Rpb3-Rpb11 heterodimer in Pol II and an Rpac40-Rpac19 heterodimer in both Pol I and Pol III are the homologs of the *E. coli* α homodimer. The *E. coli* ω subunit is homologous to Rpb6 (ABC23) in yeast and is a common subunit of Pol I, II and III (Minakhin et al. 2001). Four additional subunits are common to all three canonical polymerases in yeast, namely Rpb5 (ABC27), Rpb8 (ABC14.5), Rpb10 (ABC10β) and Rpb12 (ABC10α) (Woychik et al. 1990; Carles et al. 1991). Rpa12, Rpb9 and Rpc11 are functional equivalents specific to Pol I, II or III, respectively (Woychik et al. 1991; Nogi et al. 1993; Chedin et al. 1998). Rpb4 and Rpb7 complete the twelve subunit core Pol II enzyme, and in budding yeast, form a dissociable sub-complex (reviewed in (Choder 2004; Sampath and Sadhale 2005)). Rpc25-Rpc17 and Rpa14-Rpa43 are the functional equivalents of Rpb4-Rpb7 in yeast Pol I and III, respectively (Peyroche et al. 2002; Siaut et al. 2003). Mutations in the majority of these subunits result in lethal phenotypes in yeast. However, knockout mutations in a subset of the subunits can yield conditionally viable yeast cells, namely Rpb9, Rpb4, Rpa12 and Rpa14 (Woychik and Young 1989; Woychik et al. 1991; Nogi et al. 1993; Smid et al. 1995). Recent evidence from plants

also indicates that the second largest subunits of Pol I, II or III and several smaller, non-catalytic subunits are essential for viability, consistent with results in yeast (Onodera et al. 2008).

Although the yeast Pol I and Pol III enzymes include a twelve subunit core that includes homologs identical or paralogous to the Pol II core subunits, they differ by the addition of two non-essential subunits specific to Pol I (Rpa34 and Rpa49) and five essential subunits specific to Pol III (Rpc31, Rpc34, Rpc37, Rpc53, Rpc82) (Jasiak et al. 2006; Kuhn et al. 2007; Lorenzen et al. 2007). Rpc31, Rpc34 and Rpc82 interact to form a trimer that is important for transcriptional initiation and binding to the TFIIIB transcription factor complex involved in promoter-specific recognition (Werner et al. 1992; Werner et al. 1993; Thuillier et al. 1995; Brun et al. 1997; Wang and Roeder 1997). Rpc37 and Rpc53 form a heterodimer implicated in termination and facilitated re-initiation of transcription along with Rpc11 (Flores et al. 1999; Hu et al. 2002; Landrieux et al. 2006; Fernandez-Tornero et al. 2007). Of the twelve subunits in Pol III that are identical or homologous to the twelve subunits of Pol II, only the largest and second-largest subunits do not cross-complement between fission yeast and budding yeast. However, none of the five specific Pol III subunits in fission yeast are able to cross-complement their Pol III homologs in budding yeast, suggesting that these polymerase-specific subunits are evolving more rapidly (Proshkina et al. 2006). Moreover, homologs of the yeast Rpa14 Pol I subunit have not been identified outside of Ascomycetes, suggesting that Pol I subunit compositions may not be absolutely conserved throughout eukaryotes (Imazawa et al. 2005).

Rpa49 and Rpa43 form a heterodimer in the yeast Pol I structure and depend on each other for stability in the Pol I complex (Gadal et al. 1997; Beckouet et al. 2008).

These subunits are not essential for viability and do not affect promoter-independent Pol I transcription *in vitro*, but *rpa49* mutants are slow growing (Liljelund et al. 1992; Gadal et al. 1997). Rpa49 interacts with the Pol I transcription initiation factor, Rrn3, in an Rpa43-dependent manner, suggesting that Rpa49 is important for Pol I transcription initiation (Beckouet et al. 2008).

In *Arabidopsis thaliana*, many of the RNA polymerase subunits are encoded by multi-gene families, unlike yeast, in which RNA polymerase subunits are each encoded by a single gene. Some of this subunit diversity is explained by the evolution of Pol IV and Pol V from Pol II. Pol IV and Pol V utilize distinct largest and second-largest subunits in addition to unique Rpb4-, Rpb7- and Rpb5-related subunits (Ream et al. 2009). However, there has also been expansion of the gene families encoding predicted Pol I or Pol III subunits. Two genes related to yeast Rpac40, AtRPAC42 and AtRPAC43, were previously shown to be expressed in *Arabidopsis* suspension culture. AtRPAC42 was shown to co-purify with Pol III, but the function of AtRPAC43 has been unknown (Ulmasov et al. 1995). AtRPAC42 and AtRPAC43 may be functionally redundant or they could be functionally distinct, differentially associating with Pol I and Pol III. In Trypanosomes, different Rpb5 and Rpb6 variants associate with Pol I or Pol II, which presumably reflects their functional diversification (Devaux et al. 2007). In contrast, the Trypanosome Rpb7 homolog functions in both Pol I and Pol II, whereas in yeast, Rpa43 is the Pol I subunit equivalent of Rpb7.

Recently, we determined subunit compositions for *Arabidopsis* Pol II, Pol IV and Pol V (Ream et al. 2009). Pol IV and Pol V share extensive subunit homology with Pol II, but have unique subunit variants that reside in positions near the DNA entry and RNA

exit channels, providing clues as to their functional diversification beyond the catalytic core (Huang et al. 2009; Ream et al. 2009). Currently, equivalent information regarding the subunit compositions of Pol I and Pol III in plants is lacking.

In this paper, we have affinity purified *Arabidopsis* Pol I and Pol III and determined their subunit compositions. In addition to the subunits expected to be shared with Pol II, we show that there are multiple homologs of several yeast Pol III-specific subunits, namely the Rpc11, Rpc17, Rpc31, and Rpc53 subunits. Two subunits corresponding to the yeast Rpc53 subunit are able to incorporate into Pol III, providing the first evidence for potential functional duplication of a Pol III-specific subunit. The two homologs of yeast Rpac40 uniquely associate with Pol I and Pol III. Combined with our previous subunit determinations for Pol II, Pol IV and Pol V, the functional roles for nearly all of the RNA polymerase subunits in the *Arabidopsis* genome can be accounted, of which there are over fifty. These findings provide a platform for the studies into the evolution, expansion and functional diversification of RNA polymerases and their subunits in plants.

Results

Arabidopsis **homologs of yeast Pol I- and Pol III-specific subunits**

Previous studies have identified *Arabidopsis* homologs that are shared or are homologous in Pol I, II and III (Saez-Vasquez and Pikaard 1997; Larkin et al. 1999). Many of these subunits are organized into gene families, suggesting these members could have undergone functional duplication or diversification (Ulmasov et al. 1995; Ulmasov et al. 1996; Larkin et al. 1999; Ream et al. 2009). Less is known about the organization

of Pol I and Pol III-specific subunits in *Arabidopsis*, although one study suggested that the yeast Pol III-specific subunits are conserved in plants (Proshkina et al. 2006).

We used BLASTp searches with query sequences corresponding to yeast Pol I or Pol III-specific subunits Rpa34, Rpa49, Rpc82, Rpc53, Rpc37, Rpc34 and Rpc31 to identify putative *Arabidopsis* homologs. The results of these searches are presented in Table 1, and alignments of these proteins are shown in Supplemental Figures 1-8. Homologs to Rpc17, the five yeast Pol III-specific subunits and the Pol I-specific Rpa49 subunit were identified, but we did not identify any putative homolog for the yeast Rpa34 subunit using this method. In *Arabidopsis*, we identified single homologs of Rpa49, Rpc34, Rpc37 and Rpc82, suggesting that these subunits are encoded by single genes. Surprisingly, we identified two putative homologs of Rpc17 (At5g62950 and At3g28956), Rpc31 (At4g01590 and At4g35680) and Rpc53 (At4g25180 and At5g09380). Using the identified *Arabidopsis* subunits, we performed a second BLASTp search to identify homologs in other plants. In general, mosses, monocots and dicots contain at least one homolog of all of the identified subunits. However, we identified two putative Rpc34 homologs in maize, two Rpc82 homologs in grape, two Rpac40 homologs in rice, and at least two homologs of Rpc53 in rice, maize and grape, but not in moss.

In keeping with a new nomenclature system we have proposed in plants, the core subunits of RNA polymerases I, II and III have been named based on their homology to yeast Pol II. Pol III-specific subunits have been named NRPC13 through NRPC17, respectively, which stands for "Nuclear RNA Polymerase C (III) subunit". The numbers 13-17 were given to the *Arabidopsis* homologs based on their homology to the yeast Pol III subunits and ranked in order of their decreasing size in yeast (Table 2). These criteria

were established because the conventional yeast nomenclature, based on protein size in kilodaltons, does not translate to plants. For example, Rpc37 is predicted to be 32 kDa in yeast but the equivalent subunit, NRPC15, is predicted to be 78 kDa in *Arabidopsis*. Naming subunits based on their homology to yeast subunits provides a point of reference in regards to function and physical placement within the enzyme, whose structure is known (Jasiak et al. 2006; Fernandez-Tornero et al. 2007). Furthermore, this system precludes likely discrepancies among different plant species in which sizes of orthologous subunits may differ. Based on this convention, the *Arabidopsis* homologs of yeast Rpa49 and Rpa34 are designated as NRPA13 and NRPA14, respectively. In cases where there are multiple members of a subunit family, each subunit variant is distinguished by a lowercase letter. For example, the two Rpb6 homologs are referred to as NRPB6a and NRPB6b in the context of Pol II. When referring to these subunits in the context of Pol V, they are NRPE6a and NRPE6b.

BLASTp searches identified the *Arabidopsis* Pol I and Pol III equivalents of the Rpa14-Rpa43 and Rpc17-Rpc25 sub-complexes in yeast, which are equivalent to the Rpb4-Rpb7 sub-complex in Pol II. We identified homologs of Rpa43 (NRPA7), Rpc17 (NRPC4) and Rpc25 (NRPC7). NRPA7 and NRPC7 are encoded by single genes in *Arabidopsis* (Ream et al. 2009), whereas NRPC4 has at least two putative homologs, At5g62950 and At3g28956 (Table 1, Figure S2). Attempts to identify the *Arabidopsis* homolog of the yeast Pol I subunit Rpa14 using BLASTp searches have been unsuccessful in our hands and may be attributed to the accelerated divergence of polymerase-specific subunits among species. As noted previously, Rpa14 has only been

identified in yeast, thus far, such that it is possibly a yeast-specific subunit (Imazawa et al. 2005; Proshkina et al. 2006).

Identification of an *Arabidopsis* Pol III complex by affinity purification and LC-MS/MS analysis

The presence of homologs of yeast Pol I- and Pol III-specific genes in *Arabidopsis* suggests that these polymerases are similar in size and subunit composition in plants and yeast, as suggested by previous purifications of these enzymes (Ulmasov et al. 1995; Saez-Vasquez and Pikaard 1997). To identify the subunit compositions of Pol I and Pol III, we affinity purified the enzymes using epitope-tagged versions of their second-largest subunits, NRPA2 and NRPC2, respectively. The *NRPA2-FLAG* and *NRPC2-FLAG* transgenes were expressed from their native promoters and rescue the lethal phenotypes of the *nrpa2-1* and *nrpc2-1* mutants, respectively (Onodera et al. 2008). Proteins obtained from anti-FLAG immunoprecipitates (IPs) from leaf extracts were subjected to trypsin digestion and LC-MS/MS analysis to yield peptide sequence information. Peptide sequence coverage results for affinity purified Pol I and Pol III samples are presented in Table 2 and Figures S9-S10. A high degree of coverage was obtained for homologs of all seventeen yeast Pol III subunits, indicating that plant and yeast Pol III share a conserved subunit structure. Importantly, no peptides unique to Pol I, II, IV or V were identified, indicating that our affinity purification yielded only Pol III (Table 2).

In the affinity purified Pol III sample, the largest subunit, NRPC1, was identified with 22% unique sequence coverage and 39% coverage was observed for NRPC2, the

second largest subunit. Recall that there are two homologs of yeast Rpac40 in *Arabidopsis*. 57% unique sequence coverage for one of the homologs was obtained, NRPC3 (At1g60620, formerly AtRPAC43), but no At1g60850 (formerly AtRPAC42) peptides were detected (Table 2) (Ulmasov et al. 1995). We obtained 50% coverage for NRPC11 (formerly AtRPAC19), the *E. coli* α subunit homolog that is the dimerization partner for NRPC3 (Ulmasov et al. 1995; Larkin and Guilfoyle 1997).

Subunits shared by Pol I, II and III in yeast and other organisms include Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12. As expected from previous studies, we identified NRPC5/NRPB5/NRPD5 (At3g22320) in the affinity purified Pol III sample with 44% unique coverage (Larkin et al. 1999; Ream et al. 2009). NRPC5 is encoded by the same gene as the NRPB5 and NRPD5 subunits of Pol II and Pol IV, but is not used by Pol V, which has a unique NRPE5 variant (Table 2) (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). We identified unique peptides corresponding to NRPB6a, NRPB6b, NRPB8a and NRPB8b, suggesting that either variant of these subunit families can function in the context of Pol III, as observed in Pol II, Pol IV and Pol V (Table 2)(Ream et al. 2009). We obtained 72% coverage for NRPC10 and 16% coverage for NRPC12. Notably, NRPC10 is encoded by the same gene as NRPB10 of Pol II, NRPD10 of Pol IV and NPRE10 of Pol V. Likewise, NRPC12 is encoded by the same gene as NRPB12, NRPD12 and NRPE12. We did not identify peptides unique to the NRPB10-like gene, At1g61700, or the NRPB12-like gene, At1g53690.

The yeast Pol III subunits, Rpc25 and Rpc17, are the equivalents of the yeast Pol II Rpb7 and Rpb4 subunits, respectively. We identified 48% coverage for the Rpc25 homolog, designated NRPC7. Peptides unique to one of two Rpc17 homologs in

Arabidopsis were identified, designated NRPC4 (At5g62950), representing 41% of the total protein (Table 2). In contrast, we identified only one peptide, representing 11% coverage, matching either of two potential NRPC9 subunits, NRPC9a (At4g07950) and NRPC9b (At1g01210).

Arabidopsis contains a single homolog to the yeast Rpc82 subunit, which we designate as NRPC13 in *Arabidopsis*. Unique peptide coverage was 27% for NRPC13 (Rpc82). There are two potential Rpc53 homologs in *Arabidopsis*, designated as NRPC14a and NRPC14b. Interestingly, we obtained 57% unique coverage for NRPC14a and 32% unique coverage for NRPC14b, suggesting that either alternative subunit can incorporate into Pol III. 38% unique coverage was obtained for the Rpc37 homolog, NRPC15 (At5g49530), and 25% unique coverage was identified for NRPC16 (Rpc34). Of two Rpc31-like subunits, only unique coverage was detected for NRPC17a (At4g01590, 33%) (Table 2, Figure S10). In summary, all seventeen homologs of the core yeast Pol III were identified in affinity purified *Arabidopsis* Pol III, constituting a core complex predicted to be ~675 kDa in size.

Identification of an *Arabidopsis* Pol I complex by affinity purification and LC-MS/MS analysis

In Pol I affinity purified by virtue of an NRPA2-FLAG tag, 49% unique sequence coverage was obtained for NRPA1, the largest subunit of Pol I, and 51% coverage was observed for NRPA2, the second-largest subunit. Of the two yeast Rpac40 homologs in *Arabidopsis*, we obtained 61% unique sequence coverage for one of them (At1g60850, formerly AtRPAC42), which we designated as NRPA3, and only 3% coverage of NRPC3

(At1g60620, formerly AtRPAC43). Combined with the Pol III data, these results indicate that Pol I and Pol III use distinct Rpac40 homologs as subunits, unlike yeast. We obtained 29% unique coverage for NRPA11/NRPC11 (formerly AtRPAC19), the *E. coli* α subunit homolog that is the partner of both NRPA3 and NRPC3 in the context of Pol I and Pol III, respectively (Table 2, Figure S9)(Ulmasov et al. 1995; Larkin and Guilfoyle 1997).

In affinity purified Pol I, we identified three of the five subunits shared by Pol I, II, III, IV and V, namely the homologs of Rpb5, Rpb8 and Rpb10 (Table 2). By contrast, the Rpb6 and Rpb12 shared subunits were not detected. However, NRPA6a and NRPA6b association with Pol I was detected by co-immunoprecipitation in subsequent tests (see below). As expected, we identified NRPA5 as the Rpb5 equivalent in Pol I with 59% unique coverage, in agreement with prior studies (Saez-Vasquez and Pikaard 1997; Ream et al. 2009). We identified unique peptide sequences for both NRPB8 paralogs in Pol I, suggesting that either of the subunits can associate into Pol I, as for Pol II, III, IV and V (Table 2) (Ream et al. 2009). We detected 60% unique coverage for the Rpb9 family homolog most closely related to yeast Rpa12 in our Pol I IP, which we designated as NRPA9 (At3g25940). We observed 55% coverage for NRPA10, encoded by the same gene as the NRPB10, NRPC10, NRPD10 and NRPE10 subunits (At1g11475), but not the NRPB10-like gene, At1g61700, which has not been found in any polymerase to date (Table 2).

Rpa49 is a subunit unique to Pol I and, as expected, we observed 53% unique coverage for a homolog of this subunit, designated as NRPA13 (Table 2). Interestingly, despite our failed attempts to identify the yeast Rpa34 homolog by BLAST, we identified

a subunit similar in size to yeast Rpa34 as a protein of unknown function in our dataset (At5g64680), with 61% unique coverage. At5g64680 has low amino acid sequence conservation compared to yeast Rpa34 (only 10% identity), potentially explaining our difficulties using BLASTp search programs (Figure S12). However, there are several short patches of absolutely conserved residues present in yeast, humans, mice and plants. In addition, At5g64680 was identified in the nucleolar proteome, the sub-nuclear compartment where Pol I transcription occurs (Pendle et al. 2005). Taken together, this evidence suggests that At5g64680 is the functional homolog of Rpa34. Therefore, we designate At5g64680 as NRPA14. Searches with the *Arabidopsis* sequence in other plants revealed that rice, maize, grape and moss all contain at least one homolog to NRPA14 (Table 1).

Confirmation of subunit associations using co-immunoprecipitation tests

To confirm the subunits identified in affinity purified Pol I and Pol III, we performed co-immunoprecipitation tests with epitope-tagged proteins expressed in transgenic plants in combination with antibodies recognizing polymerase-specific subunits.

We raised polyclonal antibodies against NRPA3 and NRPC7, two subunits that are specific to Pol I or Pol III, respectively. To test the specificity of these antibodies, we immunoprecipitated FLAG-tagged largest subunits of Pol IV and Pol V or FLAG-tagged second-largest subunits of Pol I, II, and III, representing the five nuclear DNA-dependent RNA polymerases. All of these transgenes are under control of their respective endogenous promoter sequences and rescue corresponding null mutants (Pontes et al.

2006; Onodera et al. 2008). All five polymerases can be immunoprecipitated and detected on immunoblots with anti-FLAG antibody (Figure 1, top panel). No signals observed in the Col-0 wild-type control IP, with the exception of a cross-reacting band around 55 kDa (Figure 1, top panel). Probing the blots with affinity purified anti-sera to NRPC7 results in a signal of ~25 kDa specifically in the Pol III IP sample, but not in the Pol I, II, IV or V IP samples (Figure 1, middle panel). This suggests that the antibody is specific to Pol III and confirms the mass spectrometry results. Likewise, if we probe these blots with an antibody to NRPA3, we detect a signal corresponding to NRPA3 only in the Pol I IP, but not in Pol II, III, IV or V IP samples (Figure 1, bottom panel). This result agrees with the mass spectrometry data showing that NRPA3 is the predominant Rpac40-like subunit in Pol I.

We next performed anti-FLAG IPs of crude extracts of plants expressing FLAG-tagged subunits of NRPA11/NRPC11, NRPA3, NRPC3, NRPA9, NRPC9b, NRPB9a, NRPB8b, NRPB10 and NRPB6b (Figure 2A). We also performed anti-FLAG IPs of crude extracts of transgenic lines expressing FLAG-tagged subunits of NRPB11, NRPB9b, NRPC7, NRPB6a, NRPC9a or NRPB8a (Figure 2B). Each of these tagged subunits is detected with the anti-FLAG antibody after pull-down (Figure 2A or 2B, top panels). NRPB6a protein expression is low compared to the other FLAG-tagged lines, consistent with previous observations (Ream et al. 2009). Probing the blots with anti-NRPC7 detected the presence of NRPC7 in NRPC11/NRPA11, NRPC3, NRPB8b/NRPC8b, NRPB10/NRPC10, NRPB6b/NRPC6b, NRPC7, NRPB6a/NRPC6a and NRPB8a/NRPC8a FLAG IP samples, confirming that these subunits are all associated with Pol III (Figure 2A-B, second panels). Interestingly, we did not detect

NRPC7 in the NRPC9a or NRPC9b (At4g07950 or At1g01210, respectively) IP samples, despite the LC-MS/MS data that suggested either subunit variant might incorporate into Pol III. A possible explanation for this discrepancy may be that the FLAG epitope interferes with recombinant NRPC9 association with Pol III.

Probing these immunoblots with anti-NRPA3 revealed the presence of NRPA3 in NRPA11/NRPC11, NRPA3, NRPB8b/NRPA8b, NRPB10/NRPA10, NRPB6b/NRPA6b, NRPB6a/NRPA6a and NRPB8a/NRPA8a IP samples, indicating that these proteins all associate with Pol I (Figures 2A-B, third panels). Notably, although we did not detect an Rpb6 homolog by mass spectrometry, we clearly see association of both NRPB6/NRPA6 variants with Pol I by co-IP. As a control, we probed the same IPs with an NRPD2/NRPE2 antibody, which only is enriched in IPs of Pol IV or Pol V subunits (Onodera et al. 2005; Ream et al. 2009). NRPD2 is detected in NRPB6a, NRPB6b, NRPB8a, NRPB8b, NRPB9a, NRPB9b, NRPB10 and NRPB11 IP samples, confirming our previous study demonstrating that these proteins are subunits of Pol IV and Pol V (Figure 2A-B, bottom panels) (Ream et al. 2009). NRPD2 is not detected in NRPA11/NRPC11, NRPA3, NRPC3, NRPC7, NRPA9, NRPC9a or NRPC9b IP samples, consistent with our previous mass spectrometry data of Pol IV and Pol V affinity purified samples (Figure 2A-B, bottom panels) (Ream et al. 2009). For all three antibodies, no specific signals are detected in the Col-0 wild-type control IP at the size ranges corresponding to NRPC7, NRPA3 or NRPD2. Moreover, the Pol I specific subunits NRPA3 and NRPA9 do not co-IP with NRPD2 or NRPC7. Likewise, the Pol III-specific subunit NRPC3 does not co-IP with NRPA3 or NRPD2. All of these results agree with the LC-MS/MS data.

Based on our purification studies, we have observed that there are specific NRPC9-like, NRPB7-like, NRPB10-like and NRPB12-like subunits in the genome that are not found in Pol I, II, III, IV or V. One possibility is that the genes encoding these subunits are pseudogenes. Alternatively, these subunits might be expressed but are incorporated into a minor fraction of the polymerase pool, such that they are not detected by our purification and mass spectrometry approaches. This is consistent with the fact that we were able to clone the cDNAs of each of these subunits. We performed RT-PCR analysis of the *NRPB10*-like (At1g61700) and *NRPB12*-like genes (At1g53690) to examine their expression in different plant tissues. Both genes are expressed in a tissue-specific manner, suggesting that the subunits they encode might be incorporated into tissue-specific forms of one or more of the five polymerases (Figure 3A).

To test if the potentially alternative subunits can associate into any of the five polymerases, we over-expressed them using the strong constitutive 35S promoter, from cauliflower mosaic virus, and performed IPs to assay if they interact with the catalytic subunits of Pol I, II, III, IV or V. NRPC9b, NRPB7-like and NRPB12-like are all detectable with the anti-FLAG antibody by immunoblot after IP (Figure 3B, top panel). Figure 3 shows that NRPE5, NRPA9 or NRPE9a, associate with at least one of the RNA polymerases (Figure 3B, middle and bottom panels). By contrast, NRPC9b (At1g01210), NRPB7-like (At4g14520), and NRPB12-like (At1g53690) do not associate with the second-largest catalytic subunits of Pol I, II, III, IV or V (Figure 3B, middle and bottom panels). Even at high expression levels, these subunits appear to be excluded from all of the RNA polymerases, making their functions unknown.

Genetic analysis of RNA polymerase I and III subunits

Previous studies of mutations disrupting the genes encoding the second-largest subunits of Pol I, II and III suggested that these mutations are lethal (Onodera et al. 2008). Furthermore, the T-DNA mutations prevent the development of the female gametophyte, whereas male gametophytes carrying the T-DNA mutation are still viable and are able to complete fertilization, although at a reduced frequency (Onodera et al. 2008). Similar results were obtained for the gene encoding the NRPA5/NRPB5/NRPC5/NRPD5 subunit common to Pol I, II, III and IV, the NRPB11/NRPD11/NRPE11 subunit common to Pol II, IV and V, and the NRPB12/NRPC12/NRPD12/NRPE12 subunit shared by at least four out of five polymerases (Onodera et al. 2008; Ream et al. 2009). To test if mutations in genes encoding subunits that are specific to either Pol I and/or Pol III are viable, we isolated T-DNA mutations in *NRPA3* (Pol I-specific), *NRPC3* (Pol III-specific) and *NRPA11/NRPC11* (shared by Pol I/III). Segregation analysis of progeny from a single heterozygous parent resulted in zero offspring that were homozygous mutant, suggesting that these are essential genes whose mutations are lethal (Table 3).

Discussion

In conclusion, we have identified all seventeen subunits demonstrated to associate with the core RNA polymerase III in yeast. Similarly, a combination of affinity purification coupled to mass spectrometry and co-immunoprecipitation analysis identified a majority of the Pol I subunits found in yeast. Notable exceptions are the yeast Rpa14, Rpa43 and Rpb12 homologs. The purification and determination of the subunit

compositions of Pol I and Pol III, combined with our prior data for Pol II, Pol IV and Pol V, provides the first comprehensive analysis of the subunit compositions of all known DNA-directed RNA polymerases in *Arabidopsis* (Table 4). Our results reveal that many of the subunits shared by Pol I, II and III in yeast are shared by Pol I, II, III, IV and V in *Arabidopsis*, with the notable exception of the NRPB5 subunit equivalent in Pol V, NRPE5, which is unique to Pol V and is required for Pol V-mediated silencing (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). Of the two subunits uniquely shared by Pol I and III in yeast, Rpac40 and Rpac19, only one is shared by Pol I and III in *Arabidopsis*, namely NRPC11/NRPA11 (Rpac19). By contrast, two non-redundant Rpac40 homologs are present in *Arabidopsis*, NRPA3 (formerly AtRPAC42) and NRPC3 (formerly AtRPAC43), and these are differentially incorporated into Pol I or Pol III, respectively. This is the first example of a subunit shared by Pol I and Pol III that has been duplicated and functionally diverged to produce polymerase-specific paralogs. The Rpac40-like family of *Arabidopsis* can therefore be added to the growing list of subunit families that have undergone functional divergence in eukaryotes, including Rpb3 (plants), Rpb4 (plants), Rpb5 (plants, protozoans), Rpb6 (protozoans) and Rpb7 (plants, protozoans) (Devaux et al. 2007; He et al. 2009; Lahmy et al. 2009; Penate et al. 2009; Ream et al. 2009).

The extensive subunit diversity among organisms has created a nomenclature challenge for referring to functionally equivalent subunits across species. Using a nomenclature system based on protein size has disadvantages because equivalent subunits across species can vary greatly in size and do not often correlate from largest to smallest. For example, in yeast, Rpc82 is 74 kDa in size, but the NRPC13 equivalent in

Arabidopsis is only 59 kDa. Likewise, Rpc53 in yeast is predicted to be 47 kDa in size, but in *Arabidopsis* the NRPC14b equivalent is predicted to be 30 kDa in size. Therefore, naming subunits strictly by size would not easily indicate functional equivalency.

In *Arabidopsis*, we have adopted a nomenclature convention that names subunits according to their equivalents in yeast, for which structural information is available. Subunits corresponding to the twelve fundamental subunits of Pol I, II and III are number 1-12, from largest to smallest. Pol I- or Pol III-specific subunits that have no Pol II equivalent are numbered 13 or greater. Our naming of RNA polymerase subunits NRPC13-17 in Pol III or NRPA13-14 in Pol I can be applied to other plant model species in future studies in order to establish and maintain a unified nomenclature system.

The partitioning of different Rpac40 homologs into Pol I and Pol III is intriguing. These subunits are only 71% identical, in contrast to other shared subunits that have duplicated but retained very high identity with each other (over 90%), such as the variants in the NRPB6 and NRPB8 families. Our genetic evidence arguing that each subunit is lethal when mutated supports the conclusion that these subunits are non-redundant and functionally and/or structurally diverged. Why plants evolved Pol I- and Pol III-specific Rpac40 subunits remains to be answered.

A previous study reported that NRPA3 (AtNRPAC42) is a subunit in highly purified fractions of Pol III (Ulmasov et al. 1995). However, we observed that this subunit is preferentially associated with Pol I, whereas NRPC3 (AtNRPAC43) preferentially associates with Pol III. The method used by Ulmasov et al. to suggest that NRPA3 associates with Pol III was based on the mobility of a polypeptide band in a purified Pol III sample relative to recombinant NRPA3 or NRPC3 in an SDS

polyacrylamide gel. It is possible that post-translation modifications not reflected in the recombinantly expressed NRPA3 could have accounted for the size discrepancy. Another explanation that may account for our differences is that Ulmasov et al. purified Pol III from callus suspension culture cells, whereas we have purified Pol III from leaf tissue.

In addition to using alternative NRPB6 and NRPB8 subunits, Pol III utilizes two forms of NRPC14 (Rpc53). These subunits are only 33% identical, and therefore may have different roles in the Pol III complex that remain to be determined.

It is interesting that homologs of yeast Rpa14 and Rpa43 are missing from our datasets. The Rpa43 subunit interacts with the Pol I initiation factor Rrn3 in yeast, which is required for the initiation of Pol I transcription at rRNA gene promoters (Peyroche et al. 2000). A homolog of yeast Rpa43 has been identified in *Arabidopsis* (At1g75670) (Ream et al. 2009), but as stated previously we have not found a homolog to yeast Rpa14, thus far. It is possible that these subunits dissociate during the affinity purification process. Alternatively, our protocol may have selectively purified inactive Pol I complexes that are missing these subunits. If true, this is not the first report of a purified Pol I complex with missing subunits. Purification of Trypanosome Pol I complexes by two labs identified only five or nine of the predicted subunits, in contrast to our twelve (Walgraffe et al. 2005; Nguyen et al. 2006). Notably, these purifications also lacked homologs to Rpb6, Rpa14 and Rpa43, although in one case Rpb6 was later shown to interact by co-IP, similar to our study (Figure 1)(Nguyen et al. 2006).

Plants are unique in having two non-essential polymerases, Pol IV and Pol V, in addition to Pol I, II and III, thereby explaining some of the polymerase subunit diversity

in the genome. However, it is now apparent that polymerase subunit diversity is also due to duplication events resulting in redundant or tissue-specific subunit variants. Duplication of subunits may be widespread in plants, because rice contains at least two homologs to Rpac40, which are 78% identical to each other (Table 1). In addition, there are two putative Rpc82-like subunits (58% identity) and two Rpc53-like subunits (55% identity) in grape (Table 1). Future studies are necessary to tease apart any functional redundancy and diversification of these subunits in plants. Comprehensive searches that identify all the putative plant homologs to RNA polymerase subunits will place an evolutionary context on the emergence of RNA polymerase subunit families in plants.

Materials and Methods

Plant materials. *A. thaliana nrpa3-1* is T-DNA line SALK_088247 and *nrpc3-1* is T-DNA line SALK_132788 (Alonso et al. 2003). *nrpa11-1* is T-DNA line WiscDsLox_419G02 (Woody et al. 2007). All primers are listed in Table S1. Plants were grown in a greenhouse or in growth chambers under long day photoperiods (16 hr. light, 8 hr. dark).

Genotyping. Genotyping was performed as described previously (Ream et al. 2009). Cycling conditions for genotyping *nrpa3-1, nrpc3-1* and *nrpa11-1* were: 94˚C 2 min. 30 sec., 36 cycles of 94˚C 30 sec., 55˚C 30 sec. and 72˚C 1 min. 15 sec. followed by a final extension of 72˚C for 7 min.

Affinity purification of Pol I and Pol III. 150-250 g of fresh or frozen leaf tissue expressing FLAG-tagged *NRPA2* or *NRPC2* was ground in extraction buffer (300 mM NaCl, 20 mM Tris pH 7.5, 5 mM $MgCl_2$, 5 mM DTT, 1 mM PMSF and 1:200 plant protease inhibitor cocktail (Sigma)) at 4˚C, filtered through two layers of Miracloth (Calbiochem) and centrifuged twice at 7000-10,000g, 25 min, 4˚C. Supernatants were incubated with anti-FLAG-M2 resin for 3 hours in a 15 ml tube using 50 ul of resin per 14 mls of extract. Pooled resin was washed 7 times in 14 mls of extraction buffer containing 0.4% NP-40 (Sigma). Wash buffer was removed from the resin before adding 1 volume of Ag/Ab Elution Buffer (Pierce) at 4˚C. The sample was mixed thoroughly and incubated for 3 min. on ice. The resin was pelleted, the eluted complex was aliquoted into 500 ul batches and was concentrated with YM-10 Centricon columns (Millipore) at 4˚C and desalted twice using Pierce 500 ul desalting columns. The final sample, eluted in 25 mM ammonium bicarbonate, was subjected to LC-MS/MS.

Mass spectrometry. Samples adjusted to 50% (v/v) 2,2,2-Trifluoroethanol (TFE) (Sigma), were sonicated 1 min. at 0˚C then incubated 2 hr. at 60ºC with shaking at 300 rpm. Proteins were reduced with 2 mM DTT, 37ºC for 1 hr., then diluted 5-fold with 50 mM ammonium bicarbonate. 1 mM $CaCl_2$ and sequencing-grade modified porcine trypsin (Promega) was added, at a 1:50 trypsin-to-protein mass ratio. After 3 hrs. at 37˚C, samples were concentrated to ~30 µl and subjected to reversed-phase liquid chromatography (RPLC) coupled to an electrospray ionization source and LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific). Tandem mass spectra were searched

against *A. thaliana* proteins using SEQUEST and filtering criteria, which provided a False Discovery Rate (FDR) <5%. See supplemental material for details.

Cloning, vectors and transgenic lines. *NRPD1-FLAG*, *NRPE1-FLAG*, *NRPA2-FLAG*, *NRPB2-FLAG* and *NRPC2-FLAG* transgenes were previously described (Pontes et al. 2006; Onodera et al. 2008). *NRPE3a, NRPE3b, NRPE5*, *NRPB6a*, *NRPB6b*, *NRPB7*, *NRPB8a*, *NRPB8a*, *NRPB9a, NRPB10*, *NRPB11* transgenic lines have been described (Ream et al. 2009). *NRPC7 (At1g06790), NRPA9 (Ag3g25940), NRPB9b (At4g16265), NRPC9a (At4g07950), NRPC9b (At1g01210), NRPB7-like (At4g14520)* and *NRPB12-like (At1g53690)* cDNAs were amplified by RT-PCR from poly-T primed cDNA or available ABRC clones and were cloned into pENTR-D-TOPO or pENTR-TEV-TOPO. cDNAs were recombined into pEarleyGate 201 (HA tag) or 202 (FLAG tag) (Earley et al. 2006). Genomic *NRPB12*, *NRPA11*, *NRPA3* and *NRPC3* clones were amplified with *Pfu* Ultra (Stratagene) and similarly cloned into pEarleyGate 302 (FLAG tag).

RT-PCR. High molecular weight RNA was isolated from different tissues using a miRVANA kit (Ambion). First-strand cDNAs were generated using poly-T primers and PCR-amplified using gene-specific primers to each *NRPB10* and *NRPB12* family member.

Antibodies. Anti-NRPE2/NRPD2 has been described (Larkin et al. 1999; Onodera et al. 2005). Anti-FLAG antibodies were from Sigma. Anti-NRPB1-CTD (8WG16) was purchased from Abcam.

Production of antibodies. Full-length *NRPC7* and *NRPA3* cDNAs in pENTR (described above) were recombined into pDEST17 using LR Clonase (Invitrogen) and transformed into BL21ai cells. Protein expression was induced with 0.2% arabinose for 3 hours at 37˚C in 125 ml 1x LB cultures. Cells were then spun 6000 rpm for 10 min. Inclusion bodies containing the target protein were purified with B-Per II lysis reagent (Pierce) according to the manufacturer's instructions. Purified inclusion bodies were dissolved in 4 mls of 2x SDS buffer and stored at -20˚C. Expression was confirmed by western blot (data not shown). Protein concentration was estimated by running a dilution series of BSA alongside 5 ul of total insoluble protein. 3 mgs of each protein was gel purified, identified by coomassie staining, sliced out and submitted to Sigma Genosys for antibody production. Affinity purified anti-sera (1:100) were used in the experiments.

Affinity purification of antibodies. ~2 mgs of recombinant protein were separated by SDS-PAGE and transferred to PVDF membrane using standard protocols. After a brief wash in TBST, the membrane was stained with Ponceau S for 2 min. and destained with water. The region corresponding to the target protein was excised and the membrane containing the protein was completely destained in several exchanges of TBST over a 10 min. period. The membrane was then blocked in TBST+ 5% milk for 1 hr. The blocking solution was discarded and a fresh solution containing 2 mls of crude anti-sera and 8 mls of TBST+ 5% milk was incubated with the membrane on an orbital shaker at 4˚C overnight. Membranes were washed three times in TBST for 5 min. each. The membranes were cut into small strips 1 cm x 0.5 cm and transferred to 2 ml

microcentrifuge tubes. Tubes were spun briefly in a microcentrifuge at 650 rpm for 1 min to remove residual TBST. Antibodies were eluted in 1 ml of 100 mM glycine, pH 2.5 (enough to cover the membrane strips) and the tubes were mixed thoroughly. The solution containing the eluted antibodies was removed and added to a new tube containing 100 ul of 1 M Tris pH 8.0. 1 volume of glycerol was added to a final concentration of 50%. Antibodies were stored at -20˚C until needed.

Immunoprecipitation and immunoblotting. 2-4 g of leaves were ground in extraction buffer (Baumberger and Baulcombe 2005), filtered through Miracloth and centrifuged at 10,000g for 15 min. Supernatants were incubated 3 hrs. at 4˚C with 30-50 ul of anti-FLAG-M2 resin (Sigma). Beads were washed 3x in extraction buffer + 0.5% NP-40 (Sigma), eluted with two bed volumes of 2X SDS sample buffer, and 5-20 ul were subjected to SDS-PAGE and transferred to Immobilon PVDF membranes (Millipore). Blots were incubated with antibodies in TBST + 5% (w/v) non-fat dried milk. Antibody dilutions were: 1:100 (NRPC7, NRPA3), 1:250 (NRPD2/NRPE2), 1:500 (anti-Pol I, II and/or III) and 1:2000-1:20,000 (FLAG-HRP). The secondary antibody was anti-rabbit-HRP, diluted 1:5,000-1:20,000, or anti-mouse-HRP, diluted 1:5000 (GE Healthcare, Sigma). Blots were washed 4X 4 min in TBST and visualized by chemiluminescence (GE Healthcare). Blots were stripped for 35 min in 25 mM glycine pH 2.0, 1% SDS, re-equilibrated in TBST and probed with additional antibodies.

Alignments. Sequences were aligned using ClustalW and highlighted using BOXSHADE. Genbank accession nos. are listed in Table S2.

Acknowledgments

T.S.R. and C.S.P. designed the study and wrote the paper. T.S.R. performed all experiments except LC-MS/MS analyses by C.D.N., A.D.N. and L.P.-T. at Pacific Northwest National Laboratory (PNNL). J.R.H. cloned the *NRPA11*, *NRPA3* and *NRPC3* genomic fragments. We thank the Washington University greenhouse staff for plant care and Pikaard lab colleagues for discussions. T.S.R. and C.S.P. also thank Biology 4024 students who helped with cloning cDNAs: Susana San Roman (*At3g25940*), Silvano Ciani and Colin Clune (*At4g14520*), Caitlin Ramsey and Colin Orr (*At1g53690*), Wan Shi and Soon Goo Lee (*At1g06790*), Lily Momper and Charu Agrawal (*At1g01210*). Pikaard lab research is supported by National Institutes of Health grant GM077590. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the NIH. Portions of this research were supported by the NIH National Center for Research Resources (RR18522), and the W.R. Wiley Environmental Molecular Science Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at PNNL. PNNL is operated by Battelle Memorial Institute for the U.S. Department of Energy under contract DE-AC05-76RL01830.

REFERENCES

References cited

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. 2003. Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301**(5633): 653-657.

Baumberger, N. and Baulcombe, D.C. 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc Natl Acad Sci U S A* **102**(33): 11928-11933.

Beckouet, F., Labarre-Mariotte, S., Albert, B., Imazawa, Y., Werner, M., Gadal, O., Nogi, Y., and Thuriaux, P. 2008. Two RNA polymerase I subunits control the binding and release of Rrn3 during transcription. *Mol Cell Biol* **28**(5): 1596-1605.

Brun, I., Sentenac, A., and Werner, M. 1997. Dual role of the C34 subunit of RNA polymerase III in transcription initiation. *EMBO J* **16**(18): 5730-5741.

Carles, C., Treich, I., Bouet, F., Riva, M., and Sentenac, A. 1991. Two additional common subunits, ABC10 alpha and ABC10 beta, are shared by yeast RNA polymerases. *J Biol Chem* **266**(35): 24092-24096.

Chedin, S., Riva, M., Schultz, P., Sentenac, A., and Carles, C. 1998. The RNA cleavage activity of RNA polymerase III is mediated by an essential TFIIS-like subunit and is important for transcription termination. *Genes Dev* **12**(24): 3857-3871.

Choder, M. 2004. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem Sci* **29**(12): 674-681.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**(5523): 1863-1876.

Devaux, S., Kelly, S., Lecordier, L., Wickstead, B., Perez-Morga, D., Pays, E., Vanhamme, L., and Gull, K. 2007. Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. *Mol Biol Cell* **18**(4): 1293-1301.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* **45**(4): 616-629.

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., Sentenac, A., Muller, C.W., and Schoehn, G. 2007. Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol Cell* **25**(6): 813-823.

Flores, A., Briand, J.F., Gadal, O., Andrau, J.C., Rubbi, L., Van Mullem, V., Boschiero, C., Goussot, M., Marck, C., Carles, C., Thuriaux, P., Sentenac, A., and Werner, M. 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proc Natl Acad Sci U S A* **96**(14): 7815-7820.

Gadal, O., Mariotte-Labarre, S., Chedin, S., Quemeneur, E., Carles, C., Sentenac, A., and Thuriaux, P. 1997. A34.5, a nonessential component of yeast RNA polymerase I, cooperates with subunit A14 and DNA topoisomerase I to produce a functional rRNA synthesis machine. *Mol Cell Biol* **17**(4): 1787-1795.
Grummt, I. 2003. Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* **17**(14): 1691-1702.
He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev* **23**(3): 318-330.
Hu, P., Wu, S., Sun, Y., Yuan, C.C., Kobayashi, R., Myers, M.P., and Hernandez, N. 2002. Characterization of human RNA polymerase III identifies orthologues for Saccharomyces cerevisiae RNA polymerase III subunits. *Mol Cell Biol* **22**(22): 8044-8055.
Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16**(1): 91-93.
Huang, Y. and Maraia, R.J. 2001. Comparison of the RNA polymerase III transcription machinery in Schizosaccharomyces pombe, Saccharomyces cerevisiae and human. *Nucleic Acids Res* **29**(13): 2675-2690.
Imazawa, Y., Hisatake, K., Mitsuzawa, H., Matsumoto, M., Tsukui, T., Nakagawa, K., Nakadai, T., Shimada, M., Ishihama, A., and Nogi, Y. 2005. The fission yeast protein Ker1p is an ortholog of RNA polymerase I subunit A14 in Saccharomyces cerevisiae and is required for stable association of Rrn3p and RPA21 in RNA polymerase I. *J Biol Chem* **280**(12): 11467-11474.
Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. 2006. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* **23**(1): 71-81.
Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell* **131**(7): 1260-1272.
Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. 2009. PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci U S A* **106**(3): 941-946.
Landrieux, E., Alic, N., Ducrot, C., Acker, J., Riva, M., and Carles, C. 2006. A subcomplex of RNA polymerase III subunits involved in transcription termination and reinitiation. *EMBO J* **25**(1): 118-128.
Larkin, R.M. and Guilfoyle, T.J. 1997. Reconstitution of yeast and Arabidopsis RNA polymerase alpha-like subunit heterodimers. *J Biol Chem* **272**(19): 12824-12830.
Larkin, R.M., Hagen, G., and Guilfoyle, T.J. 1999. Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. *Gene* **231**(1-2): 41-47.
Liljelund, P., Mariotte, S., Buhler, J.M., and Sentenac, A. 1992. Characterization and mutagenesis of the gene encoding the A49 subunit of RNA polymerase A in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **89**(19): 9302-9305.

Lorenzen, K., Vannini, A., Cramer, P., and Heck, A.J. 2007. Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture. *Structure* **15**(10): 1237-1245.

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. 2001. Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**(3): 892-897.

Nguyen, T.N., Schimanski, B., Zahn, A., Klumpp, B., and Gunzl, A. 2006. Purification of an eight subunit RNA polymerase I complex in Trypanosoma brucei. *Mol Biochem Parasitol* **149**(1): 27-37.

Nogi, Y., Yano, R., Dodd, J., Carles, C., and Nomura, M. 1993. Gene RRN4 in Saccharomyces cerevisiae encodes the A12.2 subunit of RNA polymerase I and is essential only at high temperatures. *Mol Cell Biol* **13**(1): 114-122.

Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**(5): 613-622.

Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. 2008. Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* **180**(1): 207-218.

Penate, X., Lopez-Farfan, D., Landeira, D., Wentland, A., Vidal, I., and Navarro, M. 2009. RNA pol II subunit RPB7 is required for RNA pol I-mediated transcription in Trypanosoma brucei. *EMBO Rep* **10**(3): 252-257.

Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W., and Shaw, P.J. 2005. Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. *Mol Biol Cell* **16**(1): 260-269.

Peyroche, G., Levillain, E., Siaut, M., Callebaut, I., Schultz, P., Sentenac, A., Riva, M., and Carles, C. 2002. The A14-A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4-Rpb7 pol II subunits. *Proc Natl Acad Sci U S A* **99**(23): 14670-14675.

Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. 2000. The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* **19**(20): 5473-5482.

Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. 2006. The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* **126**(1): 79-92.

Proshkina, G.M., Shematorova, E.K., Proshkin, S.A., Zaros, C., Thuriaux, P., and Shpakovski, G.V. 2006. Ancient origin, functional conservation and fast evolution of DNA-dependent RNA polymerase III. *Nucleic Acids Res* **34**(13): 3615-3624.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**(2): 192-203.

Saez-Vasquez, J. and Pikaard, C.S. 1997. Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc Natl Acad Sci U S A* **94**(22): 11869-11874.

Sampath, V. and Sadhale, P. 2005. Rpb4 and Rpb7: a sub-complex integral to multi-subunit RNA polymerases performs a multitude of functions. *IUBMB Life* **57**(2): 93-102.

Schramm, L. and Hernandez, N. 2002. Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**(20): 2593-2620.

Siaut, M., Zaros, C., Levivier, E., Ferri, M.L., Court, M., Werner, M., Callebaut, I., Thuriaux, P., Sentenac, A., and Conesa, C. 2003. An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. *Mol Cell Biol* **23**(1): 195-205.

Smid, A., Riva, M., Bouet, F., Sentenac, A., and Carles, C. 1995. The association of three subunits with yeast RNA polymerase is stabilized by A14. *J Biol Chem* **270**(22): 13534-13540.

Thuillier, V., Stettler, S., Sentenac, A., Thuriaux, P., and Werner, M. 1995. A mutation in the C31 subunit of Saccharomyces cerevisiae RNA polymerase III affects transcription initiation. *EMBO J* **14**(2): 351-359.

Ulmasov, T., Larkin, R.M., and Guilfoyle, T.J. 1995. Arabidopsis expresses two genes that encode polypeptides similar to the yeast RNA polymerase I and III AC40 subunit. *Gene* **167**(1-2): 203-207.

-. 1996. Association between 36- and 13.6-kDa alpha-like subunits of Arabidopsis thaliana RNA polymerase II. *J Biol Chem* **271**(9): 5085-5094.

Walgraffe, D., Devaux, S., Lecordier, L., Dierick, J.F., Dieu, M., Van den Abbeele, J., Pays, E., and Vanhamme, L. 2005. Characterization of subunits of the RNA polymerase I complex in Trypanosoma brucei. *Mol Biochem Parasitol* **139**(2): 249-260.

Wang, Z. and Roeder, R.G. 1997. Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev* **11**(10): 1315-1326.

Werner, M., Chaussivert, N., Willis, I.M., and Sentenac, A. 1993. Interaction between a complex of RNA polymerase III subunits and the 70-kDa component of transcription factor IIIB. *J Biol Chem* **268**(28): 20721-20724.

Werner, M., Hermann-Le Denmat, S., Treich, I., Sentenac, A., and Thuriaux, P. 1992. Effect of mutations in a zinc-binding domain of yeast RNA polymerase C (III) on enzyme function and subunit association. *Mol Cell Biol* **12**(3): 1087-1095.

Woody, S.T., Austin-Phillips, S., Amasino, R.M., and Krysan, P.J. 2007. The WiscDsLox T-DNA collection: an arabidopsis community resource generated by using an improved high-throughput T-DNA sequencing pipeline. *J Plant Res* **120**(1): 157-165.

Woychik, N.A. and Hampsey, M. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**(4): 453-463.

Woychik, N.A., Lane, W.S., and Young, R.A. 1991. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* **266**(28): 19053-19055.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* **4**(3): 313-323.
Woychik, N.A. and Young, R.A. 1989. RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol Cell Biol* **9**(7): 2854-2859.

Table 1. Number of Pol I and Pol III-specific subunits in plants and yeast.

Table 2. Peptide coverage corresponding to RNA polymerase subunits identified in affinity purified Pol I and Pol III complexes. Relationships between *Arabidopsis* Pol I and Pol III subunits from *E. coli*, Archaeal and yeast RNA polymerase subunits are indicated. Numbers indicate percent protein coverage represented by peptides unique to that protein. "*" indicates that peptides match either of two closely-related subunit variants. Unshaded numbers represent alternative subunits detected at trace levels relative to the predominant subunit.

Figure 1. Immunoblots testing the specificity of antibodies directed against Pol I or Pol III specific subunits, NRPA3 and NRPC7, respectively. Pol I, II, III, IV or V were immunoprecipitated by virtue of their FLAG-tagged catalytic subunits. Duplicate immunoblots were probed with anti-FLAG-HRP, anti-NRPC7 or anti-NRPA3.

Figure 2. Verification of Pol I and Pol III subunit associations using co-immunoprecipitation tests. A. Col-0 wild-type, NRPA11, NRPA3, NRPC3, NRPA9, NRPC9b, NRPB9a, NRPB8b, NRPB10 and NRPB6b were immunoprecipitated by virtue of their FLAG epitopes and detected with anti-FLAG-HRP. Duplicate immunoblots were probed with anti-NRPC7, anti-NRPA3 or anti-NRPD2 as indicated. "*" represents a non-specific band detected by the FLAG antibody after immunoprecipitation. B. Col-0 wild-type, NRPB9b, NRPB11, NRPC7, NRPB6a, NRPC9a and NRPB8a were

immunoprecipitated by virtue of their FLAG epitopes and immunoblotting was performed as in Figure 2A.

Figure 3. Alternative RNA polymerase subunits are expressed but do not associate with Pol I, II, III, IV or V. (A) RT-PCR analysis of transcripts encoding both variants of *NRPB10* and *NRPB12* family subunits using gene-specific primers in roots, callus, seedlings, flowers and leaves. Actin and no RT lanes serve as controls. (B) Col-0 wild-type, NRPE5, NRPB12-like (At1g53690), NRPB7-like (At4g14520), NRPA9, NRPC9b (At1g01210) and NRPB9a were immunoprecipitated with anti-FLAG resin and tested for association with Pol I, II, III, IV or V by immunoblotting. Blots were probed with anti-FLAG-HRP, anti-NRPD2 or anti-Pol I-II-III second-largest subunit antibodies as indicated.

Table 3. Genotyping analysis of F2 progeny segregating T-DNAs in *NRPA11*, *NRPA3* or *NRPC3* genes. Offspring from single heterozygote parents were genotyped for wild-type and T-DNA alleles.

Table 4. Partitioning of DNA-dependent RNA polymerase subunits in the five *Arabidopsis* RNA polymerases compared to the equivalent subunits of yeast, archaea and bacteria. Identical color shading within each row represents subunits that are encoded by the same gene in a given species. "?" indicates that this subunit has not been shown to associate with the given RNA polymerase either by mass spectrometry or co-IP

in *Arabidopsis*. "*" indicates that a homolog to this subunit has not been identified in *Arabidopsis* using bioinformatics searches.

Table 1. Number of Pol I and Pol III specific subunits in plants and yeast

	Rpa34/NRPA14	Rpa49/NRPA13	Rpc17/NRPC4	Rpc31/NRPC17	Rpc34/NRPC16	Rpc37/NRPC15	Rpc53/NRPC14	Rpc82/NRPC13	Rpac40-like
yeast	1	1	1	1	1	1	1	1	1
Arabidopsis thaliana	1	1	2	2	1	1	2	1	2
Vitis vinifera	1	1	1	1	1	1	2	2	1
Zea mays	1	1	1	1	2	not identified	2	not identified	1
Oryza sativa (japonica)	2	1	1	1	1	1	3	1	2
Phsycomitrella patens	1	1	1	1	1	1	1	1*	1

* partial sequence

Function	Bacteria	Archaea	Sc Pol I, II, or III	Subunit/Synonyms	At Gene ID	Pol I		Pol III	
						Total	Unique	Total	Unique
Catalytic	B'	RPOA'		NRPD1	At1g63020	0	0	0	0
		RPOA"		NRPE1	At2g40030	0	0	0	0
			Rpa1	NRPA1	At3g57660	49	49	0	0
			Rpb1	NRPB1	At4g35800	0	0	0	0
			Rpc1	NRPC1	At5g60040	0.6	0	22	22
	B	RPOB'		NRPD2/NRPE2	At3g23780	0	0	0	0
		RPOB"		NRPD2b	At3g18090	0	0	0	0
			Rpa2	NRPA2	At1g29940	51	51	0	0
			Rpb2	NRPB2	At4g21710	2	2	0	0
			Rpc2	NRPC2	At5g45140	0	0	39	39
Assembly	α	RPOD	Rpb3	NRPB3/NRPD3/NRPE3a	At2g15430	4	4	0	0
				NRPE3b	At2g15400	4	4	0	0
			Rpac40	NRPA3	At1g60850	61	61	0	0
				NRPC3	At1g60620	7	3	57	57
		RPOL	Rpb11	NRPB11/NRPD11/NRPE11	At3g52090	0	0	0	0
			Rpac19	NRPA11/NRPC11	At2g29540	29	29	50	50
		RPON	Rpb10	NRPA10/NRPB10/NRPC10/NRPD10/NRPE10	At1g11475	70	55	87	72
				NRPB10-like	At1g61700	15	0	15	0
		RPOP	Rpb12	NRPB12/NRPC12/NRPD12/NRPE12	At5g41010	0	0	16	16
				NRPB12-like	At1g53690	0	0	0	0
Auxillary		RPOF	Rpa14	NRPA4	?	0	0	0	0
			Rpc17	NRPC4	At5g62950	0	0	55	41
				NRPC4-like	At3g28956	0	0	14	0
			Rpb4	NRPB4	At5g09920	0	0	0	0
				NRPE4/NRPD4	At4g15950	0	0	0	0
		RPOH	Rpb5	NRPA5/NRPB5/NRPC5/NRPD5	At3g22320	59	59	44	44
				NRPB5-like	At5g57980	0	0	0	0
				NRPE5	At3g57080	0	0	0	0
				NRPE5-like	At2g41340	0	0	0	0
				NRPE5-like	At3g54490	0	0	0	0
	ω	RPOK	Rpb6	NRPB6a/NRPE6a	At5g51940	0	0	36	16
				NRPB6b/NRPE6b	At2g04630	0	0	35	15
		RPOE		NRPE7	At4g14660	0	0	0	0
				NRPD7	At3g22900	0	0	0	0
				NRPB7-like	At4g14520	0	0	0	0
			Rpb7	NRPB7	At5g59180	0	0	0	0
			Rpc25	NRPC7	At1g06790	0	0	48	48
			Rpa43	NRPA7	At1g75670	0	0	0	0
		RPOG	Rpb8	NRPB8a/NRPE8a	At1g54250	65	56	37	28
				NRPB8b/NRPE8b	At3g59600	62	37	33	24
		TFS/RPOX	Rpb9	NRPB8a/NRPD9a/NRPE9a	At3g16980	0	0	0	0
				NRPB9b/NRPD9b/NRPE9b	At4g16265	0	0	0	0
			Rpa12	NRPA9	At3g25940	60	60	0	0
			Rpc11	NRPC9a	At4g07950	0	0	11	0
				NRPC9b	At1g01210	0	0	11	0
Pol I only			Rpa49	NRPA13	At3g13940	53	53	0	0
			Rpa34	NRPA14	At5g64680	61	61	0	0
Pol III only			Rpc82	NRPC13	At3g49000	0	0	27	27
			Rpc53	NRPC14a	At4g25180	0	0	57	57
				NRPC14b	At5g09380	0	0	32	32
			Rpc37	NRPC15	At5g49530	0	0	38	38
			Rpc34	NRPC16	At5g23710	0	0	25	25
			Rpc31	NRPC17	At4g01590	0	0	54	33
				NRPC17-like	At4g35680	0	0	8	0

Table 2. Peptide coverage corresponding to RNA polymerase subunits identified in affinity purified Pol I and Pol III complexes.

A. Co-IP tests with NRPC7 (Pol III) and NRPA3 (Pol I)

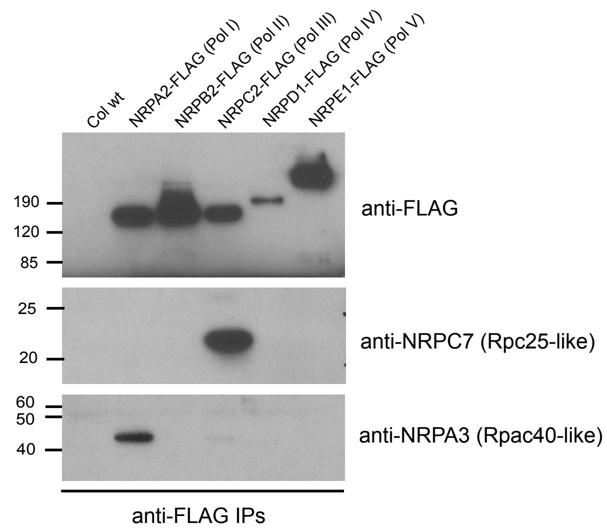
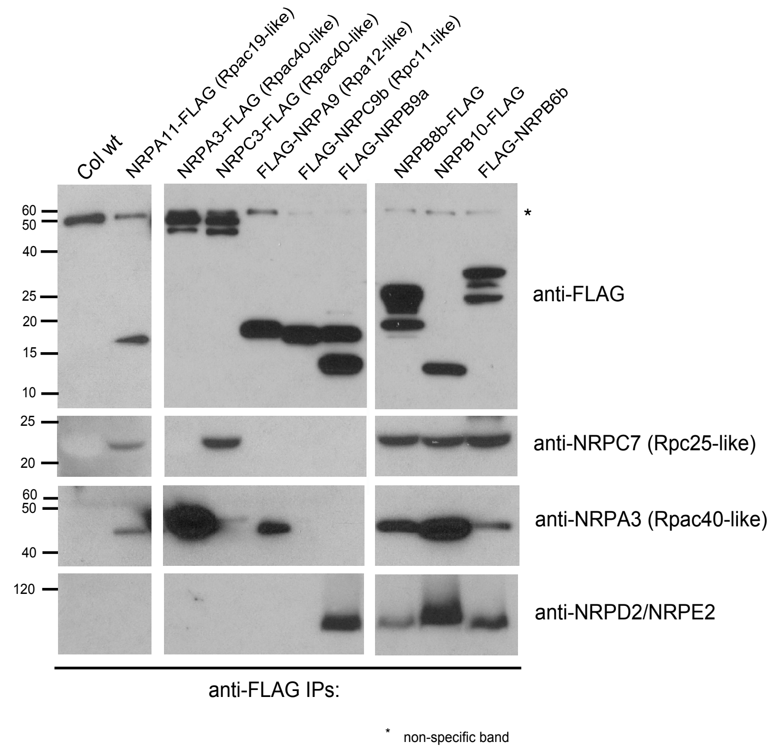


Figure 1



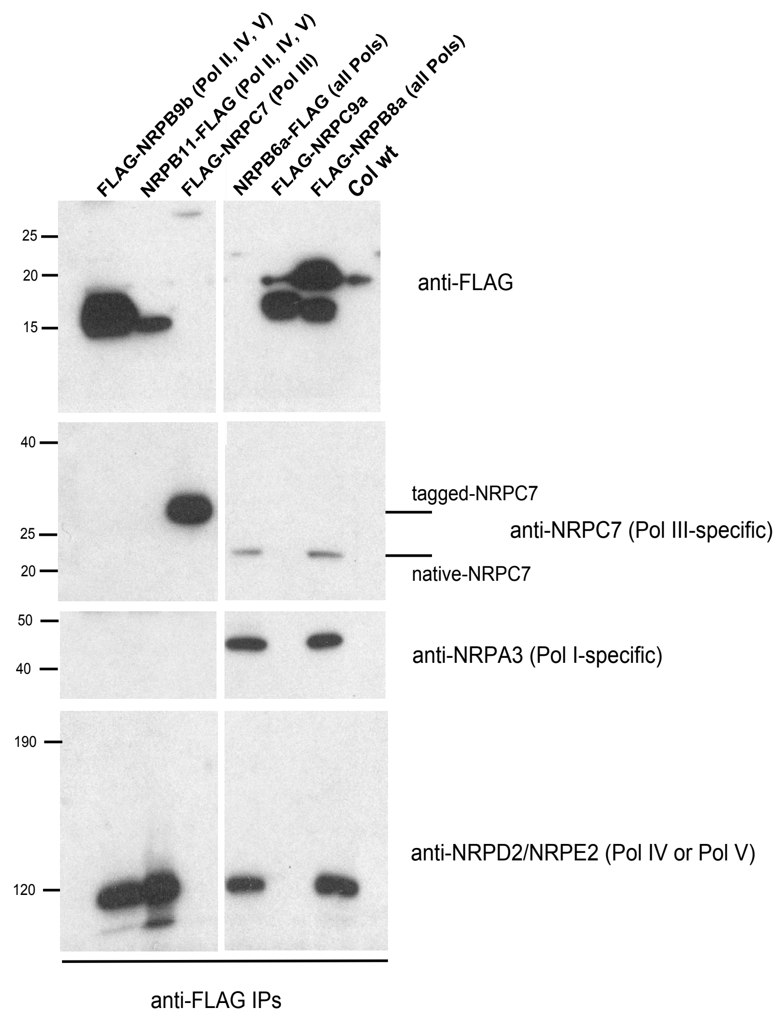
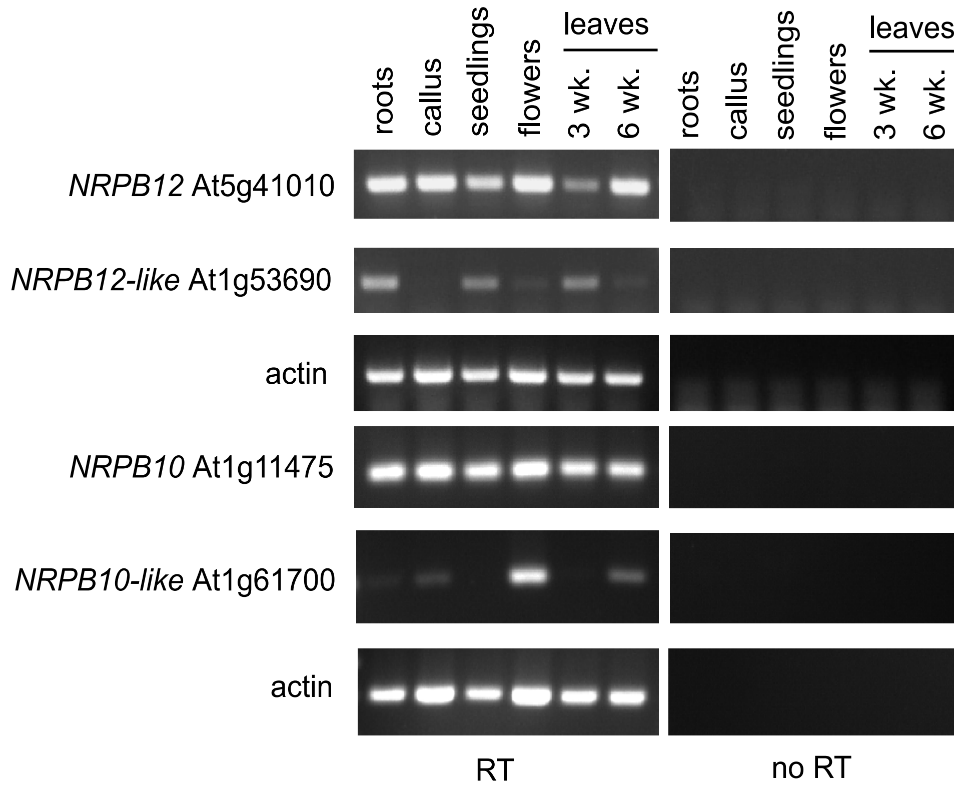


Figure 2

A. Expression of *NRPB10* and *NRPB12* subunits



B. Co-IP tests of variant RNA polymerase subunits

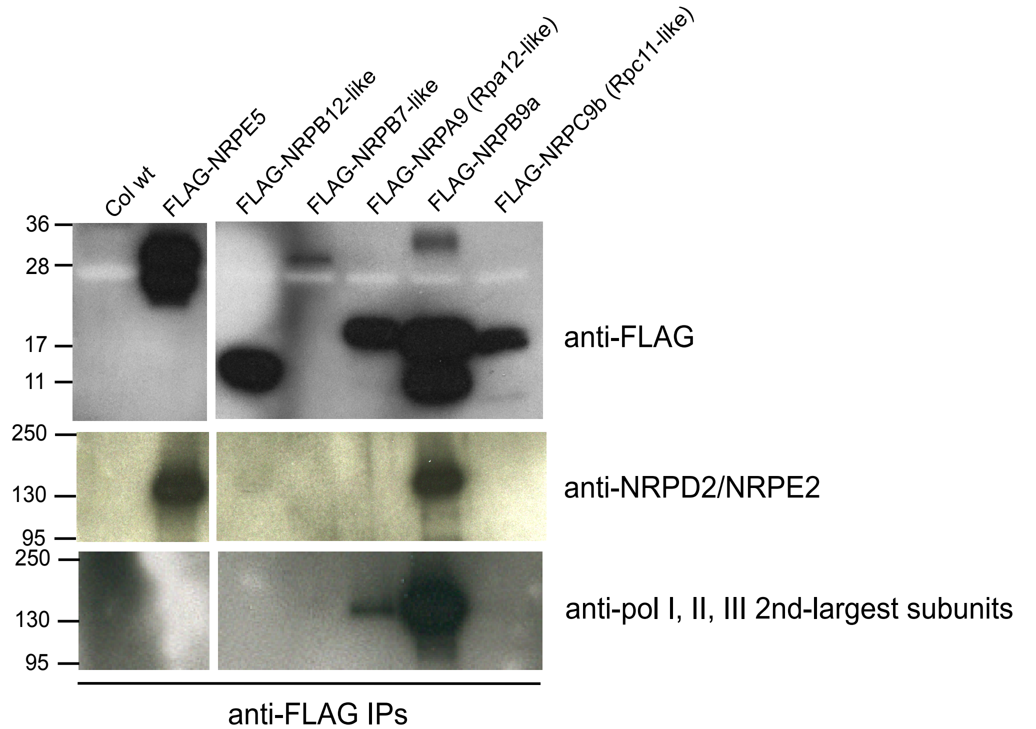


Figure 3

Gene	T-DNA allele	Location	Wild-type	heterozygotes	mutants	total
NRPA3	*nrpa3-1*	exon	39	40	0	79
NRPC3	*nrpc3-1*	intron	59	28	0	89
NRPA11/NRPC11	*nrpa11-1*	intron	58	34	0	92

Table 3. Genotyping analysis of F2 progeny segregating T-DNAs in *NRPA11*, *NRPA3* or *NRPC3* genes.

Bacteria	Archaea	yeast Pol I	yeast Pol II	yeast Pol III	*Arabidopsis thaliana* Pol I	*Arabidopsis thaliana* Pol II	*Arabidopsis thaliana* Pol III	*Arabidopsis thaliana* Pol IV	*Arabidopsis thaliana* Pol V
ß'	RpoA'	Rpa1	Rpb1	Rpc1	NRPA1	NRPB1	NRPC1	NRPD1	NRPE1
	RpoA"								
ß	RpoB'	Rpa2	Rpb2	Rpc2	NRPA2	NRPB2	NRPC2	NRPD2	NRPE2
	RpoB'								
α	RpoD	Rpac40	Rpb3	Rpac40	NRPA3	NRPB3	NRPC3	NRPD3	NRPE3a
									NRPE3b
α	RpoL	Rpac19	Rpb11	Rpac19	NRPA11	NRPB11	NRPA11	NRPD11	NRPE11
	RpoN	Rpb10	Rpb10	Rpb10	NRPA10	NRPB10	NRPC10	NRPD10	NRPE10
	RpoP	Rpb12	Rpb12	Rpb12	NRPA12?	NRPB12	NRPC12	NRPD12	NRPE12
ω	RpoK	Rpb6	Rpb6	Rpb6	NRPA6a	NRPB6a	NRPC6a	NRPD6a	NRPE6a
					NRPA6b	NRPB6b	NRPC6b	NRPD6b?	NRPE6b
	RpoG	Rpb8	Rpb8	Rpb8	NRPA8a	NRPB8a	NRPC8a	NRPD8a?	NRPE8a
					NRPA8b	NRPB8b	NRPC8b	NRPD8b	NRPE8b
	RpoH	Rpb5	Rpb5	Rpb5	NRPA5	NRPB5	NRPC5	NRPD5	NRPE5
	RpoF	Rpa14	Rpb4	Rpc17	?*	NRPB4	NRPC4	NRPD4	NRPE4
	RpoE	Rpa43	Rpb7	Rpc25	NRPA7?	NRPB7	NRPC7	NRPD7	NRPE7
	TFS/X	Rpa12	Rpb9	Rpc11	NRPA9	NRPB9a	NRPC9a?	NRPD9a	NRPE9a
						NRPB9b	NRPC9b?	NRPD9b	NRPE9b
		Rpa49			NRPA13				
		Rpa34			NRPA14				
				Rpc82			NRPC13		
				Rpc53			NRPC14a		
							NRPC14b		
				Rpc37			NRPC15		
				Rpc34			NRPC16		
				Rpc31			NRPC17		

Table 4. Partitioning of DNA-dependent RNA polymerase subunits in the five *Arabidopsis* RNA polymerases compared to the equivalent subunits of yeast, archaea and bacteria.

SUPPLEMENTAL MATERIAL

A. Supplemental Materials and Methods

Liquid chromatography and mass spectrometry details. Two independent affinity-purified NRPA2 samples and one affinity purified NRPC2 sample were analyzed by LC-MS/MS in order to identify Pol I and Pol III subunits, respectively. Affinity purified samples from pEARLEYGATE 302 vector only control lines were also analyzed to identify background contamination levels.

All samples were prepared for analysis using the following procedure: a Coomassie protein assay (Pierce, Rockford, IL) was performed to determine the initial protein concentration of the sample. 2,2,2-trifluoroethanol (TFE) (Sigma, St. Louis, MO) was then added to the sample for a final concentration of 50% TFE. The sample was sonicated in an ice-water bath for 1 min. and incubated at 60ºC for 2 hours with gentle shaking at 300 rpm. The sample was then reduced with 2mM dithiothreitol (DTT) (Sigma, St. Louis, MO) with incubation at 37ºC for 1 hr with gentle shaking at 300rpm. Samples were then diluted 5-fold with 50mM ammonium bicarbonate for preparation for digestion. 1mM $CaCl_2$ and sequencing-grade modified porcine trypsin (Promega, Madison, WI) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio for 3 h at 37˚C. The sample was concentrated in a Speed Vac (ThermoSavant, Holbrook, NY) to a volume of ~30μl and was then centrifuged at 14,000 rpm. The supernatant was removed and added to a sample vial for LC-MS/MS analysis.

Peptide samples were analyzed on a custom-built reversed-phase liquid chromatography (RPLC) system coupled via electrospray ionization (ESI) utilizing an ion

funnel to a ThermoFisher Scientific LTQ-Orbitrap mass spectrometer (ThermoFisher Scientific, San Jose, CA). Briefly, the capillary RPLC separation was performed under a constant pressure of 10,000 psi, using two ISCO (Lincoln, NE) Model 100 DM high-pressure syringe pumps and a column (60 cm × 75 μm i.d.) packed in-house (Pacific Northwest National Laboratory) with Phenomenex (Torrance, CA) Jupiter particles (C18 stationary phase, 5 μm particles, 300 Å pore size). Mobile phase A consisted of 0.1% formic acid in water, and mobile phase B consisted of 100% acetonitrile. The RPLC system was equilibrated at 10,000 psi with 100% mobile phase A. A mobile phase selection valve was switched 50 min after injection to create a near-exponential gradient as mobile phase B displaced A in a 2.5 mL mixer. A split was used to provide an initial flow rate through the column of ~ 400 nL/min. The column was coupled to the mass spectrometer using an in-house manufactured ESI interface with homemade 20 μm i.d. chemically etched emitters. The heated capillary temperature and spray voltage were 200° C and 2.2 kV, respectively. Mass spectra were acquired for 80 min over the *m/z* range 400-2000 at a resolving power of 100K. A maximum of six data-dependent LTQ tandem mass spectra were recorded for the most intense peaks in each survey mass spectrum.

Tandem mass spectra were searched against an *Arabidopsis thaliana* protein file (The Institute for Genomic Research, TIGR 2008 http://www.tigr.org/plantProjects.shtml) containing 27,854 protein sequences after the removal of duplicates. Searching was performed using SEQUEST, allowing for a dynamic oxidation of methionine. In addition, peptide cleavage events were limited to fully tryptic sequences. For the spectra acquired in the Orbitrap, the monoisotopic masses were corrected prior to generation of the dta files used for searching using the

program DeconMSN, developed in house. Peptide sequences were considered confident if the scores passed Xcorr and delcn thresholds described by Washburn et al., which gave a False Discovery Rate (FDR) for all identified peptides of less than 5% and averaged 1.5% based on a reversed database search. Proteins with at least 2 filter passing peptides were considered confidently identified.

Generation of transgenic lines. Plants were transformed by *Agrobacterium tumefaciens* strain GV3101 harboring each transgene-bearing plasmid, using the floral dip method (Clough and Bent, 1998). Transformants were selected by spraying with 0.05% Finale herbicide, containing 5.78% (w/v) glufosinate-ammonium (AgrEvo Environmental Health). Protein assays in the tagged RNA polymerase subunit lines were performed using 3- to 4-week-old pooled T2 progeny derived from single T1 plants or on pooled T1 plants.

Protein alignments presented as supplemental material. Sequences were aligned using ClustalW and highlighted using BOXSHADE. Genbank accession nos. are listed in Table S2. Sequences of RNA polymerase subunits were obtained by BLASTp searches using either *S. cerevisiae* or *A. thaliana* sequences.

Supplemental Table 1. List of primer sequences.

Primer	sequence	Used for:
cNRPB7-like At4g14520-F	CAC CAT GTT TTC AGAQ AGT AGA AAT GGC TCG	cloning *At4g14520* cDNA
cNRPB7-like At4g14520-R	TTA CAT GTA TGG CTC ATC TGA CC	cloning *At4g14520* cDNA
cNRPC7-F	CAC CAT GTT TTA TCT TAG CGA GCT AGA ACA TTC	cloning *NRPC7* cDNA
cNRPC7-R	TTA CTC TTC TTG ATC AAC CTG TTC ATA AG	cloning *NRPC7* cDNA
cNRPC9b-F	CAC CAT GGA GTT TTG TCC AAC ATG TGG G	cloning *NRPC9b* cDNA
cNRPC9b-R	TTA TTC CTC ACG CCA AGT GAA CTC	cloning *NRPC9b* cDNA
cNRPB12-like At1g53690-F	CAC CAT GGA TCT ACA ACA GTC CGA AAC	cloning *At1g53690* cDNA
cNRPB12-like At1g53690-F	TTA AAC ACC TAT ACG GGT TTC CTT C	cloning *At1g53690* cDNA
cNRPA9-F	CAC CAT GGA GAA ATC TAG AGA AAG CGA G	cloning *NRPA9* cDNA
cNRPA9-R	TTA ACC TTC TGT GAA TCT ATG AGC AC	cloning *NRPA9* cDNA
cNRPA3-F	CAC CAT GGT GAC TAA AGC AGA AAA ACA ATT CG	cloning *NRPA3* cDNA
cNRPA3-R	TCA AAA GTC GGT TAT AGC TTC ACA TTT AG	cloning *NRPA3* cDNA
gNRPB12 At5g41010-F	CAC CGG ATA TAA GGT ACA CAA TAA ACC GAT TAA G	cloning *NRPB12* genomic fragment
gNRPB12 At5g41010-R	GCG AGC TTC GTA TTG AAC AAC	cloning *NRPB12* genomic fragment
gNRPA11-F	CAC CCG GGA ATA GGA GAA GCC ATT TTT TTA G	cloning *NRPA11* genomic fragment
gNRPA11-R	GTT GTT TTC AAT GTC CAT GGA C	cloning *NRPA11* genomic fragment
gNRPA3-F	CAC CAT TAC CAA TAT ACC ATT TGC TG	cloning *NRPA3* genomic fragment
gNRPA3-R	AAA GTC GGT TAT AGC TTC ACA TTT AG	cloning *NRPA3* genomic fragment
gNRPC3-F	CAC CGG CTC ACA GCG ATC CAA GTT TGC	cloning *NRPC3* genomic fragment
gNRPC3-R	GGA GAG TTC AGA GAT TAC TCT TTC	cloning *NRPC3* genomic fragment
NRPB10-like At1g61700 F	CAC ATA CCT TGA ACT TCT CCA GGC T	RT-PCR
NRPB10-like At1g61700 R	TTA ATT GGG GTC GGA TTT CTC CAT AG	RT-PCR
NRPB10 At1g11475 F	TCA GTA TCT TGA TCT TCT CCA GCT C	RT-PCR
NRPB10 At1g11475 R	TTA ACT GTT GTC TGA TTT CTC CAG AG	RT-PCR
NRPB12 At5g41010 F	CAC CAT GGA TCC AGC GCC CGA ACC C	RT-PCR
NRPB12 At5g41010 R	TCA GCG AGC TTC GTA TTG AAC AAC	RT-PCR
NRPB12-like At1g53690 F	CAC CAT GGA TCT ACA ACA GTC CGA AAC	RT-PCR
NRPB12-like At1g53690 R	TTA AAC ACC TAT ACG GGT TTC CTT C	RT-PCR
SALK_088247 LP	GAT AAT TCT TCT GAT TAG CTT ACT TCT TTA TC	genotyping *nrpa3-1*
SALK_088247 RP	GTA GCA GAG TAA AAG GAA TGG G	genotyping *nrpa3-1*
SALK_132788 LP	CTT TGT GTC ACG CTT AGT TTT GTT GTT G	genotyping *nrpc3-1*
SALK_132788 RP	GTT TGC AAG TCA AAT TAA TTT GCC GCC	genotyping *nrpc3-1*
WiscDSLox_419G02 LP	CGC CGA CTA ATC AAG CTT TTA G	genotyping *nrpa11-1*
WiscDSLox_419G02 RP	TTC AAG GGA AGG ATG TGG GAT GG	genotyping *nrpa11-1*

Table S2. List of Genbank accession numbers for proteins used in alignments.

	Saccharomyces cerevisiae	Homo sapiens	*Arabidopsis thaliana*	*Vitis vinifera*	*Zea mays*	*Oryza sativa (japonica)*	*Phsycomitrella patens*
Rpa34/NRPAs14	NP_012387	NP_036231	NP_851270	CAO68707	ACG46868	NP_001042641 NP_001061828	XP_001757296
Rpa49/NRPAs13	NP_014151	NP_071935	NP_188010	CAO45674	ACG40611 ACG32045	NP_001068274 NP_001066791	XP_001777438 XP_001771006
Rpc17/NRPC4	NP_012523	NP_055293	NP_201100 NP_683602(NRPC4-like)	CAO42808			
Rpc31/NRPCs17	NP_014248	XP_036456	NP_001031	CAO45800	ACG38101	NP_001060249	XP_001760032
Rpc34/NRPCs16	NP_014400	NP_006457	NP_197760	CAO62747	ACF85449 ACG41460	NP_001058012	XP_001754105
Rpc37/NRPCs15	NP_012950	NP_060589	NP_199764	CAO22879	not identified	EAZ42119	XP_001757939
Rpc53/NRPCs14	NP_010131	NP_001713	NP_194248 NP_196500	CAO70140 CAO65266	NP_001132322 ACF85554	NP_001045045 NP_001046951 NP_001052654	XP_001757951
Rpc82/NRPCs13	NP_015516	NP_006459	NP_566914	CAO15472 CAO71124	not identified	NP_001043863	XP_001775586
Rpac40-like	NP_015435	NP_976035	NP_176261 NP_176282	CAO45155	NP_001130839	NP_001049823 NP_001047973	XP_001773900

Supplemental Figure 1.
ClustalW-aligned Rpa49-like sequences from plants, fungi and animals.

```
A_thaliana_At3g13940_NP_188010 MEGEEKNDYRN---------------------EEEDFNTP
V_vinifera_CAO45674            MEEEE---------------------------PEISSQTL
O_sativa_j_NP_001068274        MADHQPAASSPKRKKKHSKKPEDSNATVDDSLAAAASPSP
Z_mays_ACG40611                MADLETLATP-----------------------APASAST
P_patens_p_XP_001777438        MATILSTTTNESVREN-------CLAARAAERAIMDSTSS
S_cerevisiae_NP_014151         -------------------------------------MSV
H_sapiens_NP_071935            -------------------------------------MAA

A_thaliana_At3g13940_NP_188010 QQLQKKQRKGKKIATET----------------------E
V_vinifera_CAO45674            TLTPTKKRKKK-----------------------------
O_sativa_j_NP_001068274        KKKEKHSKKKREAIDATMAAASPKKKEKKHSKKQEDTNVP
Z_mays_ACG40611                KKKKKHSKKHKAASDAP-----------------------
P_patens_p_XP_001777438        DQHHHHKKKKKSASKV------------------------
S_cerevisiae_NP_014151         KRSVS-----------------------------------
H_sapiens_NP_071935            EVLPS-----------------------------------

A_thaliana_At3g13940_NP_188010 DENEEDKRVTLKVTQVAERPDR-ISPIVAYFSTGYDPCKV
V_vinifera_CAO45674            ------KHVNVKMEVVHEHPEK-IPPLVGYFPSGFDPHKP
O_sativa_j_NP_001068274        EKKREVVHVTVDASLTG-AAAAGAAPVVAYFPTGYDPLAA
Z_mays_ACG40611                ------TNVTVDASLMGGSSYVAAAPVVGYFPSGYDPLTT
P_patens_p_XP_001777438        ------VEQDVRMDYVPSKEGR-ARPFVAYFPTGFDPLGE
S_cerevisiae_NP_014151         ---------EIEIESVQDQPSV----AVGSFFKGFR----
H_sapiens_NP_071935            ----------ARWQYCGAPDGSQRAVLVQFSNGKLQSPGN

A_thaliana_At3g13940_NP_188010 DPKTGKRVHETPKVTVYKHKDDSKKRIQVVVSPP---GAR
V_vinifera_CAO45674            QDHESD---PPPAVKVFRNKARAN-RLQLVVSPN---DSK
O_sativa_j_NP_001068274        --GGGRKGREAPRTRLFRHTKHPS-RIELVVGAATGGGGG
Z_mays_ACG40611                --AAG--AESAPGTQLFRHEKHPT-WVDLVVRSP---GGG
P_patens_p_XP_001777438        A-DHGDAAAALSSVKVYQGVDKLKAKQTQLVAST---HGQ
S_cerevisiae_NP_014151         ---------APSDTTFDLYKKKKSEKDEFVLHGE---NER
H_sapiens_NP_071935            ---------MRFTLYENKDSTNPRKRNQRILAAET---DR

A_thaliana_At3g13940_NP_188010 VEFVG-TNYTGEQAAMQTNTYRVGVFNREAKTLRILPVAH
V_vinifera_CAO45674            VNFVG-TSYSGEAASAQVCTYALGVLDKETQTLKIVPIAS
O_sativa_j_NP_001068274        PDFVG-RSYAGEAVLPQLTGYALGVLDKASGTLKVVPIAA
Z_mays_ACG40611                HDFVG-RSYAGEAAVPQVCDYALGVLDKASGTLRIVPIAG
P_patens_p_XP_001777438        VDFVG-LNYAGEAAAWQPCSYALGVYDKEKGTLQLVPLAG
S_cerevisiae_NP_014151         LEYEG-YTDS---SSQASNQYVVGLFNPEKKSIQLYKAPV
H_sapiens_NP_071935            LSYVGNNFGTGALKCNTLCRHFVGILNKTSGQMEVYDAEL

A_thaliana_At3g13940_NP_188010 N------------------------------KIIRLEPRV
V_vinifera_CAO45674            N------------------------------KIFRLEPRV
O_sativa_j_NP_001068274        N------------------------------KKL------
Z_mays_ACG40611                K------------------------------KILRLEPHL
P_patens_p_XP_001777438        EKGIVDVVRKVRFGLVSGFMVYVGGSKTIGFDILRMEPRV
S_cerevisiae_NP_014151         LVS----------------------------KVVS-----
H_sapiens_NP_071935            FN--------------------------------------

A_thaliana_At3g13940_NP_188010 KAQE-TNEEEASGSAVVKELEELKTGER----DRYNTKKA
V_vinifera_CAO45674            GGSD-ASVDEPSSLLKGEQTAEEKAGRMRDLTNLYGTKKS
O_sativa_j_NP_001068274        -----AHSQHSGAVGEAVSSAGDADLKVQDITKAFGTQKD
Z_mays_ACG40611                EVQHPAHSQHSEVASEAASAAGNDELKVQDLTMMYGAKTD
P_patens_p_XP_001777438        RG---MSYEWTDEEERVEDTREQSLLKRRMLVETFGSQKS
```

```
S_cerevisiae_NP_014151              -----KSSKNLRGPKIKSKSDTRPSALRNALGEAFGTKKA
H_sapiens_NP_071935                 -MQPLFSDVSVESELALESQTKTYREKMDSCIEAFGTTKQ

A_thaliana_At3g13940_NP_188010      VTRDKKKR-------ALYMGDDAATQKVLDGKLSELGVNT
V_vinifera_CAO45674                 ITQAKKLQ-------ALNQKDDPRSQKDLDRKLKEVVINK
O_sativa_j_NP_001068274             KAKDIKWQ-------SLNEQRNDPS-AFMDLDLGNADTSV
Z_mays_ACG40611                     RDKDNKWR-------SLKEQRNDPS-AYEDIDLGISETNV
P_patens_p_XP_001777438             KMRAQRMERGIVKEDALGDTEDMESLFQEAGENEALLTNE
S_cerevisiae_NP_014151              KKAIADLER---NRIDSDKLTDSAIDIVDSVRTASKDLPT
H_sapiens_NP_071935                 KRALNTRRMNRVGNESLNRAVAKAAETIIDTKG-VTALVS

A_thaliana_At3g13940_NP_188010      AALEGTSSTVARNIPPYNTAATTANEAYPLEKIIEKGDWS
V_vinifera_CAO45674                 EALESTSAHSARNIPPYDDTATTPQEAYRLDKIIFKGEWD
O_sativa_j_NP_001068274             GANESQEPT-VRNIPPYDPAADTSERAYLFDEIIPKSIRP
Z_mays_ACG40611                     NANDRQEPVLVRNIPPYDPTADTSENAYLLDEIIPKSMRE
P_patens_p_XP_001777438             QALQQANLSVVRNVPPHDVSANTPEGAYPLDKLIPAEEWE
S_cerevisiae_NP_014151              RAQLDEITSNDRPTPLANIDATDVEQIYPIESIIPKKELQ
H_sapiens_NP_071935                 DAIHNDLQDDSLYLPPCYDDAAKPEDVYKFEDLLSPAEYE

A_thaliana_At3g13940_NP_188010      -----FLEDIYWLLQQETEAATEAYPVFVRNRLYRLRDI-
V_vinifera_CAO45674                 -----YLLDILDLLQAGVEVESSAYSSFVCNRIQKLKEL-
O_sativa_j_NP_001068274             -----HLVDIIGHFESG-EISSKGYGSFVSNRVNKLQEL-
Z_mays_ACG40611                     -----HFLEIVRHFESG-EISSKGYGSFVSKRVHKLKEL-
P_patens_p_XP_001777438             NLDTRELKTAAKKAKDEELLRQQKYPNFVLSRLRRIRVE-
S_cerevisiae_NP_014151              FIR---VSSILKEADKEKKLELFPYQNNSKYVAKKLDSL-
H_sapiens_NP_071935                 ALQ--SPSEAFRNVTSEEILKMIEENSHCTFVIEALKSLP

A_thaliana_At3g13940_NP_188010      KDDMKKQTVCGAATLLTHLVKFKDRN---------SMNGY
V_vinifera_CAO45674                 EDEGEKTTLACIFSYINHLLKYKDRN---------SMD-A
O_sativa_j_NP_001068274             QGE-DKERLAWILSYITHLLSLLARNSSMSKR---HRKEN
Z_mays_ACG40611                     QGE-DKKRFAWILSYIQHLLSLLARNGSMSNRKWKGRNEN
P_patens_p_XP_001777438             GDKEGNDRRAKILVYLRHLFTFHATPHYAVRK--AASDPG
S_cerevisiae_NP_014151              TQPSQMTKLQLLYYLSLLLGVYENRRVNNKTKLLERLNSP
H_sapiens_NP_071935                 SDVESRDRQARCIWFLDTLIKFRAHR----------VVKR

A_thaliana_At3g13940_NP_188010      DSAKDHKMPDIFRQKFNSMFK-DSESDRIPV--------D
V_vinifera_CAO45674                 SSAKHHKIPGILNQKFSSMFALDSDSRRLTE--------E
O_sativa_j_NP_001068274             QATSGPVIPQYVYRKMVLMFT-EPGSSALST--------E
Z_mays_ACG40611                     QISHGPATPQAVYRRLLFMFT-EPGSSVMPT--------E
P_patens_p_XP_001777438             RFSQETGIPGIIVSSFLERFAGESTTDSLNKSEIRIQSKQ
S_cerevisiae_NP_014151              PEILVDGILSRFTVIKPGQFGRSKDRSYFIDP-------Q
H_sapiens_NP_071935                 KSALGPGVPHIINTKLLKHFTCLTYNNGRLRN---LISDS

A_thaliana_At3g13940_NP_188010      KINLLISYVLVLSLHVDNFMTDPEDIAKDLRISTVELRKH
V_vinifera_CAO45674                 KKDLLISYVLVLTLYADEFRTDPSDIARDLRMSAVKLRAH
O_sativa_j_NP_001068274             KHELLINYILVLTLYADDFRSDPKDICEDLKMTRQMIKPY
Z_mays_ACG40611                     KKELLTNYILVLTLFADDFRSNPSDICEDLKMTRQNLKPY
P_patens_p_XP_001777438             HKDLLTSYCLILGLAVDDFQADPYDMALELKQTVNTIRPY
S_cerevisiae_NP_014151              NEDKILCYILAIIMHLDNFIVEITPLAHELNLKPSKVVSL
H_sapiens_NP_071935                 MKAKITAYVIIILALHIHDFQIDLTVLQRDLKLSEKRMMEI

A_thaliana_At3g13940_NP_188010      FLQLGCKFLKQ---------------NSTTVATLPTPLNF
V_vinifera_CAO45674                 FEHLGCKLVSQ---------------NKVTMATLPVPLTF
O_sativa_j_NP_001068274             YDQLGCKSSSAGA-------------FKSSVMTLPAPLKF
Z_mays_ACG40611                     YDQLGCKSVTEGA-------------FKSTFMTLPAPLKF
P_patens_p_XP_001777438             YRELGCKFEARSINERREL-GDPDSAQGKWRVTLPVPLKF
```

```
S_cerevisiae_NP_014151          FRVLGAIVKGATVAQAEAFGIPKSTAASYKIATMKVPFKL
H_sapiens_NP_071935             AKAMRLKISKRRVSVAAGS------EEDHKLGTLSLPLPP

A_thaliana_At3g13940_NP_188010  PEVNRRRRARK---
V_vinifera_CAO45674             PRLRQKRRR-----
O_sativa_j_NP_001068274         PKEATRRKRRRF--
Z_mays_ACG40611                 PKEVTRKRRRQ---
P_patens_p_XP_001777438         PEIVTAKGKGPRR-
S_cerevisiae_NP_014151          PEMTRRGRGPRR--
H_sapiens_NP_071935             AQTSDRLAKRRKIT
```

Supplemental Figure 2.
ClustalW-aligned Rpc17-like sequences from plants, fungi and animals.

```
A_thaliana_At5g62950_NP_201100  -MKIVKANAGALTNFEVLDFLNSRGASKDTTR------VI
A_thaliana_At3g28956_NP_683602  ----MKANAGALTNFELLDFLNSRGASKDTTR------VI
V_vinifera_CAO42808             MSCRKKANAGALTNFEVLDFLRSRGAAKDPTR------VI
O_sativa_j_NP_001066791         -MKIEKTNAGFLTNFEVLDFLRSRGAKTDPMG------CL
Z_mays_ACG32045                 -MKIQKANAGVLTNFEVLDFLRSRGAKIDPMG------CL
P_patens_p_XP_001771006         -MKLKNGNAGLLTNFELMDLFKSRGADRGNLG------VA
H_sapiens_NP_055293             -MEVKDANSALLSNYEVFQLLTDLKEQRKESGKNKHSSGQ
Sc_Rpc17p_NP_012523             -MKVLEERNAFLSDYEVLKFLTDLEKKHLWDQK-----SL

A_thaliana_At5g62950_NP_201100  APIARSEYK-VYDYLVETAASTQTRESINKFADKCKDFK-
A_thaliana_At3g28956_NP_683602  APIARSEYK-VYDYLVETAASTQTRESVNKSADKCKDFK-
V_vinifera_CAO42808             APIAASEFK-VYDYLVESAACNQTRESINELLEKCKKYD-
O_sativa_j_NP_001066791         GAVAASECK-VYEYLLKTPACNQTRESINEFVTRCESFK-
Z_mays_ACG32045                 GAVAVSECK-VYEYILKTPACNQTRESIYEFVKRSEGFR-
P_patens_p_XP_001771006         NTLSPAEMK-VYDYLEKTPAGSQTRENIVAFFKATEQYK-
H_sapiens_NP_055293             QNLNTITYE-TLKYISKTPCRHQSPEIVREFLTALKSHK-
Sc_Rpc17p_NP_012523             AALKKSRSKGKQNRPYNHPELQGITRNVVNYLSINKNFIN

A_thaliana_At5g62950_NP_201100  -LAKAEILNIINLRPSSIVELLPIIENLDDRE-IDTDG--
A_thaliana_At3g28956_NP_683602  -LAKAEILNIINLWPSSIVELLPIIENLDDRE-IDTDG--
V_vinifera_CAO42808             -LAKAEILNIINIRPASPVEIDPIIEESEKRMGEGIED--
O_sativa_j_NP_001066791         -LTNADKLNVINWRPSSAADAYAQNEGILRAIIQ-IWGQD
Z_mays_ACG32045                 -LAEADKLNVINWRPSSAADAYAMIEECGKRFSRDERGEA
P_patens_p_XP_001771006         -LTKAEYLQVSNLRPASAVEVHLIVEDCDERLSSDAVDQ-
H_sapiens_NP_055293             -LTKAEKLQLLNHRPVTAVEIQLMVEESEERLTEEQIEA-
Sc_Rpc17p_NP_012523             QEDEGEERESSGAKDAEKSGISKMSDESFAELMTKLNSFK

A_thaliana_At5g62950_NP_201100  ---------ILELVKDLLPPLPTTASPKD---DDEEE-TE
A_thaliana_At3g28956_NP_683602  ---------ILELVKDLLPPLPTAESPKDN--DEEEE-TE
V_vinifera_CAO42808             ---------LIEMVITVLPPPPNQMNTEEEVVAGDEE-IA
O_sativa_j_NP_001066791         CK--------------------------------------
Z_mays_ACG32045                 CDEDERVQEFLDMVKEVLPAPPKAEVETAEAEAEAEA-EA
P_patens_p_XP_001771006         ---------FLGAIEEILPPIPELEEPAAEGDAGEEDAEA
H_sapiens_NP_055293             ---------LLHTVTSILPAEPEAEQKKNTNSNVAMDEED
Sc_Rpc17p_NP_012523             LFKAEKLQIVNQLPANMVHLYSIVEECDARFDEKTIEEML
```

```
A_thaliana_At5g62950_NP_201100 NGEQS-------------
A_thaliana_At3g28956_NP_683602 NGEQS-------------
V_vinifera_CAO42808            DGEQTTTETTTTTMATEN
O_sativa_j_NP_001066791        ------------------
Z_mays_ACG32045                EAVEAEAEVEAMQE----
P_patens_p_XP_001771006        DGDEMEA-----------
H_sapiens_NP_055293            PA----------------
Sc_Rpc17p_NP_012523            EIISAYA-----------
```

Supplemental Figure 3.
ClustalW-aligned Rpc31-like sequences from plants, fungi and animals.

```
A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 MEGLDSVFLSNQTTHQPVRSASLPSRIHPLSVKLRTALSR
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 LSIWRRSSSSISVSASFGYETVLVGLVNLTELYGCVHELL
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 ESPYVKHTLLHHQEGKLLDESLDGSVLLLDVYEGTREVIV
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 AMREHVTNLKSALRRKGSLEKEAKAYFNLRKKAKKEISKQ
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------
```

```
A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 INALKKMETRDISTNTDQDSAIASTSVLRETIQITVSMFR
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 HLLLFLSTIPPPPSPAIFKTTIGLLSIPFVSPSLSDKSLI
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 LIKEMKSLDDVFLGSILDSRKTLFEVETMENEKMRRDVVE
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 DGFRDLEAELDSVSKCLVKNRVLFLNILNLVCEISQHHSS
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        ----------------------------------------
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 -----MSWKG-ARG-KPKGYGGDYG-------------KP
A_thaliana_At4g35680_isoB_NP_1 REFGSMSWKG-GRG-KPKGYGGDYG-------------KP
O_sativa_j_Os07g0611600_NP_001 -----MASRGRGRG-RGR-RGGGYGFDH--------PAKH
Z_mays_ACG38101                -----MSFRGRGRGGRGRGRGGGFGYDH--------PAKH
V_vinifera_CAO45800            -----MAFRGRGRGRGGYGRGGVGGFG---------IAKQ
P_patens_p_XP_001760032        -----MAARGRGRG-RGRGGGAQFLARDDDGNLIFNKKQE
S_cerevisiae_NP_014248         ----MSSYRGGSRG-GGSNYMSNLPFGLG-----YGDVGK
H_sapiens_XP_036456            ----MAGNKGRGRAAYTFNIEAVGFSKGEK----LPDVVL

A_thaliana_At4g01590_NP_001031 EPFVIFPEIT-LPDPK-SISTDSQL--VQSYFTFNKFWRN
A_thaliana_At4g35680_isoB_NP_1 EPFVIFPEIT-LPDPK-SISTDSQLVVVQSYFTFNKFWMN
O_sativa_j_Os07g0611600_NP_001 TPHEDFPDIT-LPEMTCARATMEEKALIQSTLKFEDFWKT
Z_mays_ACG38101                VPHEDFPDIA-LPEMKCAKASNEEKALIVSTLKLEEFWRS
V_vinifera_CAO45800            EPFVLFPDIE-LPDVS---SVPEEKNLVIRNARLQNYWKS
P_patens_p_XP_001760032        GPTPLFPKVERLPDLP--TVTKRDEMLVFRRHHLQKAWEC
S_cerevisiae_NP_014248         NHITEFPSIP-LPING--PITNKERSLAVKYINFGKTVKD
H_sapiens_XP_036456            KPPPLFPDTDYKPVPL--KTGEGEEYMLALKQELRETMKR
```

```
A_thaliana_At4g01590_NP_001031 SPYHLGDGG------------------VSKKEKESLNIER
A_thaliana_At4g35680_isoB_NP_1 SPYHLCDGG------------------VSKKEKASLDIE-
O_sativa_j_Os07g0611600_NP_001 SCYHLEED-------------------VPKKKNDDKEIER
Z_mays_ACG38101                SCYYLEED-------------------APKKKNEDKEIER
V_vinifera_CAO45800            SPYYLEDI-------------------VPRK-SQSTDIER
P_patens_p_XP_001760032        SPYYIEKI-------------------KTKTEGIVAEIDR
S_cerevisiae_NP_014248         GPFYTGSMSLIIDQQENSKSGKRKPNIILDEDDTNDGIER
H_sapiens_XP_036456            MPYFIETP------------------------EERQDIER

A_thaliana_At4g01590_NP_001031 YSDSLKPKMKS--NKNGSFFDFLVLRPDNFPKELLGD---
A_thaliana_At4g35680_isoB_NP_1 -------------------------RPDNFSKELVGD---
O_sativa_j_Os07g0611600_NP_001 YSD--RKRKTH--SKREALASYLILTPANFPVELVQG---
Z_mays_ACG38101                FSD--RKRKTQ--SKREGLASYLKLAPSNFPAELVQG---
V_vinifera_CAO45800            FSDR-VKLRTT--LKRDPLEQILRLTSNNFPLELVQG---
P_patens_p_XP_001760032        YSDR-YRTNRR--ANRVPLNSVLKLTGASFPVELLGQGWT
S_cerevisiae_NP_014248         YSDKYLKKRKIGISIDDHPYNLNLFPNELYNVMGINKKKL
H_sapiens_XP_036456            YSKRYMKVYKE-----EWIPDWRRLPREMMPRNKCKKAGP

A_thaliana_At4g01590_NP_001031 -TRREQRPVKRAKWSQ--EADLQKLDVFEKLEAKFK----
A_thaliana_At4g35680_isoB_NP_1 -TRREQRPVKRAKWSQ--EADLQKLDVFEKLESKFK----
O_sativa_j_Os07g0611600_NP_001 -SKRGQPSSKKLRWDR--SSDDQAFEVFEKLEEKHK----
Z_mays_ACG38101                -SRRGQATNKKLRWDK--ESDEQAFEAFEKLEQKHK----
V_vinifera_CAO45800            -MKGATHNKRKVQWNP--ESDMQKLELFEKLEKKLEGQD-
P_patens_p_XP_001760032        FSLSSLASMYQVCGSH--SVDLHRLDQLASLEQRRLKDEE
S_cerevisiae_NP_014248         LAISKFNNADDVFTGTGLQDENIGLSMLAKLKELAEDVDD
H_sapiens_XP_036456            KPKKAKDAGKGTPLTN-TEDVLKKMEELEKRGDGEKSDEE

A_thaliana_At4g01590_NP_001031 ----VEGKEEKEEG---EDDEEV-VES--EGEESDNGDYD
A_thaliana_At4g35680_isoB_NP_1 ----TQGNEEKEDG---EDDEQV-VES--EGEESDNGDYD
O_sativa_j_Os07g0611600_NP_001 ----DGDKKTEKDG---DDEDEH-EEEEVEEE-NSDDDYN
Z_mays_ACG38101                ----DGSK-VEKEG---DDEDEQ-EEEVQEEEENSDDDYN
V_vinifera_CAO45800            ----EKGGKEKKEGENEDEDDEG-GEEAAEEEFSDDGDYN
P_patens_p_XP_001760032        K--GDKDGEEKKPGEGDDGDGEN-EEIEDEEDEIADDDYA
S_cerevisiae_NP_014248         ASTGDGAAKGSKTGEGEDDDLAD-DDFEEDEDEEDDDDYN
H_sapiens_XP_036456            NEEKEGSKEKSKEGDDDDDDDAAEQEEYDEEEQEEENDYI

A_thaliana_At4g01590_NP_001031 Q------------NQDFD--------------DDDDDYNN
A_thaliana_At4g35680_isoB_NP_1 Q------------NQDFD--------------DDEDDYNH
O_sativa_j_Os07g0611600_NP_001 Q------------NIEFD--------------DDDDDWNQ
Z_mays_ACG38101                Q------------NIEFD--------------DDDDDWNQ
V_vinifera_CAO45800            Q------------NIDFD--------------DDEDDFNM
P_patens_p_XP_001760032        QVKSTVKAHRGVMNINLTSTSGFSSVLQTFGFDDDEEYDV
S_cerevisiae_NP_014248         A------------EKYFNN-------------GDDDDYGD
H_sapiens_XP_036456            N-------------SYFE---------------DGDDFGA

A_thaliana_At4g01590_NP_001031 E-DDG-FEEVY-----------------------------
A_thaliana_At4g35680_isoB_NP_1 E-EDGGFEEVY-----------------------------
O_sativa_j_Os07g0611600_NP_001 E-EEA-HEDYYD----------------------------
Z_mays_ACG38101                E-DEA-QEDFYD----------------------------
V_vinifera_CAO45800            E-DDNDGICTFSLLALVICIWDSQTEGRC-----------
P_patens_p_XP_001760032        D-DGGDDEALDSSLAFDSSLPLLLLFLYCGSNCMCRPVTI
S_cerevisiae_NP_014248         E-EDPNEEAAF-----------------------------
H_sapiens_XP_036456            DSDDNMDEATY-----------------------------
```

```
A_thaliana_At4g01590_NP_001031 ----------------------------------------
A_thaliana_At4g35680_isoB_NP_1 ----------------------------------------
O_sativa_j_Os07g0611600_NP_001 ----------------------------------------
Z_mays_ACG38101                ----------------------------------------
V_vinifera_CAO45800            ----------------------------------------
P_patens_p_XP_001760032        RPPHDLPDTNACGSGYLPVQKMNLYEIATGHGFLFGQPHL
S_cerevisiae_NP_014248         ----------------------------------------
H_sapiens_XP_036456            ----------------------------------------

A_thaliana_At4g01590_NP_001031 ---------------------------------
A_thaliana_At4g35680_isoB_NP_1 ---------------------------------
O_sativa_j_Os07g0611600_NP_001 ---------------------------------
Z_mays_ACG38101                ---------------------------------
V_vinifera_CAO45800            ---------------------------------
P_patens_p_XP_001760032        GQKRSKVGFCCLSTEIFQLFGDSLSVSHGQFHL
S_cerevisiae_NP_014248         ---------------------------------
H_sapiens_XP_036456            ---------------------------------
```

Supplemental Figure 4.
ClustalW-aligned Rpc34-like sequences from plants, fungi and animals.

```
P_patens_p_XP_001754105        --------MQPGLITWCSTQCTRHCTYSVHTGGFLHESPV
H_sapiens_NP_006457            ----------------------------------MAEVKV
A_thaliana_At5g23710_NP_197760 ----------------------------------------
V_vinifera_CAO62747            MKFGKRLRWEVEETIPEWRSEFISYKQLKKLLNQIDLELR
Z_mays_ACF85449                ----------------------------------------
Z_mays_ACG41460                ----------------------------------------
O_sativa_j_NP_001058012_Os06g0 ----------------------------------------
S_cerevisiae_NP_014400         --------------------------------MSGMIENG

P_patens_p_XP_001754105        IYIEHSQHLFQVEGEGKLVTAPFRRLRCRTYSAFLGIGLV
H_sapiens_NP_006457            KVQPPDADPVEIENRIIELCHQFP----------------
A_thaliana_At5g23710_NP_197760 --MSKRKRP-----DPKSSG--------------------
V_vinifera_CAO62747            ETDGSNKRPRFTTFDGLGVGGRYVHM--------------
Z_mays_ACF85449                --MPPRKRPPSPPEEPSLAS--------------------
Z_mays_ACG41460                --MPPRKRPALPPDEPSPSS--------------------
O_sativa_j_NP_001058012_Os06g0 --MPPRKRPPPPAKAPEAPP--------------------
S_cerevisiae_NP_014400         LQLSDNAKTLHSQMMSKGIGALFT----------------

P_patens_p_XP_001754105        IQGSRTLVSFRRQAVQLGEEQATWRRGWGLPQRVLDEEIP
H_sapiens_NP_006457            ---------------------------HGITDQVIQNEMP
A_thaliana_At5g23710_NP_197760 ----------------------------------------
V_vinifera_CAO62747            ---MREDKGFIRLFEGEIEKVNTFFVDKEEDYIIKLKELQ
Z_mays_ACF85449                ------------------------------------KPRP
Z_mays_ACG41460                ------------------------------------KHQP
O_sativa_j_NP_001058012_Os06g0 ------------------------------------ETKP
S_cerevisiae_NP_014400         -------------------------------QQELQKQMG
```

```
P_patens_p_XP_001754105          DLEMLS--QALNFLLSKQKLVVFTQ---------------
H_sapiens_NP_006457              HIEAQQRAVAINRLLSMGQLDLLRS---------------
A_thaliana_At5g23710_NP_197760   ---------GDLNEPH------------------------
V_vinifera_CAO62747              EMVANLDVDGDILEVQRHILDFHGEMVLLLHYSLTNFTVK
Z_mays_ACF85449                  D-AQPS---KPDPDSEQSTIN-------------------
Z_mays_ACG41460                  E-AHPSEPSKPSPDSEQPIIS-------------------
O_sativa_j_NP_001058012_Os06g0   KISPPAKPIAAAPPPPKAVLS-------------------
S_cerevisiae_NP_014400           IGSLTDLMSIVQELLDKNLIKLVKQ---------------

P_patens_p_XP_001754105          --------------------------------GDNIVYKE
H_sapiens_NP_006457              --------------------------------NTGLLYRI
A_thaliana_At5g23710_NP_197760   ----------------------------------------
V_vinifera_CAO62747              KHYKKIVEKQRQHRVAHHYIHFMPKVMQQPFFSTDLLYKL
Z_mays_ACF85449                  ----------------------------------------
Z_mays_ACG41460                  ----------------------------------------
O_sativa_j_NP_001058012_Os06g0   ----------------------------------------
S_cerevisiae_NP_014400           --------------------------------NDELKFQG

P_patens_p_XP_001754105          QSADVAVKFKGLSSEDMLVYQSIEAAANQGIWTADLKRRT
H_sapiens_NP_006457              KDSQNAGKMKGSDNQEKLVYQIIEDAGNKGIWSRDIRYKS
A_thaliana_At5g23710_NP_197760   ---------------EKKLLDLIRINQGRGATMFELKREK
V_vinifera_CAO62747              MKEYTNAASQPLLDAERVLYDLIRSKEDMGIWTRDMKRET
Z_mays_ACF85449                  -----PRILAQLPNQERLVYKLIFEAGNKGMWMLDIRKKL
Z_mays_ACG41460                  -----AAILAQLPPQERLAYKLIFEAGNKGMWMLDIRKKL
O_sativa_j_NP_001058012_Os06g0   -----DTVLAALSQHERPIYKLVFAGGDKGMSQTEIRIKT
S_cerevisiae_NP_014400           VLESEAQKKATMSAEEALVYSYIEASGREGIWSKTIKART

P_patens_p_XP_001754105          NLQQPQINKALKNLEGRSLVKAVKTVTNKNRKVYMLFDLT
H_sapiens_NP_006457              NLPLTEINKILKNLESKKLIKAVKSVAASKKKVYMLYNLQ
A_thaliana_At5g23710_NP_197760   TIPATIVTRLIASLRKKNLIKEVANMNNKGVKHYLAMEFE
V_vinifera_CAO62747              NLPDNVVTKSLKALQVKKLIKEVVNIQSKGRKHYMAVEFE
Z_mays_ACF85449                  LMAPNVATKVVRTLVASGLLKEVSDVRHRSRKIFMATDFQ
Z_mays_ACG41460                  LMGPNIATKVVRTLVTRRLLKEVSDVRHRSRKIFMATDFQ
O_sativa_j_NP_001058012_Os06g0   GMPTSTLTKHLR----------------------------
S_cerevisiae_NP_014400           NLHQHVVLKCLKSLESQRYVKSVKSVKFPTRKIYMLYSLQ

P_patens_p_XP_001754105          PSREVTGGAWYTEQDYDAEFVDVVKQQCLQFITRQGLADL
H_sapiens_NP_006457              PDRSVTGGAWYSDQDFESEFVEVLNQQCFKFLQSKAETAR
A_thaliana_At5g23710_NP_197760   PCSELTGGEWYTDGALDLSKIEDLKAKCVMILERHRQRVV
V_vinifera_CAO62747              PSKELTGGAWYVEGNLDTEYIKILKETCVKVLS--KLKVA
Z_mays_ACF85449                  PSAEITGGTWYHDGRLDTDAVTTARRCCQAQVQRLGAATA
Z_mays_ACG41460                  PSAEITGGTWYHDGRLDTDAVTTARRCCQAQVERLGAATV
O_sativa_j_NP_001058012_Os06g0   ----------------------------------------
S_cerevisiae_NP_014400           PSVDITGGPWFTDGELDIEFINSLLTIVWRFISENTFPNG

P_patens_p_XP_001754105          EAIADAIKKS-----------------GITQVELGLEEFK
H_sapiens_NP_006457              ESKQNPMIQRNSSFASSHEVWKYICELGISKVELSMEDIE
A_thaliana_At5g23710_NP_197760   TLEVLCGYFVKE-------------------EKLSVDQTK
V_vinifera_CAO62747              TAEGISESINKTG---------------LCATECSTQQIA
Z_mays_ACF85449                  QMIHQGILKEDP------------------RAGYTIDKVR
Z_mays_ACG41460                  QMIHHGILKDDP------------------RAGYTIDKIR
O_sativa_j_NP_001058012_Os06g0   ----------------------------------------
S_cerevisiae_NP_014400           FKNFENGPKKNVFYAPN-------VKNYSTTQEILEFITA
```

```
P_patens_p_XP_001754105        QIMDTLVLDGDVEETMGG------GTYSSDTLCYR-----
H_sapiens_NP_006457            TILNTLIYDGKVEMTIIAAKEGTVGSVDGHMKLYR-----
A_thaliana_At5g23710_NP_197760 EILKNLILD-NLIMEVKSNGMNEFASTRIGEVCYRLTGK-
V_vinifera_CAO62747            EILRALVLD-NEIMEVRSSGSGEFSSIPIGAVCYRCTNK-
Z_mays_ACF85449                DIIKTMVLD-KVLEEVKSTGAGDFSAVRAGTMCY------
Z_mays_ACG41460                DIIKTMVLD-KVLEEVKSTGTGEFRDVRAGTMCY------
O_sativa_j_NP_001058012_Os06g0 -------------------------DLRAGRVCY------
S_cerevisiae_NP_014400         AQVANVELTPSNIRSLCEVLVYDDKLEKVTHDCYRVTLES

P_patens_p_XP_001754105        ----VAKARIPETSALTNVPCGICPVLHECQPG-GLVSPE
H_sapiens_NP_006457            ----AVNPIIPPT-GLVRAPCGLCPVFDDCHEG-GEISPS
A_thaliana_At5g23710_NP_197760 ---K-FGNGEPRAGAFASIPCGVCPHIAICSPD-GVISPT
V_vinifera_CAO62747            ---GGLGGGGPKIGAMASIPCGVCPRITQCTPD-GIISPK
Z_mays_ACF85449                ----RLVTGAPQGGMMEGIPCGVCPRIHECSPE-GIISPS
Z_mays_ACG41460                ----RLVTGAPQGGTMEGIPCGVCPRIDECSPE-GIISPS
O_sativa_j_NP_001058012_Os06g0 ----RRG-GPVQGGMMERIPCGVCPRIDECSPD-GVISPS
S_cerevisiae_NP_014400         ILQMNQGEGEPEAGNKALEDEEEFSIFNYFKMFPASKHDK

P_patens_p_XP_001754105        TCVYFNDWLSF--
H_sapiens_NP_006457            NCIYMTEWLEF--
A_thaliana_At5g23710_NP_197760 TCVYFQKWLDF--
V_vinifera_CAO62747            SCAYYAKWLDF--
Z_mays_ACF85449                TCVYYKKWLQMDF
Z_mays_ACG41460                TCVYYKNWLQMDF
O_sativa_j_NP_001058012_Os06g0 TCVYYKKWLQMDF
S_cerevisiae_NP_014400         EVVYFDEWTI---
```

Supplemental Figure 5.
ClustalW-aligned Rpc37-like sequences from plants, fungi and animals.

```
A_thaliana_At5g49530_NP_199764 ----------------------MDFDDDDKPK--EVTKTR
V_vinifera_CAO22879            ----------------------MDLDDLDGPNRAPVRRPS
O_sativa_j_EAZ42119            ----------------------------------------
P_patens_p_XP_001757939        MAAPGGEDVEMADAAEGSGPSVFALLNPKREPAAAAPSKG
H_sapiens_NP_060589            ----------------------------------------
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 RFAPGRAGKSKPKPKPEPTADKP-----------------
V_vinifera_CAO22879            RFAP-KSSKPKLIPKPEPSSQP------------------
O_sativa_j_EAZ42119            -----------PEEPKPE----------------------
P_patens_p_XP_001757939        KFVPKVKARPGARVRTRPSAQTEPKVETAPVPVNGESAIP
H_sapiens_NP_060589            ----------------------------------------
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 --------------------------------VQPPPQSQ
V_vinifera_CAO22879            ---------------------------------EPSKLDD
O_sativa_j_EAZ42119            ---------------------------------AAAPMED
P_patens_p_XP_001757939        MVVGDNAKTSSSMDTKGAGEIPLLTVKAEPDVVKSEPTDG
H_sapiens_NP_060589            ----------------------------------------
Sc_Rpc37p_NP_012950            ----------------------------------------
```

```
A_thaliana_At5g49530_NP_199764 TESVSKTEHDVDAKFVGTKVETEVCNGSVKMEIDSKVDKE
V_vinifera_CAO22879            VTLINKKEEDSQPPLV-KKADEPFASSTQNGVIEMEVEAK
O_sativa_j_EAZ42119            SMEVDGVG--------------PSASASAEGVGAEEVE--
P_patens_p_XP_001757939        VPLLNAKEEPCDEADLKLKLESDMSKQSIKEEVKEEPPSV
H_sapiens_NP_060589            -----------------------MAN--------------
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 PEIMETELMEEDQQLPLQEEKEEEEEEDDVVVREIDVFFK
V_vinifera_CAO22879            AEVED--AMEED------------DRDEDPVVREIDVFFT
O_sativa_j_EAZ42119            -------------------------DEEDYVLREIDVYFT
P_patens_p_XP_001757939        KEVDDELKDEPMVVDSVVS-----KPGVDRVVREIDVFLT
H_sapiens_NP_060589            -------------------------EEDDPVVQEIDVYLA
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 P-SIDANTQLYVLQYPLRPSWRPYEMDERCEEVRVNPSTS
V_vinifera_CAO22879            PPPTDDNTKIYVLQYPLRPYWRPYELGERCDEVRVKPASA
O_sativa_j_EAZ42119            PKPFDEDTMLYVMQYPLRPCWRPYELNEICKEVRVKPLSS
P_patens_p_XP_001757939        P-RVDPETNLYVLQYPLRPYWRPYNLEERCLEVRIKPQQT
H_sapiens_NP_060589            K---SLAEKLYLFQYPVRPASMTYDDIP-HLSAKIKPKQQ
Sc_Rpc37p_NP_012950            ---MSIDNKLFVTE----------EDEEDRTQDRADVEDE

A_thaliana_At5g49530_NP_199764 QVEIDLSMDVHSKNYD------------------SNFG--
V_vinifera_CAO22879            EVEVDLSIDVDSKNYD------------------PDAPNA
O_sativa_j_EAZ42119            KVEVDLDINTECENYD------------------PEVPLP
P_patens_p_XP_001757939        KFEVDLDIDTTSENYD------------------QDREQH
H_sapiens_NP_060589            KVELEMAIDTLNPNYCRSKGEQIALNVDGACADETSTYSS
Sc_Rpc37p_NP_012950            SNDIDMIADENGTNSA------------------------

A_thaliana_At5g49530_NP_199764 LNMTKQTLKTTWKQPPTLDYAVGVLSGDKLHLNPVHAVAQ
V_vinifera_CAO22879            QRMKKQTLSTSWKPPQTTGYVVGVLMGNKLHLNPIHAVVQ
O_sativa_j_EAZ42119            SRLTEQTLSS-SKAADVADYAVGVLRGNLVHLNHIDAVMQ
P_patens_p_XP_001757939        LQIEKQTLAS-SKVAMTTSYAIGILRGNRLHLNPVQAVVQ
H_sapiens_NP_060589            KLMDKQTFCSSQTTSNTSRYAAALYRQGELHLTPLHGILQ
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 LRPSMQSLSSDKKKKQEEST--------------EESVG-
V_vinifera_CAO22879            LRPSMEYLSSGGAKRKNNVTGDTNSAVKLEVSNEEESVG-
O_sativa_j_EAZ42119            LRPSMLHVNSGRSNARQAHGG----------ASSDASGS-
P_patens_p_XP_001757939        FRPSMKYIDEYDAAKKKSNKEAGIADGDEEELFDAEPPES
H_sapiens_NP_060589            LRPSFSYLDKADAKHREREAANEAGDSSQDEAEDDVKQIT
Sc_Rpc37p_NP_012950            --IANEQEEKSEEVKAEDDTG-------------------

A_thaliana_At5g49530_NP_199764 ---------------TSKKQNKGVQQASTDQKPINEETWV
V_vinifera_CAO22879            ---------------SLKKQNK-IMGILDEQNTDAAESWV
O_sativa_j_EAZ42119            ---------------TMPSVKRNEHSEDSKDYTEESEPWI
P_patens_p_XP_001757939        KTELTLLQVNVRRRETERQEATRLQSHAYLKQLDEAEAWI
H_sapiens_NP_060589            VR--------FSRPESEQARQRRVQSYEFLQKKHAEEPWV
Sc_Rpc37p_NP_012950            ----------------EEEEDDPVIEEFPLKISGEEESLH

A_thaliana_At5g49530_NP_199764 SLKYHGLQSEYCSRYLNGMMANGNSSID---FNMSPGTYI
V_vinifera_CAO22879            PLKYHSSESDFSARYLQKMVAEDNSPIQ---FLMSPYEYV
O_sativa_j_EAZ42119            SLTYQQAGSNVARKYHAEMVSDDGGPID---FTMSTSDYV
P_patens_p_XP_001757939        PLEPHGTDSPITEGIRQKMVTSTQDRIN---FNLSPDAYL
H_sapiens_NP_060589            HLHYYGLRDSRSEHERQYLLCPGSSGVENTELVKSPSEYL
Sc_Rpc37p_NP_012950            VFQYAN---------RPRLVGRKPAEHP----FISAARYK
```

```
A_thaliana_At5g49530_NP_199764 NELCR--------GGSSRN-SESKETLKRVLLSLPLKERV
V_vinifera_CAO22879            NSLCP--------PGVSNDDNRSKGLSRRFLLSMPLEERF
O_sativa_j_EAZ42119            MSLCP--------GGSTNSRDKNKSHAIREMLKLPLEERL
P_patens_p_XP_001757939        NTLVPGRASTSTAGSLIEDGSGNVGISRSHLETLVLPEVR
H_sapiens_NP_060589            MMLMP----PSQEEEKDKPVAPSNVLSMAQLRTLPLADQI
Sc_Rpc37p_NP_012950            P--------------------------KSHLWEIDIPLDE

A_thaliana_At5g49530_NP_199764 QKLLCEGS--PLIRYSVLKHYAPE-FSDEDFLGALQEYGR
V_vinifera_CAO22879            KTLLCEGP--PVHRFSALKHFAPD-NSVEDVLGVLTKHAH
O_sativa_j_EAZ42119            KKWFTEVS--QVNRFDALMHLAPL-YSEDDVLKVLPVYAD
P_patens_p_XP_001757939        FEYLLRVSRVQVLQFERLMKMAPAGCSDEELLNVLSNFAV
H_sapiens_NP_060589            KILMKNVK---VMPFANLMSLLGPSIDSVAVLRGIQKVAM
Sc_Rpc37p_NP_012950            QAFYNKDK--------------------------------

A_thaliana_At5g49530_NP_199764 LVQGLWTPKTRLLKLDGP-----------VEAARDYVLSL
V_vinifera_CAO22879            LVQGLWVPKSSLLFPGGQGV---------APLARDYILLL
O_sativa_j_EAZ42119            LVRGLWVCKSSLLFDDGY------------AWKRDRILLE
P_patens_p_XP_001757939        LVQGCWVAASHIRPEYRGG----------LRAMRDYILFL
H_sapiens_NP_060589            LVQGNWVVKSDILYPKDSSSPHSGVPAEVLCRGRDFVMWK
Sc_Rpc37p_NP_012950            -AESEWNG--------------------------------

A_thaliana_At5g49530_NP_199764 FSQNTTIKYSEVE--ATGDKMKPLMERMLTEFAKERHVLK
V_vinifera_CAO22879            FSKGPLINYSQVD--VPGSLNKALKG-FLSILAVERPSFR
O_sativa_j_EAZ42119            FRKKDSIPLKDIHRIIKVKLDDNLKKKFLYPLCKIRAKLE
P_patens_p_XP_001757939        FSKNRVVKHEQLR---GLPLTKGALREIMVPIAVQRAGVG
H_sapiens_NP_060589            FTQSRWVVRKEVAT--VTKLCAEDVKDFLEHMAVVRINKG
Sc_Rpc37p_NP_012950            -VNVQTLKGVGVEN--NGQYAAFVKDMQVYLVPIERVAQL

A_thaliana_At5g49530_NP_199764 DWKFKEPTDVSFIKSYPEIVKEQDIFWTDKRENLKSRITA
V_vinifera_CAO22879            DWKFKEPTDVSFMKLHPDIVREQRRVWESLEKQITDIIHG
O_sativa_j_EAZ42119            DCKFVLPVDSSFIRRYPHIVKEQDHAWSVRETTMR-----
P_patens_p_XP_001757939        -WEFQEDTDKSFIKKHQSVVKEQTAQWVNSEPGIVGALVE
H_sapiens_NP_060589            -WEFILPYDGEFIKKHPDVVQRQHMLWTGIQAKLEKVYNL
Sc_Rpc37p_NP_012950            KPFFKYIDDANVTRKQEDARRNPN----------------

A_thaliana_At5g49530_NP_199764 QGGKSRADKRRNVVGT---SSSVTVKPEVPTTLSDK----
V_vinifera_CAO22879            GG------RRGPVTKN---SGKPSIANKTGTSTKSD----
O_sativa_j_EAZ42119            -----------ESQET---SSNTEARKTKNTTKSNI----
P_patens_p_XP_001757939        MNPSFPEVHGFEPSET---FEGNKKLNTKSGKQGNSSLVS
H_sapiens_NP_060589            VKETMPKKPDAQSGPAGLVCGDQRIQVAKTKAQQNHALLE
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 ----------------------------------------
V_vinifera_CAO22879            ----------------------------------------
O_sativa_j_EAZ42119            ----------------------------------------
P_patens_p_XP_001757939        HSPKPPSGKIQGNKAGP-----------------------
H_sapiens_NP_060589            RELQRRKEQLRVPAVPPGVRIKEEPVSEEGEEDEEQEAEE
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 ----------------------GGSSKNTIHRVVTQEMPE
V_vinifera_CAO22879            ----------------------GATRATNGTPVRRTAMSA
O_sativa_j_EAZ42119            ----------------------PSKGPDPNMNKARDGPVQ
P_patens_p_XP_001757939        --------------LGEAGPSGDGHDKSNVENGAQGTMSD
H_sapiens_NP_060589            EPMDTSPSGLHSKLANGLPLGRAAGTDSFNGHPPQGCAST
Sc_Rpc37p_NP_012950            ----------------------PSSQRAQVVTMSVKSVND
```

```
A_thaliana_At5g49530_NP_199764 ELKKALPKALKKVFQTHKVCRYETICQGLRDLAVSTSNNP
V_vinifera_CAO22879            ETREALPKALEEVFRLHKVCSFQLICQALRDLAVSKSTSP
O_sativa_j_EAZ42119            GNENLVHSVLDTIFTANKVRSIQAIRRDLRQLAAKYASDR
P_patens_p_XP_001757939        ETRMALPGALKEIFAKHHVCTMQIIVQSLRDMAIER-TVA
H_sapiens_NP_060589            PVARELKAFVEATFQRQFVLTLSELKRLFNLHLASLPPGH
Sc_Rpc37p_NP_012950            PSQNRLTGSLLAHKVADEEANIELTWAEG-----------

A_thaliana_At5g49530_NP_199764 KADS---GMAVNVALAVDAYQGELEDVINGVATNIHGSYV
V_vinifera_CAO22879            KVDP---RTFVAAAYGADAPQEELHAIISKYATTVHGVYV
O_sativa_j_EAZ42119            KDGPKLQALSNAATNCASFPLVDLQKSLNQVAVPVHGVYV
P_patens_p_XP_001757939        NSNPKSAAAAVAAAKAANAPASELTAAVCRVATNIGGVYF
H_sapiens_NP_060589            TLFSGISDRMLQDTVLAAGCKQILVPFPPQTAASPDEQKV
Sc_Rpc37p_NP_012950            ----------------------------------------

A_thaliana_At5g49530_NP_199764 SISSPDHPEYDSLREVVISLLTGSPPGTKLMKAEVFAAGR
V_vinifera_CAO22879            WKTS-EHSEHNPLRKVVVDLFCGKEPNAKLKRADIMEAAK
O_sativa_j_EAZ42119            AKPA----KPNSPRNILIKLFRDKDPDSKLTKQEILDCAA
P_patens_p_XP_001757939        LTTLN-NPALDPFREVVIALLRKMGPDAGLRKNDILLAFK
H_sapiens_NP_060589            FALWESGDMSDQHRQVLLEIFSKN---YRVRRNMIQSRLT
Sc_Rpc37p_NP_012950            --------TFEQFKDTIVKEAEDKTLVALEKQEDYIDNLV

A_thaliana_At5g49530_NP_199764 TKLEREITNNEYIKVMHEICETNSSGWVLQKAR----
V_vinifera_CAO22879            IALKKEITTHEYNKVVSELCESRGSLWVLKSGDGKPK
O_sativa_j_EAZ42119            NHLKKGLNEKDYHQVWRESVIVFKFCPVTA-------
P_patens_p_XP_001757939        SAG-REAPSATYLKVMKELCYTRGAAWYLKPGDGRPA
H_sapiens_NP_060589            QECGEDLSKQEVDKVLKDCCVSYGGMWYLKGTVQS--
Sc_Rpc37p_NP_012950            -------------------------------------
```

Supplemental Figure 6.
ClustalW-aligned Rpc53-like sequences from plants, fungi and animals.

```
O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            ----------------------------------------
V_vinifera_CAO70140            ----------------------------------------
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ----------------------------------------
O_sativa_j_NP_001052654_Os04g0 ----------------------------------------
Z_mays_ACF85554                ----------------------------------------
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        MRGNDEEHSGHTCRGSSVGNAETVGACCPAEAGVLEPQFR
H_sapiens_NP_001713            ----------------------------------------
S_cerevisiae_NP_010131         ----------------------------------------
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------
```

```
O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            ----------------------------------------
V_vinifera_CAO70140            ----------------------------------------
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ----------------------------------------
O_sativa_j_NP_001052654_Os04g0 ----------------------------------------
Z_mays_ACF85554                ----------------------------------------
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        SPRRRSCCVYCHAIQFHALTASIAARVVCGVANILCWIVG
H_sapiens_NP_001713            ----------------------------------------
S_cerevisiae_NP_010131         ---------------------------------MSSNKG
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------

O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            -------------------------------------MDK
V_vinifera_CAO70140            ----------------------------------------
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ----------------------------------------
O_sativa_j_NP_001052654_Os04g0 --------------------------------MDSKGKV
Z_mays_ACF85554                --------------------------------MDDKNKV
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        LGRGMIAAAVAMERKAVTRTVPPPPEGSNKKVQPPGPAAT
H_sapiens_NP_001713            --------------------------------PKEEVTVK
S_cerevisiae_NP_010131         NGRLPSLKDSSSNGGGSAKPSLKFKPKAVARKSKEEREAA
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------

O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            DGSSERSRPDHGDVKMLKFRPKVPRQKQKPKQQDPKPIDE
V_vinifera_CAO70140            ---MDHNESSSVSPRKVRFAPKSPPRRKPKTTAPQPVVAE
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ---MDSGEQK--SKRRFQPNPPRPSRRLPIAPTSNTEAEE
O_sativa_j_NP_001052654_Os04g0 KKETDG---LPPRKGGLKFAPKVRPKKAPKIVPKTEPAEE
Z_mays_ACF85554                KKEIDGPLGPPPRKGGLKFAPKVPQKKPAKVVPKKEPVEE
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        PLTLSAEAYTHGGVKKTKFEPKVPARRLKKLSEKVKSEPR
H_sapiens_NP_001713            KEKRERDRDRQREGHGRGRGRPEVIQSHSIFEQGPAEMMK
S_cerevisiae_NP_010131         ASKVKLEEESKRGNDKKHFNNKNKRVTGAGGQQRRMAKYL
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------

O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            DVMKILRTRQVAAKSAPNTEDECSPQKTLSTPPSADVVCL
V_vinifera_CAO70140            EEDEAKRAQYLLR----RVNEKLRRQGPKVEKTSSVQVVF
V_vinifera_CAO65266            ----------------------------MVPYVTLDVAF
A_thaliana_At4g25180_NP_194248 DEENIKASRQFDR----RIVG---RRPKTETKASSPEVAF
O_sativa_j_NP_001052654_Os04g0 SKDETVDKELLMKLKTSQSTDPFVRKFKTEKKEMRTQVAF
Z_mays_ACF85554                RKEDALDKELLMKLKMSQNKYP-SRIKSDEKPRSGTQVAF
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        EGEGGGEIPKELQRLLQQAENDGRGTFRPNQRTSKVAFGF
H_sapiens_NP_001713            KKGNWDKTVDVSDMGPSHIINIKKEKRETDEETKQILRML
S_cerevisiae_NP_010131         NNTHVISSGPLAAGNFVSEKGDLRRGFIKSEGSGSSLVQK
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------
```

```
O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            SP---------AQLGVQKQNQLPLQ---------------
V_vinifera_CAO70140            GPGAATPSDTIRTFGVHRDGNSDKSSGM------------
V_vinifera_CAO65266            AG------------GLYGSGLSD-----------------
A_thaliana_At4g25180_NP_194248 QP--SLSPLAIRSFGVPKEDDKPNS---------------
O_sativa_j_NP_001052654_Os04g0 GQ---GNSSYARSFPMQS--SADGS---------------
Z_mays_ACF85554                GQ---GNSYYARNFPMPKKDSSAGG---------------
A_thaliana_At5g09380_NP_196500 GV---VNSTRSNKYLNRS-NGAYGS---------------
P_patens_p_XP_001757951        GGFSSRSSLSSAGFQHRSSSGGGKGGGGGGGGG-----G
H_sapiens_NP_001713            EKDDFLDDPGLRNDTRNMPVQLPLAHSG------------
S_cerevisiae_NP_010131         GLETIDNGAESSENEAEDDDNEGVASKSKKKFNMGKEFEA
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------

O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            ----------------------------------------
V_vinifera_CAO70140            ----------ELKVSTPDHEEIAVSSPSTTKPDETNGYFA
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ----------DVNPSSPASILPAVSSVTAAQEDG-----E
O_sativa_j_NP_001052654_Os04g0 ----------------------------------------
Z_mays_ACF85554                ----------------------------------------
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        GGGGGGGGGGGGGGGGGGSSGGGGGGASRGVEKHEFLHAD
H_sapiens_NP_001713            ----------------------------------------
S_cerevisiae_NP_010131         RNLIEDEDDGESEKSSDVDMDDEEWRSKRIEQLFPVRPVR
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------

O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            --------------IPRSFPVPVNSGLVYEEESS-DDDDD
V_vinifera_CAO70140            DATDDSAQIRKRYKEPWDYVHSYYPTTLPLRKPH-SGDPE
V_vinifera_CAO65266            ---------HKEYKEPWDY-YTYYPVTLPLRRPY-SGNPE
A_thaliana_At4g25180_NP_194248 EVHNFVTRTGDDYVEPWDYRNSYYPTVLPLRKPN-SGDIE
O_sativa_j_NP_001052654_Os04g0 -----ASKLPKEYVEPWDYTHSDYPVTLPLRRPY-SGDPE
Z_mays_ACF85554                -----ASKLPKEYAEPWDYSH-DYPVTLPLRKPY-SGNPE
A_thaliana_At5g09380_NP_196500 -----TSTQEIEYKEPWDYYS-YYPITLPMRRPY-AGDPE
P_patens_p_XP_001757951        GIGGDSKRGVRKNVEVFDLER-YFPVTLPLRRPF-AGDSQ
H_sapiens_NP_001713            --WLFKEENDEPDVKPWLAGPKEEDMEVDIPAVK-VKEEP
S_cerevisiae_NP_010131         VRHEDVETVKREIQEALSEKPTREPTPSVKTEPVGTGLQS
O_sativa_j_NP_001046951_Os02g0 ----------MKRGGDGRACPCCYVVTEYINRSS---LVR

O_sativa_j_NP_001045045_Os01g0 -------------------------------EELNLLEQ-
Z_mays_NP_001132322            NDSDNV--GLLETQPDSIESEA----LTCPAEELDLLQQE
V_vinifera_CAO70140            ILDEAE--FGEASTNLEYDEK-----TINPASELGLLEES
V_vinifera_CAO65266            LLDEEE--FGEASESTAYDEN-----STNPAMELGLMDEN
A_thaliana_At4g25180_NP_194248 LLDQEE--FGEVAKNRDYDEN-----TINSAEELGLTSVQ
O_sativa_j_NP_001052654_Os04g0 ILNEEE--FGESSATGAQDG------ELTTAEELGLMMPA
Z_mays_ACF85554                ILDEEE--FGESSASRTQDD------ELSAAEQLGLMDMS
A_thaliana_At5g09380_NP_196500 VLDVEE--FMQ--AGGHHED------SLNTAANLGLMEDS
P_patens_p_XP_001757951        VLDEEE--FGEDDEVPEEDEED----VTEAAKELGLTEKQ
H_sapiens_NP_001713            RDEEEE--AKMKAPPKAARKTPGLPKDVSVAELLRELSLT
S_cerevisiae_NP_010131         YLEERERQVNEKLADLGLEKEFQSVDGKEAAAELELLNAD
O_sativa_j_NP_001046951_Os02g0 VVQEER--------------------WLLGSIMANILTGG
```

```
O_sativa_j_NP_001045045_Os01g0 GD------------TRRMLLFQLPKSLPLPR---------
Z_mays_NP_001132322            GS------------KERMFFFQFPNSLPLPK---------
V_vinifera_CAO70140            E-------------KGRMFLFQLPANLPLFK---------
V_vinifera_CAO65266            Q-------------EASMLFLQLPATMPMIK---------
A_thaliana_At4g25180_NP_194248 HS------------KKQMFIFKIPDCLPVVK---------
O_sativa_j_NP_001052654_Os04g0 S-------------LP--LPKQPDSVAETDKGDGVDA---
Z_mays_ACF85554                S-------------ESQFLFFQLPSSLPLPRQPQSGA---
A_thaliana_At5g09380_NP_196500 G-------------EQKMFFMRLPSVPLAS----------
P_patens_p_XP_001757951        E-------------GDQLFFFQLPKGLPVSLN--------
H_sapiens_NP_001713            K-------------EEELLFLQLPDTLPGQPPTQDIKPIK
S_cerevisiae_NP_010131         HQHILRKLKKMNNKPERFMVFQLPTRLPAFERPAVKEE--
O_sativa_j_NP_001046951_Os02g0 P------------------LFLIPTLHRHVR---------

O_sativa_j_NP_001045045_Os01g0 --------ISAAVERNGKARDKEVKEGSNLKELPQG-YLG
Z_mays_NP_001132322            --------RSSSAG-------KEVMEGSNLQQLPQG-YLG
V_vinifera_CAO70140            --------QSPTRSSLSSKSIGTSEHSCRLEDLAGG-HIG
V_vinifera_CAO65266            --------QAAT-----AEDAGSIQKTCRLEELPSG-FMG
A_thaliana_At4g25180_NP_194248 --------QTTGATTKRSVREYSSGISNPFEGLPEG-FMG
O_sativa_j_NP_001052654_Os04g0 -----EPTSTSSKEMHAGTR-PPKVLGSKLKDLPEG-FMG
Z_mays_ACF85554                -----DPNEVRD-EKREDMR-PSSHNGSKLKEVPEG-YMG
A_thaliana_At5g09380_NP_196500 -------TPTENLETRPNIKGPVEKKTVDLKALPEG-YMG
P_patens_p_XP_001757951        -----------------ASGKKEAAGSAKLEDLGKAGAMG
H_sapiens_NP_001713            TEVQGEDGQVVLIKQEKDREAKLAENACTLADLTEG-QVG
S_cerevisiae_NP_010131         ---KEDMETQASDPSKKKKNIKKKDTKDALSTRELAGKVG
O_sativa_j_NP_001046951_Os02g0 ------------------------AQLVHVVGVTPG----

O_sativa_j_NP_001045045_Os01g0 KMVVYKSGKIKMKLGDVMFDVNPGEECRMAQHVAAINTKE
Z_mays_NP_001132322            KMLVYKSGKIKMKLGDIIFDVNPGVESRMLQRAVALNTRD
V_vinifera_CAO70140            KMLVYKSGAIKLKLGEILYDVSPGLDCTCVQDVVAINTVD
V_vinifera_CAO65266            KMLVYKSGAIKLKLGDTLYDVSPGLDCVFAQDVVAINTED
A_thaliana_At4g25180_NP_194248 KMLVYKSGAVKLKVGDALFDVSPGPGTKIPNDVVAIDIKG
O_sativa_j_NP_001052654_Os04g0 KILVYRSGKVKMKIGDSLFDVSPGSNCMFVQEVAAINARE
Z_mays_ACF85554                KMLVYKSGKVKMKIGDTLFDVSFGSNCMFSQEVAAINIRE
A_thaliana_At5g09380_NP_196500 KLLVYKSGAVKMKLGEVLYDVSPGLKSEFAQDVMVVNTEQ
P_patens_p_XP_001757951        KLLIYESGAVKFKVGDVIFDALPGTEVTFAQELAAVNITT
H_sapiens_NP_001713            KLLIRKSGRVQLLLGKVTLDVTMGTACSFLQELVSVGLGD
S_cerevisiae_NP_010131         SIRVHKSGKLSVKIGNVVMDIGKGAETTFLQDVIALSIAD
O_sativa_j_NP_001046951_Os02g0 ---FHRRHEEEEQHDRRNRCVSAGMKCEFVQEVVPINTRE

O_sativa_j_NP_001045045_Os01g0 K---HCCLLGEIESRHVVVTPDVDSLLLNDNRG-------
Z_mays_NP_001132322            K---HCCLLGEIENRHVIVTPNVDS-LLNDK---------
V_vinifera_CAO70140            K---HCYALGELG-KRVIVTPDVDSLLDSMIALD------
V_vinifera_CAO65266            K---CCCVLGELK-KRAVVTPDVDSALSSMDDLG------
A_thaliana_At4g25180_NP_194248 R---NCSRIGSSA-KFVTVTPDVESLLNPASDMETQK---
O_sativa_j_NP_001052654_Os04g0 K---HCCTLGEIS-KRAIVTPDIEHLLDSFDKMEADRPEI
Z_mays_ACF85554                K---HCCTVGEIS-KRAIIITPDINYMLGSVDKMEE-----
A_thaliana_At5g09380_NP_196500 K---NCCLVGDVY-KHAVLTPDIDSILKDIENI-------
P_patens_p_XP_001757951        K---HCGFLGEVY-KRVVLTPDISNVLPDMTDLNTGP---
H_sapiens_NP_001713            SRTGEMTVLGHVK-HKLVCSPDFESLLDHKHR--------
S_cerevisiae_NP_010131         DAS-SAELLGRVD-GKIVVTPQI-----------------
O_sativa_j_NP_001046951_Os02g0 K---HFCSLGKIK-KHLIGTTDIDNLLDK-----------
```

```
O_sativa_j_NP_001045045_Os01g0 ----------------------------------------
Z_mays_NP_001132322            ----------------------------------------
V_vinifera_CAO70140            ----------------------------------------
V_vinifera_CAO65266            ----------------------------------------
A_thaliana_At4g25180_NP_194248 ----------------------------------------
O_sativa_j_NP_001052654_Os04g0 RPIGLETGPLDMQHMNLFIVFLWSSNWKVSQPSPEAKT
Z_mays_ACF85554                ----------------------------------------
A_thaliana_At5g09380_NP_196500 ----------------------------------------
P_patens_p_XP_001757951        ----------------------------------------
H_sapiens_NP_001713            ----------------------------------------
S_cerevisiae_NP_010131         ----------------------------------------
O_sativa_j_NP_001046951_Os02g0 ----------------------------------------
```

Supplemental Figure 7.
ClustalW-aligned Rpc82-like sequences from plants, fungi and animals.

```
V_vinifera_CAO15472            ----------------------------------------M
V_vinifera_CAO71124            ----------------------------------------M
A_thaliana_At3g49000_NP_566914 ---------------------------------------MS
O_sativa_j_NP_001043863_Os01g0 ----------------------------------------M
H_sapiens_NP_006459            -----------------------------------------
S_cerevisiae_Rpc82p_NP_015516  MDELLGEALSAENQTGESTVESEKLVTPEDVMTISSLEQR

V_vinifera_CAO15472            ATQYGVQLAVHLISSHFGNLVAKVCDCLLRKGTLTLASVI
V_vinifera_CAO71124            TPEYGIKLAVHLISTDFGDHAAKVCDCLLRKGTLALDDIV
A_thaliana_At3g49000_NP_566914 MSEFGIVYAIHIITVQFGSVVSKVCECLLRKGPLSSRDIS
O_sativa_j_NP_001043863_Os01g0 VSQHGILLAVAIISDHFGPLVSKVCRCLLRHGALPLQEIV
H_sapiens_NP_006459            MTQAEIKLCSLLLQEHFGEIVEKIGVHLIRTGSQPLRVIA
S_cerevisiae_Rpc82p_NP_015516  TLNPDLFLYKELVKAHLGERAASVIGMLVALGRLSVRELV

V_vinifera_CAO15472            RFTE--LPPQQVKSCILVLIQHNCVQAFAIQQEVG---FG
V_vinifera_CAO71124            RTAK--LSPQQVKTYLKVLIQHNCVQAFAMEHEGG---CG
A_thaliana_At3g49000_NP_566914 RLAESDINHNKVKDILYLLIQHNCVQAFSIEPPD-----G
O_sativa_j_NP_001043863_Os01g0 RRLE--LSPGQVKNSLLVLIQHNCVQAFNAPR-------G
H_sapiens_NP_006459            HDTG--TSLDQVKKALCVLVQHNLVSYQVHKR--------
S_cerevisiae_Rpc82p_NP_015516  EKIDG-MDVDSVKTTLVSLTQLRCVKYLQETAISGKKTTY

V_vinifera_CAO15472            EAPRVVTQYMALFHNILHRMRFAKFLAIVSEELDKECEEI
V_vinifera_CAO71124            EGPVVVTQYMLLFDNILHRMRFPKLLAIVSKELDKECEEL
A_thaliana_At3g49000_NP_566914 SESKAIVQYIVLFNNILHRVRYNKFSRIVNEELDSKCGAV
O_sativa_j_NP_001043863_Os01g0 NGDKTVTHYLAIFDNIVHRQRFSKFLSIIRADIP-ESEAL
H_sapiens_NP_006459            ----GVVEYEAQCSRVLRMLRYPRYIYTTKTLYSDTGELI
S_cerevisiae_Rpc82p_NP_015516  YYYNEEGIHILLYSGLIIDEIITQMRVNDEEEHKQLVAEI

V_vinifera_CAO15472            LEGLLQHGRLTLEQILDRAKSNQ----------------K
V_vinifera_CAO71124            FEAFLKHGWLTLQQILDRVPSNQNEG-------------E
A_thaliana_At3g49000_NP_566914 LDGLLSNGRLTLGQFIERDRDSG----------------N
O_sativa_j_NP_001043863_Os01g0 LEGLLQNGRLTFGQLVERTISKVPEGS------------I
H_sapiens_NP_006459            VEELLLNGKLTMSAVVKKVADRLTETMED---------GK
S_cerevisiae_Rpc82p_NP_015516  VQNVISLGSLTVEDYLSSVTSDSMKYTISSLFVQLCEMGY
```

```
V_vinifera_CAO15472                PVVQDALRESFLKLVRAHFVERCPASEPFLAPPSEEETPA
V_vinifera_CAO71124                PFVQDALRGSFTKLVNAHYVERCPAPDPFLVPPDEEKTPA
A_thaliana_At3g49000_NP_566914     TIGSEAIRDSLQKLVAARFVERIPSPEPVLG--NKEKEPA
O_sativa_j_NP_001043863_Os01g0     TPAREEIRMNFNKLVFAHYVERCPKPEPFFDPLVDEQSTS
H_sapiens_NP_006459                TMDYAEVSNTFVRLADTHFVQRCPS-----VPTTENSDPG
S_cerevisiae_Rpc82p_NP_015516      LIQISKLHYTPIEDLWQFLYEKHYKNIPRNSPLSDLKKRS

V_vinifera_CAO15472                RKRGAKSAKLVEEPQTI-EQRAIAAAAPMDAKRF---SVI
V_vinifera_CAO71124                RKHDANSAKALTDLCQIGEQCDIAKATPSKAKGF---PVI
A_thaliana_At3g49000_NP_566914     KRRGAKAAKILKEPETL-EEQVVEAATPVDAIRF---PLI
O_sativa_j_NP_001043863_Os01g0     SRK--RAPKTVEIALSI-DKKVVNTAALSDAERFSEIPYI
H_sapiens_NP_006459                PPPPAPTLVINEK-----DMYLVPKLSLIGKGKR------
S_cerevisiae_Rpc82p_NP_015516      QAKMNAKTDFAKIINKPNELSQILTVDPKTSLRIVKPTVS

V_vinifera_CAO15472                INTGTDVD--------------------------------
V_vinifera_CAO71124                TDTCGDINGD------------------------------
A_thaliana_At3g49000_NP_566914     FEEDSNSS--------------------------------
O_sativa_j_NP_001043863_Os01g0     MEDASNAN--------------------------------
H_sapiens_NP_006459                -RRSSDED--------------------------------
S_cerevisiae_Rpc82p_NP_015516      LTINLDRFMKGRRSKQLINLAKTRVGSVTAQVYKIALRLT

V_vinifera_CAO15472                -------DSPNVTAGEK-----VELDKETG-ASCEKDVLW
V_vinifera_CAO71124                ---KSIGNPPGVVIGKKRKQDALKLDTESG-AASEKDVPW
A_thaliana_At3g49000_NP_566914     ----LADDNSNITEGKR-----KQRDVDTS-DSSS-GVIW
O_sativa_j_NP_001043863_Os01g0     ------DSPRSSISGAKRKHNALEGDAELDSTIAENEVLW
H_sapiens_NP_006459                ---------AAGEPKAKR-----PKYTTDNKEPIPDDGIYW
S_cerevisiae_Rpc82p_NP_015516      EQKSPKIRDPLTQTGLLQDLEEAKSFQDEAELVEEKTPGL

V_vinifera_CAO15472                RANFEEFVRCLRHKACIANVKTRLDDGAAIVLSAMLEATR
V_vinifera_CAO71124                RANYKELVCHLRHKACIENVRARLDSGAAIVLSAMLEATQ
A_thaliana_At3g49000_NP_566914     RPNFEEFIHRLRHKACVEIVKERRDEGCAIVLRAMLEVGR
O_sativa_j_NP_001043863_Os01g0     RANFEKFTFCLKKKFCADRKKPKLKVGTHPIWEAFFEASL
H_sapiens_NP_006459                QANLDRFHQHFRDQAIVSAVANRMDQTSSEIVRTMLRMSE
S_cerevisiae_Rpc82p_NP_015516      TFNAIDLARHLPAELDLRGSLLSRKPSDNKKRSGSNAAAS

V_vinifera_CAO15472                SEEKKVKTENSVPLSMDTIFEGVMKSEAGRSMTLERVRGS
V_vinifera_CAO71124                SAEKNVKEEISVLLSISTITEAVMQSEAGCSMTLERIRAS
A_thaliana_At3g49000_NP_566914     SQEKKVKTDNSAPMSVGSIYEEVIKTEAGRTMLQERVEAC
O_sativa_j_NP_001043863_Os01g0     ME----RDNNSVTSPINGIMERLGQKEGGTSMTLDHITRV
H_sapiens_NP_006459                ITTSS-SAPFTQPLSSNEIFRSLP---VGYNISKQVLDQY
S_cerevisiae_Rpc82p_NP_015516      LPSKKLKTEDGFVIPALPAAVSKSLQESGDTQEEDEEEED

V_vinifera_CAO15472                LIQLGCAP--------------------CVRGAAESYSVG
V_vinifera_CAO71124                LVQLGCHP--------------------NGVGADRTYNID
A_thaliana_At3g49000_NP_566914     LDQLSATSSYLPA---------------FVNEVNDSYIVD
O_sativa_j_NP_001043863_Os01g0     LEELNCSP---------------------SSEDPDSFILD
H_sapiens_NP_006459                LTLLADDPLEFVGK--------------SGDSGGGMYVIN
S_cerevisiae_Rpc82p_NP_015516      LDADTEDPHSASLINSHLKILASSNFPFLNETKPGVYYVP

V_vinifera_CAO15472                ----IILPSSISICIMLPCMYFFIAFAACYLL--CCSLLN
V_vinifera_CAO71124                LKKIIELAQNDEVESMVLKRYGREAHRIFRLLSKSCCPLE
A_thaliana_At3g49000_NP_566914     YKSIISVAQKDEIEAVVMRRYGKEAFRMFRYLSQEGRFVE
O_sativa_j_NP_001043863_Os01g0     LSRIVEASRNEEIESLVRKKYGQEAFTIFRLLVRERGPVE
H_sapiens_NP_006459                LHKALASLATATLESVVQERFGSRCARIFRLVLQKKHIEQ
S_cerevisiae_Rpc82p_NP_015516      YSKLMPVLKSSVYEYVIASTLGPSAMRLSRCIR-DNKLVS
```

```
V_vinifera_CAO15472            GDSISDTTFVEKKDTAKILYQLWKDEYLQMEKLNLHGS-R
V_vinifera_CAO71124            TTKISDTTFIEKKEAAKILYKLWKDECLHMEKIVLNES-R
A_thaliana_At3g49000_NP_566914 TDKIADAALTEKKDTPQFLLKMWKDGYLHMQKLAVTAPGM
O_sativa_j_NP_001043863_Os01g0 TDKIIDTTILDKQIVHGTLYKLWKDDYIDTERILSGTGTG
H_sapiens_NP_006459            -KQVEDFAMIPAKEAKDMLYKMLSENFMSLQEIPKTPDHA
S_cerevisiae_Rpc82p_NP_015516  EKIINSTALMKEKDIRSTLASLIRYNSVEIQEVPRTADRS

V_vinifera_CAO15472            QSQ-FLLWKVTKDTLWGHVLNEMYHSALNLSLRVAYELEQ
V_vinifera_CAO71124            QSQ-FMLWKVKKDTFWRHILDEMYHAALNLSIQIIEHYGY
A_thaliana_At3g49000_NP_566914 YTP-FLLWKVNKLIVTTQMLDEMYHASLNLSLRLAHELEA
O_sativa_j_NP_001043863_Os01g0 NTQ-YFVWRVKN-TFREQFIDNLCHAALNLRQMVNYIAE-
H_sapiens_NP_006459            PSRTFYLYTVNILSAARMLLHRCYKSIANLIERRQFETKE
S_cerevisiae_Rpc82p_NP_015516  ASRAVFLFRCKETHSYNFMRQNLEWNMANLLFKKEKLKQE

V_vinifera_CAO15472            EQEIIQLPR-----------------------EKRVGALG
V_vinifera_CAO71124            LCNILQLRR-----------------------DKHVGALR
A_thaliana_At3g49000_NP_566914 EKELLLLPL-----------------------DKLEGPLK
O_sativa_j_NP_001043863_Os01g0 ----LLLEG-----------------------SKDD----
H_sapiens_NP_006459            NKRLLEKSQRVEAIIASMQATGAEEAQLQEIEEMITAPER
S_cerevisiae_Rpc82p_NP_015516  NSTLLKKANRDDVK-----------------GRENELLLP

V_vinifera_CAO15472            NRFERLRKVRILLESSLMKLDDAIMLFNDF--------
V_vinifera_CAO71124            ---ERDVRLAILLQASLMKLDDALMLFHDF--------
A_thaliana_At3g49000_NP_566914 ERLKKVKAKRLLLSSTMFKLDDAIMLFHDF--------
O_sativa_j_NP_001043863_Os01g0 ---TKLRNRKNILILALTRHDDSLMLFQDF--------
H_sapiens_NP_006459            QQLETLKRNVNKLDASEIQVDETIFLLESYIECTMKRQ
S_cerevisiae_Rpc82p_NP_015516  SELNQLKMVNERELNVFARLSRLLSLWEVFQMA-----
```

Supplemental Figure 8.
ClustalW-aligned Rpac40-like sequences from plants, fungi and animals.

```
A_thaliana_At1g60620_NP_176261 MGTNEVTRIVTDEEKREAKNFNIFDLPDVPTG-LPPHLEL
A_thaliana_At1g60850_NP_176282 --------MVTKAEKQFAKNFNIDDLPDVPAG-LPPHLKA
V_vinifera_CAO45155            -----------------MHRSPALDMPDVPKGQLPPHIEL
O_sativa_j_NP_001049823_Os03g0 --------MPKKGTKSRKGTDD--EPKQEEKNKLPEHLEV
O_sativa_j_NP_001047973_Os02g0 -----------MGSKSSKDTDDDHEPKQQEK-KLPEHLEV
Z_mays_NP_001130839            --------MPKKGTKSKKEKVDD-VPNEQEN-KLPDYLEL
P_patens_p_XP_001773900        ---------------------MADASSGEHEPELPPHLEL
H_sapiens_NP_976035            -------------------------------MAASQAVEE
S_cerevisiae_NP_015435         ---------------------------------MSNIVGI

A_thaliana_At1g60620_NP_176261 QRTRVVCKKDSNIHPTAITFSGAYSSMG-VDNSVRLENFS
A_thaliana_At1g60850_NP_176282 QQTRVVSKNNAPAHTASAIYSGTYVSSTEEDDNVKLGNFY
V_vinifera_CAO45155            QRTRVLCNFDAPTHTENVQYSGAYSSMG-VDNSLRMDQFC
O_sativa_j_NP_001049823_Os03g0 QRTRVVCKGDAPVNTEGFQYAGAFAAMG-IDNSVSADKFC
O_sativa_j_NP_001047973_Os02g0 QRTRVVCKADAPVNTEGFQYAGAYSAMG-IDNSVSAEKFC
Z_mays_NP_001130839            QRTRVVCNADAPIHTQGFQYSGAFAAMG-IDNSVSVDKFC
P_patens_p_XP_001773900        QRTRVVCAPDAPSYTESVQYSGAYMAYG-VDNSLRIEDFC
H_sapiens_NP_976035            MRSRVVLGEFGVRNVHTTDFPGNYSGYD---DAWDQDRFE
S_cerevisiae_NP_015435         EYNRVT-------NTTSTDFPG-FSKDA--ENEWNVEKFK
```

```
A_thaliana_At1g60620_NP_176261 EDFKVDVISLTETDMVFDMIGVHAGIANAFRRILLAELPS
A_thaliana_At1g60850_NP_176282 DNFKVDVVSLTKTDMEFDMIGIDAAFANAFRRILIAEVPS
V_vinifera_CAO45155            NNFRVEVIRLSEDDMEFDMIGIDPSIANAFRRILIAELPT
O_sativa_j_NP_001049823_Os03g0 KNFKVVVNRLTEDDMEFDMIGIDASMANAFRRILIAEVPT
O_sativa_j_NP_001047973_Os02g0 KNFKVEISRLTEDDMEFDMIGIDASIANAFRRILIAELPT
Z_mays_NP_001130839            KNFKIEINRLTEDYMEFDMIGIDASIANAFRRILISEVPT
P_patens_p_XP_001773900        KNLRVEIISMNEEDIEFDIVGIDAAIANAFRRILIAEVPT
H_sapiens_NP_976035            KNFRVDVVHMDENSLEFDMVGIDAAIANAFRRILLAEVPT
S_cerevisiae_NP_015435         KDFEVNISSLDAREANFDLINIDTSIANAFRRIMISEVPS

A_thaliana_At1g60620_NP_176261 MAIEKVYVANNTSVIQDEVLAHRLGLIPIAADPRLFEYLS
A_thaliana_At1g60850_NP_176282 MAIEKVLIAYNTSVIIDEVLAHRMGLIPIAADPRLFEYLS
V_vinifera_CAO45155            MAIEKVLIANNTSIVQDEVLAHRLGLIPINADPRLFDDLS
O_sativa_j_NP_001049823_Os03g0 MAIEKVLMADNTSVIADEVLSHRLGLIPLDADPRLFEYIS
O_sativa_j_NP_001047973_Os02g0 MAIEKVLMVDNTSVIADEVLSHRLGLIPLDADPRHFEYMS
Z_mays_NP_001130839            MAIEKVFMADNNSVIADEVLSHRLGLIPLDADPRLFEYIS
P_patens_p_XP_001773900        MAIEKVYIANNTSVIQDEVLAHRLGLLPIKADPRLFDYKS
H_sapiens_NP_976035            MAVEKVLVYNNTSIVQDEILAHRLGLIPIHADPRLFEYRN
S_cerevisiae_NP_015435         VAAEYVVYFFNNTSVIQDEVLAHRIGLVPLKVDPDMLTWVD

A_thaliana_At1g60620_NP_176261 E---NDQP-NEKNTIVFKLHVKCLKGDPRRK---------
A_thaliana_At1g60850_NP_176282 E---HDQA-NEKNTIVFKLHVKCPKNRPRLK---------
V_vinifera_CAO45155            E---NDTP-NEKNTIVFNLHVRCERGGPRLT---------
O_sativa_j_NP_001049823_Os03g0 E---NDVP-TERNTIVYKLHVSCKKGSPRLT---------
O_sativa_j_NP_001047973_Os02g0 E---NDVP-NERNTIVYKLHVSCKKGSPRLTAQACIVFRR
Z_mays_NP_001130839            E---NDVP-NERNTIVYKLHVSCPKNSQRIT---------
P_patens_p_XP_001773900        E---EENA-NEKNTIVFKIDVVCTRKGERVLN--------
H_sapiens_NP_976035            Q---GDEEGTEIDTLQFRLQVRCTRNPHAAK---------
S_cerevisiae_NP_015435         SNLPDDEKFTDENTIVLSLNVKCTRNPDAPK---------

A_thaliana_At1g60620_NP_176261 ------------------VLTSELKWLPNGSELIKESGGS
A_thaliana_At1g60850_NP_176282 ------------------VLTSDLKWLPNGSELLRESENK
V_vinifera_CAO45155            ------------------GMVIVLCIYAVLNNAVK-----
O_sativa_j_NP_001049823_Os03g0 ------------------VKSGDLEWLPEGSQLPLASPAQ
O_sativa_j_NP_001047973_Os02g0 MLIDIFYRIKFSHSFLCTVKSGDLEWLPEGSRLPLASPAQ
Z_mays_NP_001130839            ------------------VKSSELEWLPEGSQLSMSAPAQ
P_patens_p_XP_001773900        ----------------SSVTSDQMKWLPEGSEMPFN----
H_sapiens_NP_976035            -------------------DSSDPNELYVNHKVYT-----
S_cerevisiae_NP_015435         -------------------GSTDPKELYNNAHVYA-----

A_thaliana_At1g60620_NP_176261 T-TTPKTYTSFNHSQDSFPEFAENPIRPTLKDILIAKLGP
A_thaliana_At1g60850_NP_176282 T-SKPKTYTSFSCSQDSLPEFANNPITPCDLDILIAKLAP
V_vinifera_CAO45155            --SNELKWLPHGSDQDSLPEFSDKSIGPKHPDIIIAKLGP
O_sativa_j_NP_001049823_Os03g0 SGDKQKTYTSFSQSQK---DILEKPLGVKFKDITIARLGP
O_sativa_j_NP_001047973_Os02g0 SRYKQKTYTSFSQSQK---DILEKPLGVKFKDITIARLGP
Z_mays_NP_001130839            SGDKQKTFTSFSQSQK---EILEKPLGVKYKDITIARLGP
P_patens_p_XP_001773900        ---TMTKFTSFGSSQKNLPGLPAEGISCQYPDIIVAKLRP
H_sapiens_NP_976035            ---RHMTWIPLGNQAD---LFPEGTIRPVHDDILIAQLRP
S_cerevisiae_NP_015435         ---RDLKFEPQGRQST---TFADCPVVPADPDILLAKLRP
```

```
A_thaliana_At1g60620_NP_176261 GQEIELEAHAVKGIGKTHAKWSPVATAWYRMLPEVVLLKE
A_thaliana_At1g60850_NP_176282 GQEIELEAHAVKGIGKTHAKWSPVGTAWYRMHPEVVLRGE
V_vinifera_CAO45155            GQEIELEAHAVKGMGKTHAKWSPVATAWYRMLPEVVLLQE
O_sativa_j_NP_001049823_Os03g0 GQAIELEVHAVKGIGKVHAKWSPVATAWYRMLPEVVLREE
O_sativa_j_NP_001047973_Os02g0 GQ--------------------------------VVFRKE
Z_mays_NP_001130839            GQAIELEAHAVKGTGNVHAKWSPVATAWYRMLPEVVILKE
P_patens_p_XP_001773900        GQAIELEGHAIKGSGNIHAKWSPVATARYRMLPEVNLMED
H_sapiens_NP_976035            GQEIDLLMHCVKGIGKDHAKFSPVATASYRLLPDITLLEP
S_cerevisiae_NP_015435         GQEISLKAHCILGIGGDHAKFSPVSTASYRLLPQINILQP

A_thaliana_At1g60620_NP_176261 FEGKHAEELVKVCPKKVFDIEDMGQGRKRATVARPRDCSL
A_thaliana_At1g60850_NP_176282 VEDELAERLVNVCPQNVFDIEDMGKGKKRATVAQPRKCTL
V_vinifera_CAO45155            IEDEKAEELVKKCPVNVFDIEDIAKGKKKATVARPRACTL
O_sativa_j_NP_001049823_Os03g0 IKDADAEMLVAKCPVNVFDIEDLGNGEKRAVVARPRACTL
O_sativa_j_NP_001047973_Os02g0 IKGDNAEKLVKKCPVNVFDIEDLGN---------------
Z_mays_NP_001130839            VEGPQAEELVNKCRVNVFDIEDLANGGKRAVVARPRACTL
P_patens_p_XP_001773900        IQGKLAEELVQKCPTGVFDIEDVGT-VKKAIVARPRNCTL
H_sapiens_NP_976035            VEGEAAEELSRCFSPGVIEVQEVQG-KKVARVANPRLDTF
S_cerevisiae_NP_015435         IKGESARRFQKCFPPGVIGIDEGSD---EAYVKDARKDTV

A_thaliana_At1g60620_NP_176261 CRECIRDGVEWED-----QVDLRRVKNHFIFTIESTGSQP
A_thaliana_At1g60850_NP_176282 CKECVRD-DDLVD-----HVDLGSVKNHFIFNIESTGSLP
V_vinifera_CAO45155            CRECIRG-DDWDK-----HVAIRRVKNHFIFTIESTGALP
O_sativa_j_NP_001049823_Os03g0 CRQCVMGPTGEVMGPTVEQVELRRVRDHFIFTIESTGALP
O_sativa_j_NP_001047973_Os02g0 ------------------------------VTIESTGGLP
Z_mays_NP_001130839            CRECVMGPNEK-------QVELRRVRDHFIFTIESTGSMP
P_patens_p_XP_001773900        CRECVRG-TDWDK-----RVQLTRVKDHFIFTIESTGALR
H_sapiens_NP_976035            SREIFRN-EKLKK-----VVRLARVRDHYIFSVESTGVLP
S_cerevisiae_NP_015435         SREVLRYEEFADK------VKLGRVRNHFIFNVESAGAMT

A_thaliana_At1g60620_NP_176261 PEVLFNEAVKILEDKCERVISELS-----
A_thaliana_At1g60850_NP_176282 PEVLFTEAVKILEAKCE-AITDF------
V_vinifera_CAO45155            PEVLFNEAVKVLEDKCERVITELS-----
O_sativa_j_NP_001049823_Os03g0 PEMLFTEAVRILEEKCERVISELS-----
O_sativa_j_NP_001047973_Os02g0 PEALFTEAVRILEEKCERVISGLS-----
Z_mays_NP_001130839            PEVLFTEAVKILEEKCARVISELS-----
P_patens_p_XP_001773900        PEVLFTEAIKILNEKCHRIITELS-----
H_sapiens_NP_976035            PDVLVSEAIKVLMGKCRRFLDELDAVQMD
S_cerevisiae_NP_015435         PEEIFFKSVRILKNKAEYLKNCPITQ---
```

Figure S9. Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V (NRPA2-FLAG). In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPA1 (At3g57660)

MAHAQTTEVCLSFHRSLLFPMGASQVVESVRFSFMTEQDVRKHSFLKVTSPILHD
NVGNPFPGGLYDLKLGPKDDKQACNSCGQLKLACPGHCGHIELVFPIYHPLLFNL
LFNFLQRACFFCHHFMAKPEDVERAVSQLKLIIKGDIVSAKQLESNTPTKSKSSDE
SCESVVTTDSSEECEDSDVEDQRWTSLQFAEVTAVLKNFMRLSSKSCSRCKGINP
KLEKPMFGWVRMRAMKDSDVGANVIRGLKLKKSTSSVENPDGFDDSGIDALSE
VEDGDKETREKSTEVAAEFEEHNSKRDLLPSEVRNILKHLWQNEHEFCSFIGDLW
QSGSEKIDYSMFFLESVLVPPTKFRPPTTGGDSVMEHPQTVGLNKVIESNNILGNA
CTNKLDQSKVIFRWRNLQESVNVLFDSKTATVQSQRDSSGICQLLEKKEGLFRQ
KMMGKRVNHACRSVISPDPYIAVNDIGIPPCFALKLTYPERVTPWNVEKLREAIIN
GPDIHPGATHYSDKSSTMKLPSTEKARRAIARKLLSSRGATTELGKTCDINFEGKT
VHRHMRDGDIVLVNRQPTLHKPSLMAHKVRVLKGEKTLRLHYANCSTYNADFD
GDEMNVHFPQDEISRAEAYNIVNANNQYARPSNGEPLRALIQDHIVSSVLLTKRD
TFLDKDHFNQLLFSSGVTDMVLSTFSGRSGKKVMVSASDAELLTVTPAILKPVPL
WTGKQVITAVLNQITKGHPPFTVEKATKLPVDFFKCRSREVKPNSGDLTKKKEID
ESWKQNLNEDKLHIRKNEFVCGVIDKAQFADYGLVHTVHELYGSNAAGNLLSV
FSRLFTVFLQTHGFTCGVDDLIILKDMDEERTKQLQECENVGERVLRKTFGIDVD
VQIDPQDMRSRIERILYEDGESALASLDRSIVNYLNQCSSKGVMNDLLSDGLLKT
PGRNCISLMTISGAKGSKVNFQQISSHLGQQDLEGKRVPRMVSGKTLPCFHPWD
WSPRAGGFISDRFLSGLRPQEYYFHCMAGREGLVDTAVKTSRSGYLQRCLMKNL
ESLKVNYDCTVRDADGSIIQFQYGEDGVDVHRSSFIEKFKELTINQDMVLQKCSE
DMLSGASSYISDLPISLKKGAEKFVEAMPMNERIASKFVRQEELLKLVKSKFFAS
LAQPGEPVGVLAAQSVGEPSTQMTLNTFHLAGRGEMNVTLGIPRLQEILMTAAA
NIKTPIMTCPLLKGKTKEDANDITDRLRKITVADIIKSMELSVVPYTVYENEVCSIH
KLKINLYKPEHYPKHTDITEEDWEETMRAVFLRKLEDAIETHMKMLHRIRGIHND
VTGPIAGNETDNDDSVSGKQNEDDGDDDGEGTEVDDLGSDAQKQKKQETDEM
DYEENSEDETNEPSSISGVEDPEMDSENEDTEVSKEDTPEPQEESMEPQKEVKGV
KNVKEQSKKKRRKFVRAKSDRHIFVKGEGEKFEVHFKFATDDPHILLAQIAQQT
AQKVYIQNSGKIERCTVANCGDPQVIYHGDNPKERREISNDEKKASPALHASGV
DFPALWEFQDKLDVRYLYSNSIHDMLNIFGVEAARETIIREINHVFKSYGISVSIRH
LNLIADYMTFSGGYRPMSRMGGIAESTSPFCRMTFETATKFIVQAATYGEKDTLE
TPSARICLGLPALSGTGCFDLMQRVEL

815/1670=49% coverage
all unique

NRPA2 (At1g29940)

MVVNAKDSTVPTMEDFKELHNLVTHHIESFDYMTLKGLDVMFNRIKPVSVYDP
NTENELSIWLENPLVFAPQKESFKSTSRKEPLLPFECRQAKISYTGTFMADVCFKY
NDGVVVRDKFDFGQFPIMLMSKLCSLKGADCRKLLKCKESTSEMGGYFILNGIE
RVFRCVIAPKRNHPTSMIRNSFRDRKEGYSSKAVVTRCVRDDQSSVTVKLYYLR
NGSARVGFWIVGREYLLPVGLVLKALTNSCDEEIYESLNCCYSEHYGRGDGAIGT
QLVRERAKIILDEVRDLGLFTREQCRKHLGQHFQPVLDGVKKESLSIVAEAVLRD
YLFVHLDNDHDKFNLLIFIIQKLYSLVDQTSLPDNPDSLQNQEILVPGHVITIYLKE
KLEEWLRKCKSLLKDELDNTNSKFSFESLADVKKLINKNPPRSIGTSIETLLKTGA
LKTQSGLDLQQRAGYTVQAERLNFLRFLSFFRAVHRGASFAGLRTTTVRKLLPES
WGFLCPVHTPDGTPCGLLNHMTRTSRITSQFDSKGNIRDFLKIRKSVVDVLTGAG
MVPSLPKLVRAGPPKVIHVLLDGQVVGTLSSNLVTKVVSYIRRLKVEAPSVIPED
LEVGYVPTSMGGSYPGLYLASCPARFIRPVKNISIPSDNIELIGPFEQVFMEISCPD
GGNGGRNNSSLATHEEIHPTGMISVVANLTPWSDHNQSPRNMYQCQMAKQTM
AYSTQALQFRADQKIYHLQTPQSPVVRTKTYTTYSIDENPTGTNAIVAVLAHTGF
DMEDAMILNKSSVERGMCHGQIYQTENIDLSDQNSRFDSGSKSFRRSTNKAEHF
RIDADGLPSVGQKLYPDEPYCSIYDEVTNKTRHMKRKGTDPVIVDFVSVDMKSK
KHPQRANIRFRHARNPIIGDKFSSRHGQKGVCSQLWPDIDMPFNGVTGMRPDLII
NPHAFPSRMTIAMLLESIAAKGGSLHGKFVDATPFRDAVKKTNGEEESKSSLLVD
DLGSMLKEKGFNHYGTETLYSGYLGVELKCEIFMGPVYYQRLRHMVSDKFQVR
STGQVDQLTHQPIKGRKRGGGIRFGEMERDSLLAHGASYLLHDRLHTSSDHHIA
DVCSLCGSLLTSSVVNVQQKKLIQEIGKLPPGRTPKKVTCYSCKTSKGMETVAM
PYVFRYLAAELASMNIKMTLQLSDREGVTD

603/1178=51% coverage
all unique

NRPA49 (At3g13940)

MEGEEKNDYRNEEEDFNTPQQLQKKQRKGKKIATETEDENEEDKRVTLKVTQV
AERPDRISPIVAYFSTGYDPCKVDPKTGKRVHETPKVTVYKHKDDSKKRIQVVVS
PPGARVEFVGTNYTGEQAAMQTNTYRVGVFNREAKTLRILPVAHNKIIRLEPRV
KAQETNEEEASGSAVVKELEELKTGERDRYNTKKAVTRDKKKRALYMGDDAA
TQKVLDGKLSELGVNTAALEGTSSTVARNIPPYNTAATTANEAYPLEKIIEKGDW
SFLEDIYWLLQQETEAATEAYPVFVRNRLYRLRDIKDDMKKQTVCGAATLLTHL
VKFKDRNSMNGYDSAKDHKMPDIFRQKFNSMFKDSESDRIPVDKINLLISYVLVL
SLHVDNFMTDPEDIAKDLRISTVELRKHFLQLGCKFLKQNSTTVATLPTPLNFPEV
NRRRRARK

234/442=53% coverage
all unique

NRPA34 (At5g64680)

MDFDFKVSGDFIVSGAEQLDDTDLTRSDEFWLIQAPLGQFPEIEENTLKIEPDKDG
LFGEFKDSNGAKYDLASFHSQDAGAELIIPSEESMIVGKITRRVALVRYPEPNELL
QKMKARTQQKLVGSVTNSSKKSSNLTQSSRHKSGTRSSREKSMFSGFTETPKSPK
RKNSESSSGKHRSSTSTVSGSSERSAKSKKKVKKEE

124/203=61% coverage
all unique

NRPB3 (NRPD3, NRPE3a) (At2g15430)

MDGATYQRFPKIKIRELKDDYAKFELRETDVSMANALRRVMISEVPTVAIDLVEI
EVNSSVLNDEFIAHRLGLIPLTSERAMSMRFSRDCDACDGDGQCEFCSVEFRLSS
KCVTDQTLDVTSRDLYSADPTVTPVDFTIDSSVSDSSEHKGIIIVKLRRGQELKLR
AIARKGIGKDHAKWSPAATVTFMYEPDIIINEDMMDTLSDEEKIDLIESSPTKVFG
MDPVTRQVVVVDPEAYTYDEEVIKKAEAMGKPGLIEISPKDDSFIFTVESTGAVK
ASQLVLNAIDLLKQKLDAVRLSDDTVEADDQFGELGAHMRGG

13/319=4% coverage
all unique

NRPE3b (At2g15400)

MDGVTYQRFPTVKIRELKDDYAKFELRETDVSMANALRRVMISEVPTMAIHLVK
IEVNSSVLNDEFIAQRLSLIPLTSERAMSMRFCQDCEDCNGDEHCEFCSVEFPLSA
KCVTDQTLDVTSRDLYSADPTVTPVDFTSNSSTSDSSEHKGIIIAKLRRGQELKLK
ALARKGIGKDHAKWSPAATVTYMYEPDIIINEEMMNTLTDEEKIDLIESSPTKVF
GIDPVTGQVVVVDPEAYTYDEEVIKKAEAMGKPGLIEIHPKHDSFVFTVESTGAL
KASQLVLNAIDILKQKLDAIRLSDNTVEADDQFGELGAHMREG

13/319=4% coverage
all unique

NRPA3 (At1g60850)

MVTKAEKQFAKNFNIDDLPDVPAGLPPHLKAQQTRVVSKNNAPAHTASAIYSGT
YVSSTEEDDNVKLGNFYDNFKVDVVSLTKTDMEFDMIGIDAAFANAFRRILIAEV
PSMAIEKVLIAYNTSVIIDEVLAHRMGLIPIAADPRLFEYLSEHDQANEKNTIVFKL
HVKCPKNRPRLKVLTSDLKWLPNGSELLRESENKTSKPKTYTSFSCSQDSLPEFA
NNPITPCDLDILIAKLAPGQEIELEAHAVKGIGKTHAKWSPVGTAWYRMHPEVVL
RGEVEDELAERLVNVCPQNVFDIEDMGKGKKRATVAQPRKCTLCKECVRDDDL
VDHVDLGSVKNHFIFNIESTGSLPPEVLFTEAVKILEAKCEAITDF

229/375=61% coverage

all unique

NRPC3 (At1g60620)

MGTNEVTRIVTDEEKREAKNFNIFDLPDVPTGLPPHLELQRTRVVCKKDSNIHPT
AITFSGAYSSMGVDNSVRLENFSEDFKVDVISLTETDMVFDMIGVHAGIANAFRR
ILLAELPSMAIEKVYVANNTSVIQDEVLAHRLGLIPIAADPRLFEYLSENDQPNEK
NTIVFKLHVKCLKGDPRRKVLTSELKWLPNGSELIKESGGSTTTPKTYTSFNHSQ
DSFPEFAENPIRPTLKDILIAKLGPGQEIELEAHAVKGIGKTHAKWSPVATAWYR
MLPEVVLLKEFEGKHAEELVKVCPKKVFDIEDMGQGRKRATVARPRDCSLCRE
CIRDGVEWEDQVDLRRVKNHFIFTIESTGSQPPEVLFNEAVKILEDKCERVISELS

26/385=7% coverage
13/385=3% unique

NRPA5 (NRPB5, NRPC5, NRPD5) (At3g22320)

MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDL
VTLKAKRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLT
PFARTCISEISSKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKE
TQLPRIQVTDPIARYFGLKRGQVVKIIRPSETAGRYVTYRYVV

120/205=59% coverage
all unique

NRPA8a (NRPB8a, NRPC8a, NRPD8a, NRPE8a) (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAV
GDKFTLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTP
KAELYVSFGGLLMLLKGDPAHISHFELDQRLFLLMRKL

95/146=65% coverage
82/146=56% unique

NRPA8b (NRPB8b, NRPC8b, NRPD8b, NRPE8b) (At3g59600)

MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAV
GDKFTLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTP
KAELYVSFGGLLMLLQGDPAHISHFELDQRLFLLMRKL

91/146=62% coverage
54/146=37% unique

NRPA9 (At3g25940)

MEKSRESEFLFCNLCGTMLVLKSTKYAECPHCKTTRNAKDIIDKEIAYTVSAEDI

RRELGISLFGEKTQAEAELPKIKKACEKCQHPELVYTTRQTRSADEGQTTYYTCP
NCAHRFTEG

71/119=60% coverage
all unique

NRPA10 (NRPB10, NRPC10, NRPD10, NRPE10) (At1g11475)

MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHV
DLIEKLLNYNTLEKSDNS

50/71=70% coverage
39/71=55% unique

NRPB10-like (At1g61700)

MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTH
VDLIEKLLNYNTMEKSDPN

11/71=15% coverage
0/71=0% unique

NRPA11 (NRPC11) (At2g29540)

MLVYIHVMLLEILIVFNECHIILSLRLMELICWCVDNDDYVNNQYCFQFCSPRVT
VAAYTIPHPSLEQVNIRVQTTGDPAREVFKDACQELMQMNRHVRSVFDKAVAE
YKDEQKRKEEAEEEELKRQRDLFGSMDIENN

41/139=29% coverage
all unique

Other RNA polymerase subunits:

NRPC1 (At5g60040)

METKMEIEFTKKPYIEDVGPLKIKSINFSVLSDLEVMKAAEVQVWNIGLYDHSFK
PYENGLLDPRMGPPNKKSICTTCEGNFQNCPGHYGYLKLDLPVYNVGYFNFILDI
LKCICKRCSNMLLDEKLYEDHLRKMRNPRMEPLKKTELAKAVVKKCSTMASQR
IITCKKCGYLNGMVKKIAAQFGIGISHDRSKIHGGEIDECKSAISHTKQSTAAINPL
TYVLDPNLVLGLFKRMSDKDCELLYIAYRPENLIITCMLVPPLSIRPSVMIGGIQSN
ENDLTARLKQIILGNASLHKILSQPTSSPKNMQVWDTVQIEVARYINSEVRGCQN
QPEEHPLSGILQRLKGKGGRFRANLSGKRVEFTGRTVISPDPNLKITEVGIPILMA
QILTFPECVSRHNIEKLRQCVRNGPNKYPGARNVRYPDGSSRTLVGDYRKRIADE
LAIGCIVDRHLQEGDVVLFNRQPSLHRMSIMCHRARIMPWRTLRFNESVCNPYN
ADFDGDEMNMHVPQTEEARTEAITLMGVQNNLCTPKNGEILVASTQDFLTSSFLI

TRKDTFYDRAAFSLICSYMGDGMDSIDLPTPTILKPIELWTGKQIFSVLLRPNASIR
VYVTLNVKEKNFKKGEHGFDETMCINDGWVYFRNSELISGQLGKATLGNGNKD
GLYSILLRDYNSHAAAVCMNRLAKLSARWIGIHGFSIGIDDVQPGEELSKERKDSI
QFGYDQCHRKIEEFNRGNLQLKAGLDGAKSLEAEITGILNTIREATGKACMSGLH
WRNSPLIMSQCGSKGSPINISQMVACVGQQTVNGHRAPDGFIDRSLPHFPRMSKS
PAAKGFVANSFYSGLTATEFFFHTMGGREGLVDTAVKTASTGYMSRRLMKALE
DLLVHYDNTVRNASGCILQFTYGDDGMDPALMEGKDGAPLNFNRLFLKVQATC
PPRSHHTYLSSEELSQKFEEELVRHDKSRVCTDAFVKSLREFVSLLGVKSASPPQV
LYKASGVTDKQLEVFVKICVFRYREKKIEAGTAIGTIGAQSIGEPGTQMTLKTFHF
AGVASMNITQGVPRINEIINASKNISTPVISAELENPLELTSARWVKGRIEKTTLGQ
VAESIEVLMTSTSASVRIILDNKIIEEACLSITPWSVKNSILKTPRIKLNDNDIRVLD
TGLDITPVVDKSRAHFNLHNLKNVLPNIIVNGIKTVERVVVAEDMDKSKQIDGKT
KWKLFVEGTNLLAVMGTPGINGRTTTSNNVVEVSKTLGIEAARTTIIDEIGTVMG
NHGMSIDIRHMMLLADVMTYRGEVLGIQRTGIQKMDKSVLMQASFERTGDHLF
SAAASGKVDNIEGVTECVIMGIPMKLGTGILKVLQRTDDLPKLKYGPDPIIS

9/1376=0.6% coverage
0/1376=0% unique

NRPB2 (At4g21710)

MEYNEYEPEPQYVEDDDDEEITQEDAWAVISAYFEEKGLVRQQLDSFDEFIQNT
MQEIVDESADIEIRPESQHNPGHQSDFAETIYKISFGQIYLSKPMMTESDGETATLF
PKAARLRNLTYSAPLYVDVTKRVIKKGHDGEEVTETQDFTKVFIGKVPIMLRSSY
CTLFQNSEKDLTELGECPYDQGGYFIINGSEKVLIAQEKMSTNHVYVFKKRQPNK
YAYVGEVRSMAENQNRPPSTMFVRMLARASAKGGSSGQYIRCTLPYIRTEIPIIIV
FRALGFVADKDILEHICYDFADTQMMELLRPSLEEAFVIQNQLVALDYIGKRGAT
VGVTKEKRIKYARDILQKEMLPHVGIGEHCETKKAYYFGYIIHRLLLCALGRRPE
DDRDHYGNKRLDLAGPLLGGLFRMLFRKLTRDVRSYVQKCVDNGKEVNLQFAI
KAKTITSGLKYSLATGNWGQANAAGTRAGVSQVLNRLTYASTLSHLRRLNSPIG
REGKLAKPRQLHNSQWGMMCPAETPEGQACGLVKNLALMVYITVGSAAYPILE
FLEEWGTENFEEISPSVIPQATKIFVNGMWVGVHRDPDMLVKTLRRLRRRVDVN
TEVGVVRDIRLKELRIYTDYGRCSRPLFIVDNQKLLIKKRDIYALQQRESAEEDG
WHHLVAKGFIEYIDTEEEETTMISMTISDLVQARLRPEEAYTENYTHCEIHPSLILG
VCASIIPFPDHNQSPRNTYQSAMGKQAMGIYVTNYQFRMDTLAYVLYYPQKPLV
TTRAMEHLHFRQLPAGINAIVAISCYSGYNQEDSVIMNQSSIDRGFFRSLFFRSYR
DEEKKMGTLVKEDFGRPDRGSTMGMRHGSYDKLDDDGLAPPGTRVSGEDVIIG
KTTPISQDEAQGQSSRYTRRDHSISLRHSETGMVDQVLLTTNADGLRFVKVRVRS
VRIPQIGDKFSSRHGQKGTVGMTYTQEDMPWTIEGVTPDIIVNPHAIPSRMTIGQL
IECIMGKVAAHMGKEGDATPFTDVTVDNISKALHKCGYQMRGFERMYNGHTGR
PLTAMIFLGPTYYQRLKHMVDDKIHSRGRGPVQILTRQPAEGRSRDGGLRFGEM
ERDCMIAHGAAHFLKERLFDQSDAYRVHVCEVCGLIAIANLKKNSFECRGCKNK
TDIVQVYIPYACKLLFQELMSMAIAPRMLTKHLKSAKGRQ

27/1188=2% coverage

all unique

The following RNA polymerase subunits had zero coverage:

NRPB1 (At4g35800)
NRPE1 (At2g40030)
NRPD1 (At1g63020)

NRPC2 (At5g45140)
NRPD2a (At3g23780)
NRPD2b (At3g18090)

NRPD4 (At4g15950)
NRPB4 (At5g09920)
NRPC4 (At5g62950)
NRPC4-like (At3g28956)

NRPB5-like (At5g57980)
NRPE5 (At3g57080)
NRPE5-like (At2g41340)
NRPE5-like (At3g54490)

NRPA6a (NRPB6a, NRPC6a, NRPD6a, NRPE6a) (At5g51940)
NRPA6b (NRPB6b, NRPC6b, NRPD6b, NRPE6b) (At2g04630)

NRPE7 (At4g14660)
NRPD7 (At3g22900)
NRPB7-like (At4g14520)
NRPB7 (At5g59180)
NRPA7 (At1g75670)

NRPB9a (At3g16980)
NRPB9b (At4g16265)
NRPC9a (At4g07950)
NRPC9b (At1g01210)

NRPB11 (At3g52090)

NRPA12 (NRPB12, NRPC12, NRPD12, NRPE12) (At5g41010)
NRPB12-like (At1g53690)

NRPC82 (At3g49000)

NRPC53 (At4g25180)
NRPC53 (At5g09380)
NRPC37 (At5g49530)
NRPC34 (At5g23710)
NRPC31 (At4g01590)
NRPC31-like (At4g35680)

Figure S10. Peptide coverage maps of DNA-directed RNA polymerase subunits detected by LC-MS/MS in affinity purified Pol V (NRPC2-FLAG). In the full-length protein sequences that follow, peptides highlighted in yellow or green indicate sequenced tryptic peptides that do not overlap with other sequenced peptides. Cyan highlighting denotes sequences represented by two overlapping peptides. Magenta highlighting indicates regions corresponding to three or more overlapping peptide sequences.

NRPC1 (At5g60040)

METKMEIEFTKKPYIEDVGPLKIKSINFSVLSDLEVMKAAEVQVWNIGLYDHSFK
PYENGLLDPRMGPPNKKSICTTCEGNFQNCPGHYGYLKLDLPVYNVGYFNFILDI
LKCICKRCSNMLLDEKLYEDHLRKMRNPRMEPLKKTELAKAVVKKCSTMASQR
IITCKKCGYLNGMVKKIAAQFGIGISHDRSKIHGGEIDECKSAISHTKQSTAAINPL
TYVLDPNLVLGLFKRMSDKDCELLYIAYRPENLIITCMLVPPLSIRPSVMIGGIQSN
ENDLTARLKQIILGNASLHKILSQPTSSPKNMQVWDTVQIEVARYINSEVRGCQN
QPEEHPLSGILQRLKGKGGRFRANLSGKRVEFTGRTVISPDPNLKITEVGIPILMA
QILTFPECVSRHNIEKLRQCVRNGPNKYPGARNVRYPDGSSRTLVGDYRKRIADE
LAIGCIVDRHLQEGDVVLFNRQPSLHRMSIMCHRARIMPWRTLRFNESVCNPYN
ADFDGDEMNMHVPQTEEARTEAITLMGVQNNLCTPKNGEILVASTQDFLTSSFLI
TRKDTFYDRAAFSLICSYMGDGMDSIDLPTPTILKPIELWTGKQIFSVLLRPNASIR
VYVTLNVKEKNFKKGEHGFDETMCINDGWVYFRNSELISGQLGKATLGNGNKD
GLYSILLRDYNSHAAAVCMNRLAKLSARWIGIHGFSIGIDDVQPGEELSKERKDSI
QFGYDQCHRKIEEFNRGNLQLKAGLDGAKSLEAEITGILNTIREATGKACMSGLH
WRNSPLIMSQCGSKGSPINISQMVACVGQQTVNGHRAPDGFIDRSLPHFPRMSKS
PAAKGFVANSFYSGLTATEFFFHTMGGREGLVDTAVKTASTGYMSRRLMKALE
DLLVHYDNTVRNASGCILQFTYGDDGMDPALMEGKDGAPLNFNRLFLKVQATC
PPRSHHTYLSSEELSQKFEEELVRHDKSRVCTDAFVKSLREFVSLLGVKSASPPQV
LYKASGVTDKQLEVFVKICVFRYREKKIEAGTAIGTIGAQSIGEPGTQMTLKTFHF
AGVASMNITQGVPRINEIINASKNISTPVISAELENPLELTSARWVKGRIEKTTLGQ
VAESIEVLMTSTSASVRIILDNKIIEEACLSITPWSVKNSILKTPRIKLNDNDIRVLD
TGLDITPVVDKSRAHFNLHNLKNVLPNIIVNGIKTVERVVVAEDMDKSKQIDGKT

KWKLFVEGTNLLAVMGTPGINGRTTTSNNVVEVSKTLGIEAARTTIIDEIGTVMG
NHGMSIDIRHMMLLADVMTYRGEVLGIQRTGIQKMDKSVLMQASFERTGDHLF
SAAASGKVDNIEGVTECVIMGIPMKLGTGILKVLQRTDDLPKLKYGPDPIIS

304/1376=22% coverage
304/1376=22% unique

NRPC2 (At5g45140)

MGLDQEDLDLTNDDHFIDKEKLSAPIKSTADKFQLVPEFLKVRGLVKQHLDSFN
YFINVGIHKIVKANSRITSTVDPSIYLRFKKVRVGEPSIINVNTVENINPHMCRLAD
MTYAAPIFVNIEYVHGSHGNKAKSAKDNVIIGRMPIMLRSCRCVLHGKDEEELA
RLGECPLDPGGYFIIKGTEKVLLIQEQLSKNRIIIDSDKKGNINASVTSSTEMTKSK
TVIQMEKEKIYLFLHRFVKKIPIIIVLKAMGMESDQEIVQMVGRDPRFSASLLPSIE
ECVSEGVNTQKQALDYLEAKVKKISYGTPPEKDGRALSILRDLFLAHVPVPDNNF
RQKCFYVGVMLRRMIEAMLNKDAMDDKDYVGNKRLELSGQLISLLFEDLFKTM
LSEAIKNVDHILNKPIRASRFDFSQCLNKDSRYSISLGLERTLSTGNFDIKRFRMHR
KGMTQVLTRLSFIGSMGFITKISPQFEKSRKVSGPRSLQPSQWGMLCPCDTPEGES
CGLVKNLALMTHVTTDEEEGPLVAMCYKLGVTDLEVLSAEELHTPDSFLVILNG
LILGKHSRPQYFANSLRRLRRAGKIGEFVSVFTNEKQHCVYVASDVGRVCRPLVI
ADKGISRVKQHHMKELQDGVRTFDDFIRDGLIEYLDVNEENNALIALYESDGTTE
LDEGAEAAKADTTHIEIEPFTILGVVAGLIPYPHHNQSPRNTYQCAMGKQAMGNI
AYNQLNRMDTLLYLLVYPQRPLLTTRTIELVGYDKLGAGQNATVAVMSFSGYDI
EDAIVMNKSSLDRGFGRCIVMKKIVAMSQKYDNCTADRILIPQRTGPDAEKMQIL
DDDGLATPGEIIRPNDIYINKQVPVDTVTKFTSALSDSQYRPAREYFKGPEGETQV
VDRVALCSDKKGQLCIKYIIRHTRRPELGDKFSSRHGQKGVCGIIIQQEDFPFSEL
GICPDLIMNPHGFPSRMTVGKMIELLGSKAGVSCGRFHYGSAFGERSGHADKVE
TISATLVEKGFSYSGKDLLYSGISGEPVEAYIFMGPIYYQKLKHMVLDKMHARGS
GPRVMMTRQPTEGKSKNGGLRVGEMERDCLIAYGASMLIYERLMISSDPFEVQV
CRACGLLGYYNYKLKKAVCTTCKNGDNIATMKLPYACKLLFQELQSMNVVPRL
KLTEA

457/1161=39% coverage
457/1161=39% unique

NRPC3 (At1g60620)

MGTNEVTRIVTDEEKREAKNFNIFDLPDVPTGLPPHLELQRTRVVCKKDSNIHPT
AITFSGAYSSMGVDNSVRLENFSEDFKVDVISLTETDMVFDMIGVHAGIANAFRR
ILLAELPSMAIEKVYVANNTSVIQDEVLAHRLGLIPIAADPRLFEYLSENDQPNEK
NTIVFKLHVKCLKGDPRRKVLTSELKWLPNGSELIKESGGSTTTPKTYTSFNHSQ
DSFPEFAENPIRPTLKDILIAKLGPGQEIELEAHAVKGIGKTHAKWSPVATAWYR
MLPEVVLLKEFEGKHAEELVKVCPKKVFDIEDMGQGRKRATVARPRDCSLCRE
CIRDGVEWEDQVDLRRVKNHFIFTIESTGSQPPEVLFNEAVKILEDKCERVISELS

221/385=57% coverage

221/385=57% unique

NRPC4 (At5g62950)

MKIVKANAGALTNFEVLDFLNSRGASKDTTRVIAPIARSEYKVYDYLVETAASTQTRESINKFADKCKDFKLAKAEILNIINLRPSSIVELLPIIENLDDREIDTDGILELVKDLLPPLPTTASPKDDDEEETENGEQS

76/139=55% coverage
0/139=41% unique

NRPC4-like (At3g28956)

MKANAGALTNFELLDFLNSRGASKDTTRVIAPIARSEYKVYDYLVETAASTQTRESVNKSADKCKDFKLAKAEILNIINLWPSSIVELLPIIENLDDREIDTDGILELVKDLLPPLPTAESPKDNDEEEETENGEQS

19/137=14% coverage
0/137=0% unique

NRPC5 (NRPA5, NRPB5, NRPD5) (At3g22320)

MLTEEELKRLYRIQKTLMQMLRDRGYFIADSELTMTKQQFIRKHGDNMKREDLVTLKAKRNDNSDQLYIFFPDEAKVGVKTMKMYTNRMKSENVFRAILVVQQNLTPFARTCISEISSKFHLEVFQEAEMLVNIKEHVLVPEHQVLTTEEKKTLLERYTVKETQLPRIQVTDPIARYFGLKRGQVVKIIRPSETAGRYVTYRYVV

91/205=44% coverage
91/205=44% unique

NRPC6a (NRPA6a, NRPB6a, NRPD6a, NRPE6a) (At5g51940)

MADEDYNDVDDLGYEDEPAEPEIEEGVEEDVEMKENDDVNGEPIEAEDKVETEPVQRPRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYLPDGSFEEWGVDELIVEDSWKRQVGGD

52/144=36% coverage
23/144=16% unique

NRPC6b (NRPA6b, NRPB6b, NRPD6b, NRPE6b) (At2g04630)

MADDDYNEVDDLGYEDEPAEPEIEEGVEEDADIKENDDVNVDPLETEDKVETEPVQRPRKTSKFMTKYERARILGTRALQISMNAPVMVELEGETDPLEIAMKELRQRKIPFTIRRYLPDMSYEEWGVDELIVEDSWKRQVGGD

51/144=35% coverage

22/144=15% unique

NRPC7 (At1g06790)

MFYLSELEHSLRVPPHLLNLPLEDAIKSVLQNVFLDKVLADLGLCVSIYDIKSVEG
GFVLPGDGAATYKVGLRIVVFRPFVGEVIAAKFKESDANGLRLTLGFFDDIYVPA
PLMPKPNRCEPDPYNRKQMIWVWEYGEPKEDYIVDDACQIKFRVESISYPSVPTE
RAEDAKPFAPMVVTGNMDDDGLGPVSWWDSYEQVDQEE

98/204=48% coverage
98/204=48% unique

NRPC8a (NRPA8a, NRPB8a, NRPD8a, NRPE8a) (At1g54250)

MASNIILFEDIFVVDQLDPDGKKFDKVTRVQATSHNLEMFMHLDVNTEVYPLAV
GDKFTLALAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTP
KAELYVSFGGLLMLLKGDPAHISHFELDQRLFLLMRKL

54/146=37% coverage
41/146=28% unique

NRPC8b (NRPA8b, NRPB8b, NRPD8b, NRPE8b) (At3g59600)

MASNIIMFEDIFVVDKLDPDGKKFDKVTRVEARSHNLEMFMHLDVNTEVYPLAV
GDKFTLAMAPTLNLDGTPDTGYFTPGAKKTLADKYEYIMHGKLYKISERDGKTP
KAELYVSFGGLLMLLQGDPAHISHFELDQRLFLLMRKL

48/146=33% coverage
35/146=24% unique

NRPC9a (At1g01210)

MEFCPTCGNLLRYEGGGNSRFFCSTCPYVAYIQRQVEIKKKQLLVKKSIEAVVTK
DDIPTAAETEAPCPRCGHDKAYFKSMQIRSADEPESRFYRCLKCEFTWREE

12/106=11% coverage
0/106=0% unique

NRPC9b (At4g07950)

MEFCPTCGNLLRYEGGGSSRFFCSTCPYVANIERRVEIKKKQLLVKKSIEPVVTK
DDIPTAAETEAPCPRCGHDKAYFKSMQIRSADEPESRFYRCLKCEFTWREE

12/106=11% coverage
0/106=0% unique

NRPC10 (NRPA10, NRPB10, NRPD10, NRPE10) (At1g11475)

MIIPVRCFTCGKVIGNKWDQYLDLLQLDYTEGDALDALQLVRYCCRRMLMTHV
DLIEKLLNYNTLEKSDNS

62/71=87% coverage
51/71=72% unique

NRPB10-like (At1g61700)

MIVPVRCFTCGKVIGNKWDTYLELLQADYAEGDALDALGLVRYCCRRMLMTH
VDLIEKLLNYNTMEKSDPN

11/71=15% coverage
0/71=0% unique

NRPC11 (NRPA11) (At2g29540)

MLVYIHVMLLEILIVFNECHIILSLRLMELICWCVDNDDYVNNQYCFQFCSPRVT
VAAYTIPHPSLEQVNIRVQTTGDPAREVFKDACQELMQMNRHVRSVFDKAVAE
YKDEQKRKEEAEEEELKRQRDLFGSMDIENN

70/139=50% coverage
70/139=50% unique

NRPC12 (NRPA12, NRPB12, NRPD12, NRPE12) (At5g41010)

MDPAPEPVTYVCGDCGQENTLKSGDVIQCRECGYRILYKKRTRRVVQYEAR

8/51=16% coverage
8/51=16% unique

NRPC82 (At3g49000)

MSMSEFGIVYAIHIITVQFGSVVSKVCECLLRKGPLSSRDISRLAESDINHNKVKDI
LYLLIQHNCVQAFSIEPPDGSESKAIVQYIVLFNNILHRVRYNKFSRIVNEELDSKC
GAVLDGLLSNGRLTLGQFIERDRDSGNTIGSEAIRDSLQKLVAARFVERIPSPEPV
LGNKEKEPAKRRGAKAAKILKEPETLEEQVVEAATPVDAIRFPLIFEEDSNSSLAD
DNSNITEGKRKQRDVDTSDSSSGVIWRPNFEEFIHRLRHKACVEIVKERRDEGCAI
VLRAMLEVGRSQEKKVKTDNSAPMSVGSIYEEVIKTEAGRTMLQERVEACLDQL
SATSSYLPAFVNEVNDSYIVDYKSIISVAQKDEIEAVVMRRYGKEAFRMFRYLSQ
EGRFVETDKIADAALTEKKDTPQFLLKMWKDGYLHMQKLAVTAPGMYTPFLL
WKVNKLIVTTQMLDEMYHASLNLSLRLAHELEAEKELLLLPLDKLEGPLKERLK

KVKAKRLLLSSTMFKLDDAIMLFHDF

140/523=27% coverage
140/523=27% unique

NRPC53 (At4g25180)

MDSGEQKSKRRFQPNPPRPSRRLPIAPTSNTEAEEDEENIKASRQFDRRIVGRRPK
TETKASSPEVAFQPSLSPLAIRSFGVPKEDDKPNSDVNPSSPASILPAVSSVTAAQE
DGEEVHNFVTRTGDDYVEPWDYRNSYYPTVLPLRKPNSGDIELLDQEEFGEVAK
NRDYDENTINSAEELGLTSVQHSKKQMFIFKIPDCLPVVKQTTGATTKRSVREYS
SGISNPFEGLPEGFMGKMLVYKSGAVKLKVGDALFDVSPGPGTKIPNDVVAIDIK
GRNCSRIGSSAKFVTVTPDVESLLNPASDMETQK

176/311=57% coverage
176/311=57% unique

NRPC53 (At5g09380)

MEQKPPVRKMKFAPKAPPKRVPKPEVKPEVVEDNSNSAQASELLRRVNERSLRK
PKADKKVPASQVAWLGGVVNSTRSNKYLNRSNGAYGSTSTQEIEYKEPWDYYS
YYPITLPMRRPYAGDPEVLDVEEFMQAGGHHEDSLNTAANLGLMEDSGEQKMF
FMRLPSVPLASTPTENLETRPNIKGPVEKKTVDLKALPEGYMGKLLVYKSGAVK
MKLGEVLYDVSPGLKSEFAQDVMVVNTEQKNCCLVGDVYKHAVLTPDIDSILK
DIENI

88/272=32% coverage
88/272=32% unique

NRPC37 (At5g49530)

MDFDDDDKPKEVTKTRRFAPGRAGKSKPKPKPEPTADKPVQPPPQSQTESVSKT
EHDVDAKFVGTKVETEVCNGSVKMEIDSKVDKEPEIMETELMEEDQQLPLQEEK
EEEEEEDDVVVREIDVFFKPSIDANTQLYVLQYPLRPSWRPYEMDERCEEVRVNP
STSQVEIDLSMDVHSKNYDSNFGLNMTKQTLKTTWKQPPTLDYAVGVLSGDKL
HLNPVHAVAQLRPSMQSLSSDKKKKQEESTEESVGTSKKQNKGVQQASTDQKPI
NEETWVSLKYHGLQSEYCSRYLNGMMANGNSSIDFNMSPGTYINELCRGGSSRN
SESKETLKRVLLSLPLKERVQKLLCEGSPLIRYSVLKHYAPEFSDEDFLGALQEYG
RLVQGLWTPKTRLLKLDGPVEAARDYVLSLFSQNTTIKYSEVEATGDKMKPLME
RMLTEFAKERHVLKDWKFKEPTDVSFIKSYPEIVKEQDIFWTDKRENLKSRITAQ
GGKSRADKRRNVVGTSSSVTVKPEVPTTLSDKGGSSKNTIHRVVTQEMPEELKK
ALPKALKKVFQTHKVCRYETICQGLRDLAVSTSNNPKADSGMAVNVALAVDAY
QGELEDVINGVATNIHGSYVSISSPDHPEYDSLREVVISLLTGSPPGTKLMKAEVF
AAGRTKLEREITNNEYIKVMHEICETNSSGWVLQKAR

260/689=38% coverage

260/689=38% unique

NRPC34 (At5g23710)

MSKRKRPDPKSSGGDLNEPHEKKLLDLIRINQGRGATMFELKREKTIPATIVTRLI
ASLRKKNLIKEVANMNNKGVKHYLAMEFEPCSELTGGEWYTDGALDLSKIEDL
KAKCVMILERHRQRVVTLEVLCGYFVKEEKLSVDQTKEILKNLILDNLIMEVKSN
GMNEFASTRIGEVCYRLTGKKFGNGEPRAGAFASIPCGVCPHIAICSPDGVISPTT
CVYFQKWLDF

57/230=25% coverage
57/230=25% unique

NRPC31 (At4g01590)

MSWKGARGKPKGYGGDYGKPEPFVIFPEITLPDPKSISTDSQLVQSYFTFNKFWR
NSPYHLGDGGVSKKEKESLNIERYSDSLKPKMKSNKNGSFFDFLVLRPDNFPKEL
LGDTRREQRPVKRAKWSQEADLQKLDVFEKLEAKFKVEGKEEKEEGEDDEEVV
ESEGEESDNGDYDQNQDFDDDDDDYNNEDDGFEEVY

107/199=54% coverage
66/199=33% unique

NRPC31-like (At4g35680)

MEGLDSVFLSNQTTHQPVRSASLPSRIHPLSVKLRTALSRLSIWRRSSSSISVSASF
GYETVLVGLVNLTELYGCVHELLESPYVKHTLLHHQEGKLLDESLDGSVLLLDV
YEGTREVIVAMREHVTNLKSALRRKGSLEKEAKAYFNLRKKAKKEISKQINALK
KMETRDISTNTDQDSAIASTSVLRETIQITVSMFRHLLLFLSTIPPPPSPAIFKTTIGL
LSIPFVSPSLSDKSLILIKEMKSLDDVFLGSILDSRKTLFEVETMENEKMRRDVVE
DGFRDLEAELDSVSKCLVKNRVLFLNILNLVCEISQHHSSREFGSMSWKGGRGKP
KGYGGDYGKPEPFVIFPEITLPDPKSISTDSQLVVVQSYFTFNKFWMNSPYHLCD
GGVSKKEKASLDIERPDNFSKELVGDTRREQRPVKRAKWSQEADLQKLDVFEKL
ESKFKTQGNEEKEDGEDDEQVVESEGEESDNGDYDQNQDFDDDEDDYNHEEDG
GFEEVY

41/503=8% coverage
0/503=0% unique

The following non-pol III subunits had zero coverage:

NRPA1 (At3g57660)
NRPB1 (At4g35800)

NRPE1 (At2g40030)
NRPD1 (At1g63020)

NRPA2 (At1g29940)
NRPB2 (At4g21710)
NRPD2a (At3g23780)
NRPD2b (At3g18090)

NRPA49 (At3g13940)

NRPA34 (At5g64680)

NRPB3a (At2g15430)
NRPB3b (At2g15400)
NRPA3 (At1g60850)

NRPD4 (At4g15950)
NRPB4 (At5g09920)

NRPB5-like (At5g57980)
NRPE5 (At3g57080)
NRPE5-like (At2g41340)
NRPE5-like (At3g54490)

NRPE7 (At4g14660)
NRPD7 (At3g22900)
NRPB7-like (At4g14520)
NRPB7 (At5g59180)
NRPA7 (At1g75670)

NRPB9a (At3g16980)
NRPB9b (At4g16265)
NRPA9 (At3g25940)

NRPB11 (At3g52090)

NRPB12-like (At1g53690)

Supplemental Figure 11.
ClustalW-aligned Rpa34/PAF49 sequences from plants, fungi and animals.

```
O_sativa_j_NP_001042641_Os01g0 -------------MR-------------------------
O_sativa_j_NP_001061828_Os08g0 MAAAAAAAEEERRMRGVYPSVWIQSMDSVCGSVILLSPHP
Z_mays_ACG46868                ------MVKAKARVR-------------------------
V_vinifera_CAO68707            -------MDAVSEAD-------------------------
A_thaliana_At5g64680_NP_851270 -----------MDFD-------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 --MEEPQAGDAARFSCPPN---------------------
P_patens_XP_001757296          -------MKEAKMLDGYSG---------------------
S_cerevisiae_NP_012387         --MSKLSKDYVSDSDSDDEVISN-----------------

O_sativa_j_NP_001042641_Os01g0 -------SYEPGPAFEENSEEA-----MLDISQTESTELW
O_sativa_j_NP_001061828_Os08g0 PQTKPHFFYEPGPAFEENSEEA-----MLDISQTESTELW
Z_mays_ACG46868                --------YEPGPAFEEVKEEA-----MLDISPTDSTEFW
V_vinifera_CAO68707            --------YKPPPEFIEDTKDS-----LVDLSMTDSKELW
A_thaliana_At5g64680_NP_851270 --------FKVSGDFIVSGAEQ-----LDDTDLTRSDEFW
H_sapiens_PAF49_CAST_CD3e_NP_0 --------FTAKPPASESPRFS-----LEALTGPDT-ELW
P_patens_XP_001757296          --------YSPGDGFRQVELLAP----LVEEAASSNTELW
S_cerevisiae_NP_012387         -------EFSIPDGFKKCKHLKNFPLNGDNKKKAKQQQVW

O_sativa_j_NP_001042641_Os01g0 LIQWPLNQL-------------------------------
O_sativa_j_NP_001061828_Os08g0 LIQWPLNQL-------------------------------
Z_mays_ACG46868                LIQWPKDQI-------------------------------
V_vinifera_CAO68707            LIQWPVN---------------------------------
A_thaliana_At5g64680_NP_851270 LIQAPLG---------------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 LIQAPADFAPECFNGRHVPLSGSQIVKGKLAGKRHRYRVL
P_patens_XP_001757296          LVQLPVN---------------------------------
S_cerevisiae_NP_012387         LIKFPSN---------------------------------

O_sativa_j_NP_001042641_Os01g0 ------------DASDFHGQELTLKLHRDGKLSSLESSS-
O_sativa_j_NP_001061828_Os08g0 ------------DASDFHGQEVTLKLHRDGKLSSLDSSS-
Z_mays_ACG46868                ------------DVLDFHGKEVSLKLHSDGNLGNLENSS-
V_vinifera_CAO68707            ------------QHPDFDGQELSLKLHQDGQLGKFEGSS-
A_thaliana_At5g64680_NP_851270 ------------QFPEIEENTLKIEPDKDGLFGEFKDSN-
H_sapiens_PAF49_CAST_CD3e_NP_0 SSCPQAGEATLLAPSTEAGGGLTCASAPQGTLRILEGPQQ
P_patens_XP_001757296          -----------EVSPTVSVSLLAFVSWTCGFFSLWEGYRV
S_cerevisiae_NP_012387         -----------VDISKLKSLPVDFESSTTMTIDKHDYKI-

O_sativa_j_NP_001042641_Os01g0 --------------------GKSYDLVSFAAQQPDATVFL
O_sativa_j_NP_001061828_Os08g0 --------------------GKSYDLVSFAAQQPDATVFL
Z_mays_ACG46868                --------------------GKSYEIASFAAQKPDATVFL
V_vinifera_CAO68707            --------------------GKLYNVVSFASQDPDATVFI
A_thaliana_At5g64680_NP_851270 --------------------GAKYDLASFHSQDAGAELII
H_sapiens_PAF49_CAST_CD3e_NP_0 SLSGSPLQPIPASPPPQIPPGLRPRFCAFGGNPPVTGPRS
P_patens_XP_001757296          TDYG------------LFVYARAVILSMCPKDLMDKQWVI
S_cerevisiae_NP_012387         --------------------MDDTDIESSLTQDNLSNMTL

O_sativa_j_NP_001042641_Os01g0 PSGPEAKAVGKIARRVSLVRYPDPEELEK-----------
O_sativa_j_NP_001061828_Os08g0 PSGPEAKAVGKIARRVSLVRYPDPEEPEK-----------
Z_mays_ACG46868                PSGSETKPVGKISRRVSLVRYPKPEEFTK-----------
V_vinifera_CAO68707            SSPSESKIVGKISRRVSLVHYPEPDELEN-----------
A_thaliana_At5g64680_NP_851270 PS-EESMIVGKITRRVALVRYPEPNELLQ-----------
H_sapiens_PAF49_CAST_CD3e_NP_0 ALAPNLLTSGKKKKEMQVTEAPVTQEAVNGHGALEVDMAL
P_patens_XP_001757296          QPPDSDGKIGHFYSIRGDICNVVTENVDRKK---------
S_cerevisiae_NP_012387         LVPSESKESLKIASTAKDNAPLQFDKVFS-----------
```

```
O_sativa_j_NP_001042641_Os01g0 ---------------------------------PGLGSLT
O_sativa_j_NP_001061828_Os08g0 ---------------------------------PGLGSLT
Z_mays_ACG46868                ---------------------------------PSFGSLT
V_vinifera_CAO68707            ---------------------------------QSANNLR
A_thaliana_At5g64680_NP_851270 ---------------------------------KMKARTQ
H_sapiens_PAF49_CAST_CD3e_NP_0 GSPEMDVRKKKKKKNQQLKEPEAAGPVGTEPTVETLEPLG
P_patens_XP_001757296          ----------------------------LYAIMPGSSQHS
S_cerevisiae_NP_012387         --------------------------------VSETAKIP

O_sativa_j_NP_001042641_Os01g0 PSSKKSAGSSKKTRSRFTSGSKNRSSQGSAQSLGQQSAEP
O_sativa_j_NP_001061828_Os08g0 PSSKKSAGSSKKTRSRFSSGSKNRSSQGSALSLGQQSAEP
Z_mays_ACG46868                PSIKKSAGS-KKTMSRFTGASKNPSSQGSALSLGQRSAEP
V_vinifera_CAO68707            KMYQRSGGSSLTHSSHHYSTPSHSTKLRNPQSVSGRSAS-
A_thaliana_At5g64680_NP_851270 QKLVGSVTNSSKKSSNLTQSSRHKSGTRSSREKSMFSGFT
H_sapiens_PAF49_CAST_CD3e_NP_0 VLFPSTTKKRKKPKGKETFEPEDKTVKQEQINTEPLEDTV
P_patens_XP_001757296          VRRITSKVCFRRQLEIQKETSLGGRSIRSGITSGEKSRGD
S_cerevisiae_NP_012387         AIDYSKVRVPRKDVPKVEGLKLEHFATGYDAEDFHVAEEV

O_sativa_j_NP_001042641_Os01g0 AHKHNQKRKDESSLGHSNVSGKSAE---------------
O_sativa_j_NP_001061828_Os08g0 AHKHNQKRKDESSLGHSNVSGKSAE---------------
Z_mays_ACG46868                TPKHKGKRKDESSLGHSNVSGKATE---------------
V_vinifera_CAO68707            THSSRHKRRHADKPATS-----INQ---------------
A_thaliana_At5g64680_NP_851270 ETPKSPKRKN------------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 LSPTKKRKRQKGTEGMEPEEGVTVESQPQVKVEPLEEAIP
P_patens_XP_001757296          STRSVQKRALTDEKQTSGVTEGSVGNSEVKK---------
S_cerevisiae_NP_012387         KENKKEPKKRSHHD--------------------------

O_sativa_j_NP_001042641_Os01g0 ----------------------------------------
O_sativa_j_NP_001061828_Os08g0 ----------------------------------------
Z_mays_ACG46868                ----------------------------------------
V_vinifera_CAO68707            ----------------------------------------
A_thaliana_At5g64680_NP_851270 ----------------------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 LPPTKKRKKEKGQMAMMEPGTEAMEPVEPEMKPLESPGGT
P_patens_XP_001757296          ----------------------------------------
S_cerevisiae_NP_012387         ----------------------------------------

O_sativa_j_NP_001042641_Os01g0 -GSQVRGGDSGTTSEVPQTPVEKSKKKNKKVRIAE-----
O_sativa_j_NP_001061828_Os08g0 -GSQVRGGDSGTTLEVPQTPGEKSKKKNKKVRIAE-----
Z_mays_ACG46868                -GSEARGAGSNTASEMPQSPPEKSKKRRKKDKIVE-----
V_vinifera_CAO68707            -LTQDSGRGHSTVTSSGSLGLSHQGKSTKKVKLEG-----
A_thaliana_At5g64680_NP_851270 --SESSSGKHRSSTSTVSGSSERSAKSKKKVKKEE-----
H_sapiens_PAF49_CAST_CD3e_NP_0 MAPQQPEGAKPQAQAALAAPKKKTKKEKQQDATVEPETEV
P_patens_XP_001757296          -ESKKKKKKTDLSATEGSIGNSEIEKDSKKKKKKEKRSKS
S_cerevisiae_NP_012387         --DEEESSEKKKKKKEKREKREKKDKKDKKKKHRD-----

O_sativa_j_NP_001042641_Os01g0 ----------------------------------------
O_sativa_j_NP_001061828_Os08g0 ----------------------------------------
Z_mays_ACG46868                ----------------------------------------
V_vinifera_CAO68707            ----------------------------------------
A_thaliana_At5g64680_NP_851270 ----------------------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 VGPELPDDLEPQAAPTSTKKKKKKKERGHTVTEPIQPLEP
P_patens_XP_001757296          S---------------------------------------
S_cerevisiae_NP_012387         ----------------------------------------
```

```
O_sativa_j_NP_001042641_Os01g0 -----------------------------------------
O_sativa_j_NP_001061828_Os08g0 -----------------------------------------
Z_mays_ACG46868                -----------------------------------------
V_vinifera_CAO68707            -----------------------------------------
A_thaliana_At5g64680_NP_851270 -----------------------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 ELPGEGQPEARATPGSTKKRKKQSQESRMPETVPQEEMPG
P_patens_XP_001757296          -----------------------------------------
S_cerevisiae_NP_012387         -----------------------------------------

O_sativa_j_NP_001042641_Os01g0 ---------------------------
O_sativa_j_NP_001061828_Os08g0 ---------------------------
Z_mays_ACG46868                ---------------------------
V_vinifera_CAO68707            ---------------------------
A_thaliana_At5g64680_NP_851270 ---------------------------
H_sapiens_PAF49_CAST_CD3e_NP_0 PPLNSESGEEAPTGRDKKRKQQQQQPV
P_patens_XP_001757296          ---------------------------
S_cerevisiae_NP_012387         ---------------------------
```

Supplemental References

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J *16*, 735-743.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2009). Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. Mol Cell *33*, 192-203.

CHAPTER FIVE

CONCLUSIONS AND FUTURE DIRECTIONS

Prologue

Solving the subunit compositions of *Arabidopsis* RNA polymerases I, II, III, IV and V has opened up a new field of research in plants, namely the functional characterization of RNA polymerase subunit families. In the case of Pol IV and Pol V, the roles of their subunits in transcription and silencing will be important to determine. In addition, transgenic lines and antibodies will provide valuable tools to purify Pol IV and Pol V such that their activities, locations in the nucleus and additional interacting partners can be deciphered. Finally, the duplications of RNA polymerase subunits in *Arabidopsis*, combined with the knowledge of the subunit compositions of all five DNA-dependent RNA polymerases, provides a foundation for understanding the evolution of RNA polymerase subunits in plants. The following outlook presents ideas and experiments that will give greater insight into our understanding of multi-subunit RNA polymerase functions.

i.

ELUCIDATING THE ROLE OF NON-CATALYTIC SUBUNITS IN MULTI-SUBUNIT DNA-DEPENDENT RNA POLYMERASE FUNCTION IN PLANTS

Introduction

Chapter 2 describes the initial studies on Pol IV that revealed its role in RNA-directed DNA methylation and siRNA production (Onodera et al. 2005). At the same time, several other groups discovered Pol IV or Pol V, although at the time it was unclear that Pol IV and Pol V constituted two functionally distinct polymerases with unique

activities and subunit compositions (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). In the four years since these initial reports, only one screen has identified a non-catalytic subunit in Pol IV and Pol V, namely NRPD4/NRPE4, as being partially required for RNA-directed DNA methylation and silencing of a reporter gene (He et al. 2009). This raises the question as to whether any of the other small, non-catalytic subunits are involved in gene silencing. Based on what is known about the contributions of different subunits in the yeast RNA polymerases, however, one would expect that null mutations in nearly any subunit might indirectly impair the activity of the enzyme and lead to a loss of silencing phenotype. If this is true, why have screens not uncovered more mutations in the non-catalytic subunits of Pol IV and Pol V? Several reasons may account for this observation. First, functional redundancy may occur with some of the subunits for which there are multiple genes. Second, some of the subunits are encoded by very short genes with less than 500 bp of coding sequence, making them smaller targets for mutagens compared to the largest and second largest subunits, which have multiple kilobases of coding sequence. Third, it is possible that some of these subunits, while important for the function of Pol IV and Pol V, are not absolutely essential and therefore when mutated show subtle phenotypes that would be missed in screens, depending on the sensitivity of their reporters and criteria for scoring a mutant phenotype. Finally, if any non-catalytic subunits of Pol IV and Pol V are shared with Pol I, II or III, then loss of function mutations in these subunits would not be recovered in genetic screens if they result in lethality, as previously demonstrated for several non-catalytic Pol I, II and III subunits in *Arabidopsis* (Chapter 3, Appendix C) (Onodera et al. 2008).

Testing for functional redundancy between subunits

Of the Pol IV and Pol V non-catalytic subunits identified in Chapter 3, at least eight are part of multi-gene families, namely the NRPE3, NRPE5, NRPE6, NRPE7, NRPE8, NRPE9, NRPE10 and NRPE12 families. Of these eight, at least five showed some evidence of redundancy. For example, NRPE3a and NRPE3b were identified by mass spectrometry and confirmed by co-IP to be *bona fide* subunits of Pol V. The same is true for NRPE6a and NRPE6b, NRPE8a and NRPE8b, NRPE9a and NRPE9b, and possibly NRPE7 and NRPD7, as some unique peptides corresponding to NRPE7 were identified in affinity purified Pol IV (Chapter 3) (Ream et al. 2009). At present, the evidence that NRPE7 or NRPD7 associate with Pol IV or Pol V is based only on peptides identified by mass spectrometry of the respective complexes. Therefore, co-IP experiments are needed to confirm the results (elaborated in the section below on NRPD7 and NRPE7 function).

NRPE6 and NRPE8 families

An interesting future direction will be to test for functional redundancies among paralogous subunit family members detected in Pol IV and Pol V, such as NRPE6a and NRPE6b or NRPE8a and NRPE8b. One approach to testing redundancy, therefore, is to obtain single and double mutants in these duplicated genes and to test them for their Pol IV- or Pol V-dependent RNA silencing phenotypes. However, a caveat to this approach is that each member of the NRPE6 and NRPE8 subunit families are shared by all five polymerases. If redundancy does exist among NRPE6a and NRPE6b subunits or NRPE8a and NRPE8b subunits, the double mutants are expected to be lethal because in

yeast, *rpb6* and *rpb8* mutants are both lethal (Woychik et al. 1990). In *Arabidopsis*, T-DNA insertion lines within both *NRPE6* genes are available and could be crossed to assay if the double mutant is lethal (Table 1). If so, this would confirm that the two genes are redundant in the context of Pol I, II or III. For *NRPB8* genes, I have identified a homozygous T-DNA insertion within an exon of the *NRPB8a/NRPE8a* gene (Table 1). There are several T-DNA insertions in either the 5'UTR or gene body for *NRPB8a*, but these mutants remain to be confirmed (Table 1). Crossing these lines to generate double mutants would assay for their redundancy, which would likely readout as a lethal phenotype due to their associations with the three essential RNA polymerases I, II and III (Appendix C)(Onodera et al. 2008).

Table 1. T-DNA insertion lines available for RNA polymerase subunits.

Protein	Gene	insertion	T-DNA location	allele	confirmed?
NRPB3a	At2g15430	SALK_008220	exon		
		SALK_002682	3' UTR		
NRPE3b	At2g15400	SALK_099705	exon		
		SALK_065619	exon		
NRPA3	At1g60850	SALK_088247	exon	*nrpa3-1*	lethal
NRPC3	At1g60620	SALK_132788	intron	*nrpc3-1*	lethal
		SALK_071883	intron		
		WsDsLox_343H05	exon		
NRPB4	At5g09920	SALK_122761	exon	*nrpb4-1*	yes
		SALK_012220	intron	*nrpb4-2*	yes
NRPD4/NRPE4	At4g15950	WsDsLox_476D09	exon	*nrpd4-2*	yes
		SAIL_1156B01	intron	*nrpd4-3*	yes
NRPC4	At5g62950				
NRPC4-like	At3g28956	FLAG_080G04	exon		

Table 1 (cont'd)

Protein	Gene	insertion	T-DNA location	allele	confirmed?
NRPB5	At3g22320	SAIL_786E02	5'UTR	*nrpb5-1*	lethal
		SALK_065080	5'UTR		yes
NRPE5	At5g57080	GABI_237A08	intron	*nrpe5-1*	yes
		FLAG_158A03	intron	*nrpe5-2*	yes
NRPE5-like	At2g41340	SALK_134107	intron		yes
		SALK_041506	intron		yes
NRPE5-like	At3g54490	SALK_135511	5'UTR		yes
NRPB5-like	At5g57980	SAIL_783G03	5'UTR		
NRPB6a	At5g51940	SALK_115863	5' UTR		
		SALK_013548	5' UTR		yes
		SALK_017716	5' UTR		
NRPB6b	At2g04630	SALK_064868	exon	*nrpb6b-1*	yes
		SALK_021069	exon		
NRPA7	At1g75670	SALK_023013	intron		
		FLAG_228D10	intron		
NRPB7	At5g59180	SAIL_592G11	5' UTR		
NRPC7	At1g06790	SALK_097958	promoter		yes
NRPD7	At3g22900	SALK_109173	promoter		yes
NRPE7	At4g14660	CSHL_13987	exon		yes
		GABI_808H11	exon		
NRPD7-like	At4g14520	FLAG_489D08	exon		
NRPB8a	At1g54250	SALK_151800	exon		yes
NRPB8b	At3g59600	SALK_014995	3' UTR		yes
		SAIL_237E10	promoter		yes
NRPB9a	At3g16980	SALK_032670	intron		yes
NRPB9b	At4g16265	SALK_031043	intron		yes
NRPA9	At3g25940	SALK_062311	3' UTR		yes
NRPC9a	At4g07950	SAIL_823B06	5' UTR		
NRPC9b	At1g01210	SALK_146765	5' UTR		yes
NRPB10	At1g11475	SALK_146725	5' UTR		yes
		SALK_114301	3' UTR		yes
		SALK_026610	5' UTR		yes
		WsDsLox_443C12	promoter		yes
NRPB10-like	At1g61700	SAIL_527F07	intron		
NRPB11	At3g52090	SALK_100563	intron	*nrpb11-1*	lethal
NRPAC19	At2g29540	SALK_007414	exon		
		WsDsLox_419G02	intron	*nrpac19-1*	lethal
NRPB12	At5g41010	SALK_049327	intron	*nrpb12-1*	lethal
NRPB12-like	At1g53690	SALK_090950	5' UTR		yes
		FLAG_069E08	intron		

Table 1 (cont'd)

Protein	Gene	insertion	T-DNA location	allele	confirmed?
NRPA13	At3g13940	SALK_145539	exon		
		SALK_054381	exon		
NRPA14	At5g64680				
NRPC13	At3g49000	SALK_027102	exon		
		SAIL_317F06	exon		
		SAIL_519A11	intron		
		FLAG_570D05	intron		
		SALK_008989	intron		
NRPC14a	At4g25180	SALK_125873	intron		
		SALK_078747	3' exon		
		SALK_025232	intron		
		SALK_073697	intron		
		SAIL_659_A08	exon		
NRPC14b	At5g09380	SALK_113934	intron		
		SALK_002157	intron		
		SALK_098495	intron		
		GABI_358F12	intron		
NRPC15	At5g49530	SAIL_1246_D09	intron		
		SAIL_35_A02	exon		
		FLAG_322F11	intron		
NRPC16	At5g23710	GABI_683A01	exon		
		SALK_087300	5'UTR		
NRPC17	At4g01590	SAIL_106_D06	intron		
NRPC17-like	At4g35680	SALK_061195	exon		
		SAIL_669G10	exon		
		SAIL_841D02	exon		
		SALK_139015	exon		

The NRPE9 family

NRPE9a and NRPE9b subunits are shared by Pol II, Pol IV and Pol V (Chapters 3-4)(Ream et al. 2009). In yeast, *rpb9* mutants are viable, but temperature sensitive (Woychik et al. 1991). Therefore, it is possible that a double *nrpe9a nrpe9b* mutant could be recovered from *Arabidopsis*, although I would expect the plants would not tolerate this mutation very well. T-DNA insertions in both *NRPE9* genes are available and I have genotyped homozygous mutants (Table 1). In preliminary tests, the single

mutants display wild-type DNA methylation and siRNA levels, supporting the co-immunoprecipitation and mass spectrometry data suggesting that the alternative *NRPE9* genes are redundant (Figure 1)(Chapter 3). These single mutants have been crossed and progeny will genotyped to identify double mutants. If double mutants are viable, it will be interesting to test if they have any reduction in DNA methylation or siRNAs levels at Pol IV- or Pol V-dependent loci that are coincident with loss of silencing. This would support a critical function for the NRPD9/NRPE9 subunit in Pol IV or Pol V function.

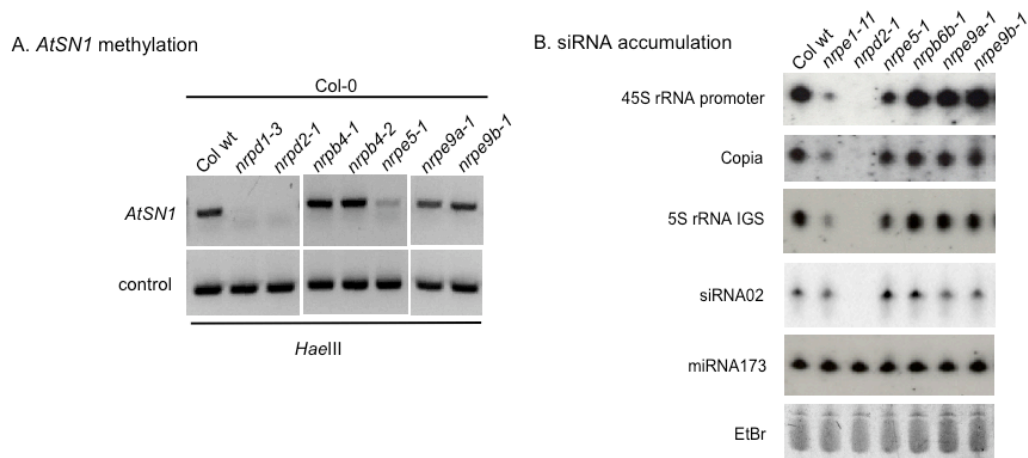


Figure 1. DNA methylation and siRNA accumulation in *nrpe9* mutants is identical to wild-type plants. A. *AtSN1* DNA methylation in the genotypes shown was assayed by digesting genomic DNA with *Hae*III and PCR amplifying DNA with primers to *AtSN1* or a control locus lacking *Hae*III sites, as described previously (Ream et al. 2009). B. siRNAs were isolated from each genotype and run on denaturing polyacrylamide gels, transferred and blotted with radio-labeled probes to the given sequences as described previously (Zilberman et al. 2004; Ream et al. 2009).

The NRPE3 family

Approaches similar to those described above can be used for studying the roles of the NRPE3 subunits, particularly NRPE3b. Whereas NRPE3a appears to be the canonical NRPE3 shared by Pol IV, Pol V and Pol II, it is interesting that NRPE3b is also significantly associated with Pol V, but not with Pol IV or Pol II. These NRPE3 variants are nearly 90% identical, yet Pol II and Pol IV have presumably evolved a mechanism to distinguish between the two subunits during polymerase assembly, whereas Pol V has not. Although it is likely that both subunits are redundant in the context of Pol V based on co-IP and mass spectrometry data, it will still be worthwhile to test if an *nrpe3b* single mutant displays any Pol V mutant phenotypes. To this end, it is possible that this subunit contributes to Pol V silencing in a locus-dependent manner. In yeast, the Rpb3 subunit is proposed to be a docking site for Mediator (Davis et al. 2002). To this end, utilization of NRPE3b by Pol V may be regulated by selective binding of NRPE3b to locus-specific transcription activators. In addition, it will be interesting to test for redundancy between NRPE3a and NRPE3b in the context of Pol II. Because only NRPB3a/NRPE3a was found associated with Pol II, one might expect that loss-of-function mutations in this subunit are lethal. If NRPB3a/NRPE3a T-DNA mutants are viable, then it is possible that in the absence of NRPB3a, Pol II can make use of the NRPE3b subunit. This latter hypothesis can be tested by co-IP experiments in homozygous *nrpb3a* mutants that have been transformed with an epitope-tagged *NRPE3b* gene. Because *E. coli* α homologs of Pol I, II or III are lethal in yeast, it is likely that either *nrpe3a* mutants or the *nrpe3a nrpe3b* double mutant is lethal (Mann et al. 1987; Kolodziej and Young 1989; Dequard-Chablat et al. 1991; Woychik et al. 1993). If the latter case proves true, this would

suggest a conditional functional redundancy of NRPE3b with NRPE3a, which is only observed when NRPE3a is absent.

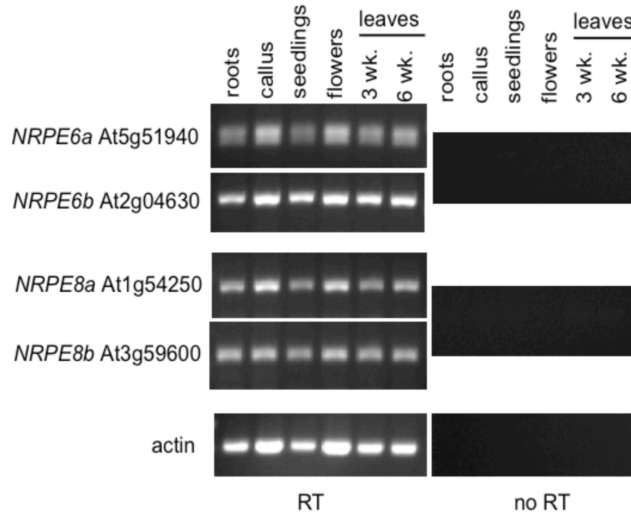


Figure 2. RT-PCR of *NRPE6* and *NRPE8* genes shows both variants are ubiquitously expressed in the tissues tested. RT-PCR-CAPS was carried out by PCR amplifying poly-T generated cDNA (Ream et al. 2009). Both *NRPE6* genes were amplified with the following primer set: F-gat gag cca gca gag cct gag and R-cga tca gtt cat caa ctc ccc. Cleavage with *Hha*I cuts At5g51940 at position 185 of a 352 bp product. Both *NRPE8* genes were amplified with the following primer set: F-gtt cga gga tat ctt cgt ggt cg and R-ctc tgc ttt tgg agt ttt gcc. Cleavage with *Hinc*II cuts At3g59600 at position 200 of a 313 bp product.

Isoforms of RNA polymerases?

For subunits resulting from gene duplications, as discussed above, it is likely that the two alternative variants are present in different polymerases, because all polymerase subunits have a stoichiometry of one in yeast (Cramer et al. 2001). Therefore, polymerases with unique combinations of alternative subunits may exist in a cell-type or

tissue-specific manner, or within polymerase enzymes localized in discrete nuclear compartments in the cell. Gene expression patterns assayed by microarray (Atgenexpress website: http://jsp.weigelworld.org/expviz/expviz.jsp) and by RT-PCR in different plant tissues indicates that both *NRPE6* and *NRPE8* paralogs are ubiquitously expressed, as expected for polymerase subunits shared by Pol I, II and III (Figure 2), suggesting that there is no spatial separation of expression of these genes on a macro scale in plants, although its always possible that individual cells within sampled tissues, or tissues not yet tested, show differential expression. Therefore, it may be informative to test the sub-cellular localization, and possible co-localization, of NRPE6a and NRPE6b proteins. This can be accomplished by performing immunolocalization of different epitope-tagged NPRE6a and NRPE6b subunits in the same genetic background. Likewise, a similar approach should immunolocalize both NRPE8 subunits relative to each other and both NRPE9 subunits relative to each other. Indeed, many of these transgenic lines have already been constructed and now need only to be crossed and analyzed for expression (Table 2). Making antibodies specific to each native subunit may be challenging given their high protein sequence identities. Nonetheless, I have made polyclonal antibodies to the NRPB3, NRPB6 and NRPB8 variants, but I have not tested their specificities in *Arabidopsis* (Table 2). If variant subunits never co-localize, then one could argue that there are different polymerase isoforms that incorporate these different subunits in each cell, and it would be interesting to understand the functional consequences of using alternative subunits. If the alternative subunits of a family do co-localize, this result will be harder to interpret, because it could still be possible, and quite likely, that they are incorporated separately into different polymerases but are active at the same locations.

However, this latter result might suggest that the variants are simply redundant and that there are no polymerase assembly mechanisms to distinguish between the two variants, which are over 88% identical in each of the NRPE3, NRPE6 and NRPE9 families.

Table 2. Transgenic lines and antibodies available for studying RNA polymerase subunits.

				Transgenic lines available		
		antibodies		genomic	35S:cDNA	
subunit	gene identifier	confirmed	not tested	FLAG	FLAG	HA
NRPE3a/NRPB3a	At2g15430		x			x
NRPE3b	At2g15400		x	x		x
NRPA3	At1g60850	x		JH		
NRPC3	At1g60620			JH		
NRPB4	At5g09920		x	x	x	
NRPC4	At5g62950					
NRPC4-like	At3g28956					
NRPB5	At3g22320	TG	x	x		
NRPE5	At3g57080	TG	x		x	
NRPB5-like	At5g57980			x		
NRPE5-like	At2g41340		x			x
NRPE5-like	At3g54490			x		
NRPB6a	At5g51940		x	x	x	
NRPB6b	At2g04630		x		x	
NRPB7	At5g59180		x		x	
NRPE7	At4g14660				x	
NRPD7	At3g22900				x	
NRPB7-like	At4g14520		x		x	
NRPA7	At1g75670				FP	
NRPC7	At1g06790	x			x	
NRPB8a	At1g54250		x		x	
NRPB8b	At3g59600		x	x		x
NRPB9a	At3g16980		x		x	
NRPB9b	At4g16265				x	
NRPA9	At3g25940				x	
NRPC9a	At4g07950				x	
NRPC9b	At1g01210		x		x	
NRPB10	At1g11475		x	x	x	
NRPB10-like	At1g61700		x	x	x	
NRPB11	At3g52090		x	x		
NRPA11/NRPC11	At2g29540		x	x		x
NRPB12	At5g41010		x	x		
NRPB12-like	At1g53690		x		x	
NRPA13	At3g13940				FP	
NRPA14	At5g64680					

Table 2 (cont'd)

subunit	gene identifier	antibodies: confirmed	antibodies: not tested	Transgenic lines available: genomic FLAG	Transgenic lines available: 35S:cDNA FLAG	Transgenic lines available: 35S:cDNA HA
NRPC1	At5g60040					
NRPC2	At5g45140			JH		
NRPB1	At4g35800					
NRPB2	At4g21710			JH		
NRPA1	At3g57660					
NRPA2	At1g29940			JH		
NRPD1	At1g63020	JH		JH		
NRPE1	At2g40030	JH		JH		
NRPD2	At3g23780	x		JH		
NRPC13	At3g49000					
NRPC14a	At4g25180					
NRPC14b	At5g09380					
NRPC15	At5g49530					
NRPC16	At5g23710					
NRPC17	At4g01590					
NRPC17-like	At4g35680					

Contributed by:

TG	Thomas Guilfoyle
JH	Jeremy Haag
FP	Fred Pontvianne

While the experiments above may simply reinforce the mass spectrometry data indicating that two different variants of NRPB3, NRPB6, NRPB8 and NRPB9 can integrate into the same class of polymerase(s), one question that is not addressed by the mass spectrometry data is whether there is a preference for particular variants of different subunit families to associate together in the same polymerase, forming different Pol IV or Pol V isoforms. To address this question, a brute force approach could be taken to affinity purify several of the different variants from different subunit families and identify the associated RNA polymerase subunits by mass spectrometry. Given the improvements in mass spectrometry analyses and the relatively quick purification of the polymerase complexes by using their epitopes, the limiting costs in the experiment are money spent

on reagents and time spent by researchers analyzing LC-MS/MS and looking for correlations between different subunit variants in a given complex. In the event that there are preferences for certain variants to associate together in a polymerase, a long-term goal would be to understand if there are functional consequences of these associations.

Roles of the NRPB5 family

An interesting subunit family is the NRPB5 family of *Arabidopsis*, which contains five members. This family can be divided into two classes—the NRPB5-like and NRPE5-like subfamilies (Devaux et al. 2007; Ream et al. 2009). The NRPB5-like class has two members, the canonical NRPB5/NRPD5 protein that associates with Pol I, II, III and IV, and At5g57980 (Chapter 3)(Saez-Vasquez and Pikaard 1997; Larkin et al. 1999; Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). The NRPE5-like class consists of NRPE5, At2g41340 and At3g54490 (Chapter 3)(Lahmy et al. 2009; Ream et al. 2009). NRPE5 is unique to Pol V and *nrpe5* mutants have defects in RNA-directed DNA methylation and siRNA production (Chapter 3) (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). The phenotypes of two different *nrpe5* mutants are similar and include reduced DNA methylation and siRNA production, although not as strongly as in an *nrpe1* mutant (Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). *nrpe5-2* (FLAG insertion line 158_A03) has been shown to be a null allele based on protein immunoblotting of wild-type vs. mutant extracts (Lahmy et al. 2009). It is possible that this subunit is not required for the full activity of Pol V, which might explain the intermediate phenotype. Alternatively, functional redundancy may exist with one of the other NRPE5-like subunits. The best candidate for redundancy with NRPE5 is

At2g41340, which is 70% identical to NRPE5. At2g41340 is expressed in a tissue-specific manner at much lower levels than NRPE5, arguing against redundancy in all cells, but perhaps could be redundant in certain tissues, such as leaf where both genes are expressed. An *nrpe5 At2g41340* double mutant has been constructed and is being assayed for functional redundancy. As an adjunct experiment to the double mutant analysis, the At2g41340 protein is being over-expressed in the *nrpe5* mutant and will be tested to see if it can associate with Pol V and rescue the *nrpe5* mutant phenotype. One caveat to this experimental approach is that over-expression of *At2g41340* from a constitutive promoter (35S) does not accurately reflect its endogenous tissue-specific gene expression pattern, and therefore interpretation of any positive result will require caution (Lahmy et al. 2009; Ream et al. 2009).

It is possible that redundancy exists between all of the NRPE5-like family subunits. A triple *nrpe5 At2g41340 At3g54490* mutant might tease apart these redundancies, but will be a challenge to obtain because *NRPE5* and *At3g54490* are closely linked on the same chromosome. Over-expression of *At3g54490* in the *nrpe5* mutant will address if this subunit is capable of rescuing *nrpe5* phenotypes and associating with Pol V. It is likely that the NRPB5-like *At5g57980* gene is not redundant with *NRPB5* in the context of Pol II because *nrpb5* mutants are lethal and *At5g57980* is not ubiquitously expressed, in contrast to NRPB5. However, tissue-specific forms of Pol II carrying At5g57980 cannot be ruled out (Onodera et al. 2008; Ream et al. 2009). Whether there is any redundancy of At5g57980 and NRPB5/NRPD5 in the context of Pol IV is an open question that will be difficult to assay using double mutants because *NRPB5/NRPD5* is an essential gene and only one T-DNA insertion in the 5'UTR of

At5g57980 is available (Chapter 3)(Ream et al. 2009). Therefore, localization of At5g57980 relative to Pol II and Pol IV may be insightful, as would RNA polymerase activity assays using immunoprecipitates of the tagged At5g57980 protein.

So-called "jaw" and "assembly" domains are present in the Rpb5 protein in yeast, and these domains are also recognizable in Rpb5 orthologs in higher eukaryotes. The assembly domain mediates the bulk of the contacts with the largest polymerase subunit, as shown by the yeast crystal structure, and shares a high degree of conservation among subunits in this domain across species (Cramer et al. 2001; Ream et al. 2009). There is a short patch of residues at the very C-terminus of NRPE5-like sequences that has significantly diverged from the NRPB5/Rpb5 orthologs in plants and animals (Chapter 3)(Ream et al. 2009). In addition, NRPE5-like sequences are unique to plants (Chapter 3)(Ream et al. 2009). One hypothesis is that these variant, C-terminal residues mediate the specific interaction with the Pol V largest subunit. In an effort to address this question, I performed a domain-swap experiment in which the C-terminal assembly domain of NRPE5 was replaced with the C-terminal assembly domain of NRPB5. This "hybrid" protein was transformed into wild-type and *nrpe5* mutants to see if it is still functional in the context of Pol V. Although the results of these experiments are still pending, the outcomes may address if this domain is sufficient for mediating the specific interaction of NRPE5 with Pol V.

Yeast Rpb5 interacts with a variety of transcriptional factors, mostly mediated by the N-terminal jaw domain that is exposed in the yeast Pol II crystal structure (Lin et al. 1997; Makino et al. 1999; Cramer et al. 2001; Soutourina et al. 2006). It is possible that Pol IV and Pol V use their distinct NRPD5/NRPB5 or NRPE5 subunits, respectively, to

interact with different activator or repressor proteins. Such differences might contribute to their functional diversification. Yeast-two-hybrid studies using NRPD5 or NRPE5 as bait proteins might be a promising approach to potentially identify weak or transient interactions that occur *in vivo* and that were missed in the mass spectrometry analyses of purified Pol IV and Pol V. In summary, the NRPB5 family of *Arabidopsis* is unique due to its numerous family members, unlike other eukaryotes, providing a potential model system for studying functional diversification among an RNA polymerase subunit family.

NRPD4/NRPE4, NRPD7 and NRPE7 functions

As stated previously, the NRPD7 and NRPE7 proteins were identified by mass spectrometry of proteins associated with Pol IV and Pol V affinity purified samples, but have not been confirmed to interact with Pol IV and Pol V by co-IP. Therefore, it would be useful to confirm these interactions using epitope-tagged lines of NRPD7 and NRPE7. To this end, I have constructed FLAG-tagged NRPD7 and NRPE7 lines and have obtained transformants. If NRPD7 and NRPE7 co-IP with both Pol IV and Pol V, this would suggest that these subunits are redundant. However, I expect the co-IP studies will confirm the mass spectrometry data, which suggest that NRPD7 and NRPE7 are specific to Pol IV and Pol V, respectively. It will be useful to obtain single mutants in *nrpd7* and *nrpe7* and to assay these mutants for loss of RNA-directed DNA methylation, siRNA production and gene silencing. Any such phenotypes should be rescued using the FLAG-tagged NRPD7 or NRPE7 constructs to confirm that the transgenes function correctly in the respective mutant background. Because these genes are specific to Pol IV or Pol V, and not shared by Pol I, II or III, their disruption should not be lethal. The *nrpd7* and

nrpe7 mutant phenotypes are expected to reflect the level of requirement for these subunits in Pol IV and V function, respectively. For example, if the NRPD7 subunit is absolutely required for Pol IV, then *nrpd7* mutants should completely eliminate heterochromatic siRNAs and decrease DNA methylation corresponding to these loci. Likewise, *nrpe7* mutants should be deficient for DNA methylation at Pol V-dependent loci. Currently, only one of two insertion alleles is available for *NRPE7* and no mutant is available for *NRPD7* (Figure 3A). However*, nrpd7* mutants might be generated by TILLING. I have started characterization of one of the *nrpe7* mutants, *nrpe7-1*. This mutant shows a defect in DNA methylation similar to the *nrpe5* mutants, namely levels of DNA methylation that are intermediate between wild-type and *nrpe1* mutants (Figure 3B-C). However, this mutant is in the Landsberg (Ler) ecotype, whereas most of the Pol V mutants studied are in the Columbia ecotype (Col-0) phenotype. Therefore, isolation of an *nrpe7* mutant in the Col-0 background would be better to directly compare DNA methylation, siRNA levels and retrotransposon reactivation. However, the intermediate DNA methylation phenotype suggests several possible scenarios. First, NRPE7 may be redundant with NRPD7. The co-IP data will certainly be informative towards this end. If true, then a double *nrpd7 nrpe7* mutant might display a more severe phenotype in regards to loss of DNA methylation and siRNA production compared to either single mutant. More likely, it is possible that the intermediate phenotype of the *nrpe7* mutant reflects a non-essential role for this subunit in Pol V-directed DNA methylation. Testing the accumulation of Pol V-dependent transcripts in an *nrpe7* mutant will provide insight into NRPE7's role in Pol V transcription.

It is interesting that Pol IV and Pol V may utilize distinct NRPB7-like subunits. In yeast, fewer pre-siRNA transcripts corresponding to the centromere repeats accumulate in *rpb7* mutants, apparently as a result of defective initiation specifically at the centromere repeat promoter (Djupedal et al. 2005). As a result, no siRNAs are produced in the *rpb7* mutants, presumably because the lower levels of Pol II transcripts in these mutants do not produce enough dsRNA to effectively feed into downstream dicing steps (Djupedal et al. 2005). Based on these results, it is possible that the NRPD7 and NRPE7 regulate recruitment of Pol IV and Pol V to specific promoters.

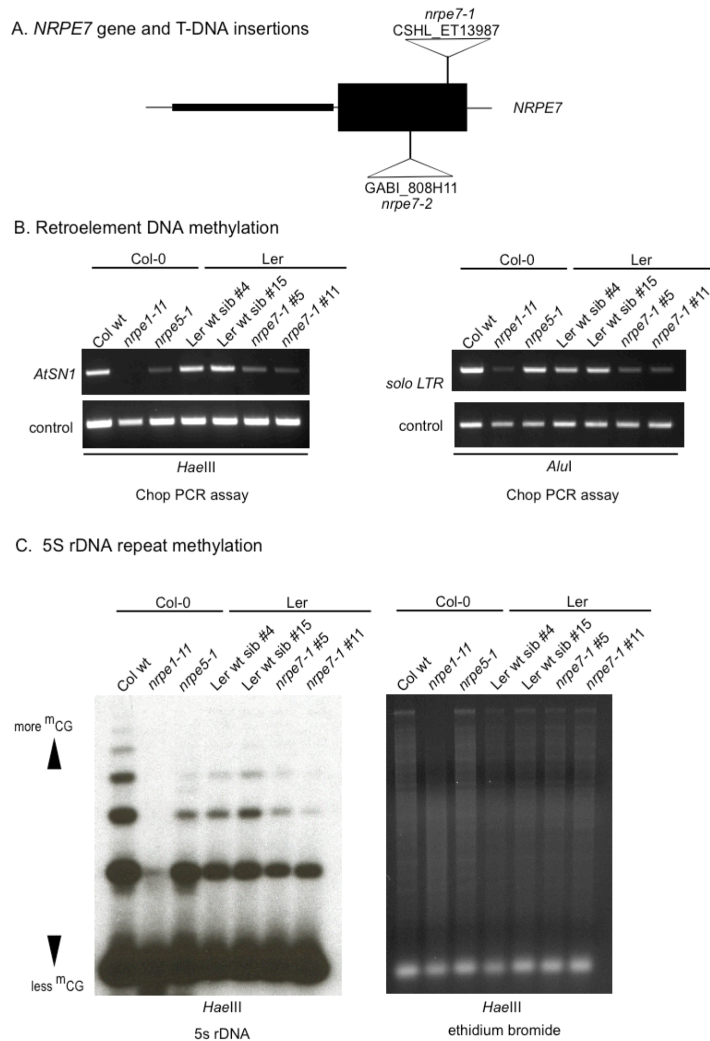


Figure 3. *nrpe7* mutants affect DNA methylation at Pol V-dependent loci. A. Gene structure of *NRPE7* showing the locations of two T-DNA insertions. B. DNA methylation analysis by Chop-PCR at *AtSN1* and the *solo LTR* loci, performed as in Figure 1A. C. Methylation-sensitive Southern blot analysis of the 5S rRNA genes in *nrpe7* mutants. *Hae*III digestion assays CHH (asymmetric) DNA methylation.

NRPE4/NRPD4 was identified in a screen for loss of RNA-directed DNA methylation of a reporter gene (He et al. 2009). Subsequent analyses of this mutant revealed that its phenotypes are not as severe as *nrpd1* or *nrpe1* mutants, similar to the case of *nrpe5* and *nrpe7* (Figure 3)(He et al. 2009). Because loss of Rpb4 leads to undetectable levels of Rpb7 in yeast Pol II (Edwards et al. 1991), it is difficult to distinguish between the direct and indirect effects of loss of Rpb4 in Pol II. Over-expression of Rpb7 partially rescues the temperature-sensitive phenotypes of an *rpb4* mutant and results in detectable levels of Rpb7 with Pol II, suggesting that Rpb4 is important for stabilizing the association of Rpb7 with Pol II (Sheffer et al. 1999). However, over-expressing Rpb7 in *rpb4* mutants is not sufficient to restore wild-type growth, arguing that Rpb4 functions independently of Rpb7 or is required for Rpb7 function outside the context of Pol II transcription (Sheffer et al. 1999). A similar scenario may play out with NRPE7 and NRPD4/NRPE4 or NRPD7 and NRPD4/NRPE4. The loss of NRPD4 in the *nrpd4* mutant may dissociate NRPD7 or NRPE7 from Pol IV or Pol V, respectively. This could be tested by co-IP of NRPD7 or NRPE7 in the *nrpd4* mutant. If this is true, then over-expression of NRPD7 or NRPE7 in the *nrpd4* mutant might provide sufficient levels of NRPD7 or NRPE7 to bind to Pol IV or Pol V and

restore their function. If the *nrpd4* phenotypes are rescued in these plants, then it can be argued that NRPD4 primarily functions to stabilize the NRPD7 or NRPE7 subunit and that the RNA-directed DNA methylation and silencing phenotypes in *nrpd4* mutants indirectly result from the loss of NRPD7 or NRPE7.

The similar methylation phenotypes of *nrpe4/nrpd4* and *nrpe7* mutants are consistent with the idea that these subunits require each other to function. This type of interplay has been demonstrated for the Rpb4 and Rpb7 subunits in yeast, where both are required for transcription initiation and mRNA processing in the nucleus, and mRNA decay in the cytoplasm (see Introduction) (Edwards et al. 1991; Orlicky et al. 2001; Goler-Baron et al. 2008; Runner et al. 2008). If this is true, then double *nrpe7 nrpe4* or *nrpd7 nrpd4* mutants should have the same mutant phenotypes as either single mutant. Currently, there are ecotype differences in the *nrpe7* and *nrpd4* alleles. Therefore, alleles from a common ecotype should be isolated and crossed. Alternatively, one can make an *nrpe7* x *nrpd4* inter-ecotype cross, followed by a series of backcrosses to one of the wild-type parent ecotypes.

The experiments proposed above suggest that the knowledge of Rpb4-Rpb7 sub-complex functions in yeast might translate to Pol IV and Pol V (Rpb4-Rpb7 roles in *Arabidopsis* Pol II are discussed later). However, this has not been demonstrated. Therefore, I propose to test the hypothesis that NRPD4-NRPD7 and NRPE4-NRPE7 heterodimers form in the context of Pol IV and Pol V, analogous to the Rpb4-Rpb7 sub-complex in yeast. To test this hypothesis, one can perform *in vitro* pull down assays with recombinantly expressed NRPD4/NRPE4, NRPD7 and NRPE7. Specificity of these putative heterodimers could be demonstrated by testing for interactions among

NRPD4/NRPE4, NRPD7, NRPE7, NRPB4 and NRPB7. NRPB4 and NRPB7 should be absent in Pol IV and Pol V based on the mass spectrometry and co-IP evidence, and therefore the interactions between NRPE4/NRPD4 and either NRPD7 or NRPE7 are expected to be specific (Chapter 3) (Ream et al. 2009). NRPB4 and NRPB7 have previously been shown to interact (Larkin and Guilfoyle 1998). The interaction is most robust when the recombinant proteins, which localize to insoluble inclusion bodies, are renatured together instead of separately (Larkin and Guilfoyle 1998). This data suggests that the proteins stabilize each other and perhaps become more soluble when together. Indeed, once formed, the sub-complex is quite stable even in high concentrations of urea (Larkin and Guilfoyle 1998). These results have implications for testing interactions of NRPE4/NRPD4 with NRPE7 and NRPD7. Co-expression of these subunits may make them more soluble and alleviate the need for a renaturation step. If sufficient quantities are generated, the sub-complexes could potentially be crystallized. Crystal structures of the Pol I, II and III Rpb4-Rpb7 equivalents in yeast have shown differences in structural conformation that may translate to their polymerase specific functions (Armache et al. 2005; Jasiak et al. 2006; Kuhn et al. 2007). Structural data might be informative towards understanding how the Pol IV and Pol V sub-complexes function in their respective polymerases.

The positions of Rpb4-Rpb7 in the crystal structure of Pol II (and the equivalent subunits in Pol I and Pol III) allow for interaction with a variety of transcription initiation and RNA processing factors (reviewed in (Choder 2004; Sampath and Sadhale 2005)). In budding yeast, the Rpb4-Rpb7 sub-complex associates stoichiometrically with Pol II during stationary phase growth but Rpb4-Rpb7 is sub-stoichiometric during log phase

(Choder and Young 1993). Models have proposed that specific factors or post-translation modifications recruit the Rpb4-Rpb7 or Rpa14-Rpa43 sub-complexes to Pol II or Pol I, respectively, as they are needed for transcription, or vice versa (Rosenheck and Choder 1998; Peyroche et al. 2000). These models suggest that the stalk sub-complexes Rpa14-Rpa43 and Rpb4-Rpb7-like subunits mediate cross-talk with the core polymerase and associated factors. Therefore, it will be interesting to determine which proteins in *Arabidopsis* directly interact with NRPD4/NRPE4, NRPD7 and NRPE7. One approach is to perform yeast-two-hybrid assays to identify interactors of these subunits. Alternatively, epitope tagged lines of NRPD7, NRPE7 or NRPE4/NRPD4 could be used to identify novel interactors by mass spectrometry. While it is likely that subunits of Pol IV or Pol V will be identified using this approach, proteins that interact with NRPD7, NRPE7 or NRPE4 outside the context of Pol IV and Pol V might also be identified. These latter subunits might have been missed in the purification of Pol IV or Pol V by two different groups that immunoprecipitated the largest subunits (Huang et al. 2009; Ream et al. 2009). It is difficult to predict the proteins that may interact with the Pol IV or Pol V Rpb4-Rpb7-like sub-complexes. However, several clues narrow down the field. Pol V-dependent transcripts are not poly-adenylated, in contrast to Pol II transcripts. The 5' ends of Pol V transcripts are either tri-phosphorylated or capped, which is typical of RNA polymerase transcripts. The machinery required to perform these processes may interact with the sub-complexes based on several lines of evidence. First, yeast Rpb7 interacts with the nascent RNA (Ujvari and Luse 2006). Second, Rpb4 is required for recruiting 3' end processing factors in yeast, such as Rna14 and Rna15 (Runner et al. 2008). Although these factors are enriched at poly-adenylated regions that signify

termination, which are probably lacking at least for Pol V transcript units, analogous factors may be involved in signaling Pol IV or Pol V termination (Runner et al. 2008; Wierzbicki et al. 2008).

As stated previously, yeast Rpb7 interacts with the nascent RNA. Therefore, it is plausible to assume that RNA emerging from the Pol IV or Pol V active sites, via the RNA exit channel, also interacts with NRPD7 or NRPE7. Crystallization of Pol IV and Pol V with bound RNA would provide insight into this process, and should be a long-term goal for Pol IV and Pol V research. In the meantime, exploiting other strategies may give hints to this process. *In vitro*, electrophoretic mobility shift assays would test if NRPE4, NRPD7 or NRPE7 can bind RNA, using yeast Rpb7 as a control. *In vivo*, FRET approaches may be informative. Recently, FRET approaches were used to demonstrate that Rpb4, but not Rpb3, is close to RNA emerging from the active site (Chen et al. 2009).

Localization of the NRPD4-NRPD7 and NRPE4-NRPE7 sub-complexes may also reveal insights into the functions of these proteins. In yeast, only a portion of the Rpb4-Rpb7 sub-complex is associated with Pol II, whereas the remaining complex can function in isolation from core Pol II in the nucleus and cytoplasm (Lotan et al. 2005; Selitrennik et al. 2006). Using FLAG-tagged NRPD7 or NRPE7 transgenic lines and antibodies to native NRPD4 (He et al. 2009) in *Arabidopsis* nuclei, one can ask if these subunits significantly co-localize and, in separate experiments, what fraction of the sub-complex co-localizes with Pol IV and Pol V. If a significant portion of the sub-complexes localize independent of Pol IV and Pol V, this would argue that they have additional functions outside their roles in core Pol IV and Pol V. It will be also be interesting to determine if

any of these subunits co-localize in the nucleolus. The nucleolus is a sub-nuclear compartment where many different types of RNA processing occur, including siRNA and miRNA production (Appendix B)(Li et al. 2006; Pontes et al. 2006; Fujioka et al. 2007; Song et al. 2007). Indeed, if RNA transcripts generated by Pol IV or Pol V are targeted to the nucleolus for Dicer destruction, then the stalk sub-complexes (NRPD4-NRPD7, NRPE4-NRPE7) of Pol IV and Pol V may guide them to the nucleolus along with other factors. This might be analogous to the localization of Rpb4-Rpb7 to P-bodies in yeast (Sheth and Parker 2003; Lotan et al. 2005; Lotan et al. 2007). It would be especially interesting if NRPD7 localizes to the nucleolus, where NRPD1 is excluded (Appendix B)(Pontes et al. 2006). Again, this would argue for an independent role for NRPD7-NRPD4.

Another potential role for NRPD4 and NRPD7 might be in the spread of silencing between cells. Catalytic subunits of Pol IV are required for the short range spreading of silencing and for the reception of the long distance silencing signal (see Introduction) (Brosnan et al. 2007; Dunoyer et al. 2007; Smith et al. 2007). Given that yeast Rpb4-Rpb7 is capable of shuttling from the nucleus to the cytoplasm, it is possible that NRPD4-NRPD7 could receive the long distance silencing signal as it is off-loaded from the phloem and transfer it back to the nucleus, where Pol IV and RDR2 amplify the signal and channel it to DCL3 for siRNA production. During short range spread of silencing, the mechanism of movement of the silencing signal is unknown. Although siRNAs and the plasmodesmata are implicated, its possible that longer RNAs move between cells. Could NRPD4 and NRPD7 acts as inter-cellular shuttles of the mobile, short range silencing signals? Crossing *nrpd4* and *nrpd7* mutants into reporter lines that

assay for short range silencing movement would be a first step towards answering these questions. Testing for localization of the subunits in the cytoplasm would also be interesting. If these subunits localize to the cytoplasm, live cell imaging techniques may be able to test for shuttling between the nucleus and the cytoplasm.

The roles of NRPB4 and NRPB7

Other than the initial report characterizing NRPB4-NRPB7 heterodimers in *Arabidopsis*, nothing is known about the role of these subunits in Pol II (Larkin and Guilfoyle 1998). To understand the role of NRPB4, I have isolated two viable, homozygous insertion mutants, *nrpb4-1* and *nrpb4-2*. The recovery of viable alleles is expected based on observations in budding yeast, where *rpb4* mutants are temperature sensitive, but not from fission yeast, where *rpb4* mutants are lethal (Sakurai et al. 1999). Neither *Arabidopsis* allele has been tested for temperature sensitivity. In addition, it is not known if either allele results in a complete loss-of-function at the protein level. Antibodies I have generated to NRPB4 now make it possible to address this question. Both mutations cause the plants to be short and stunted in development (Figure 4A). In addition, the mutants display a wide array of floral organ identity and organ number phenotypes and the roots appear to be shorter than wild-type plants (data not shown). The plants eventually flower but have such extremely low seed set that they are essentially sterile. However, for any future studies with this mutant, a more thorough qualitative or quantitative comparison of these phenotypes to wild-type plants is necessary.

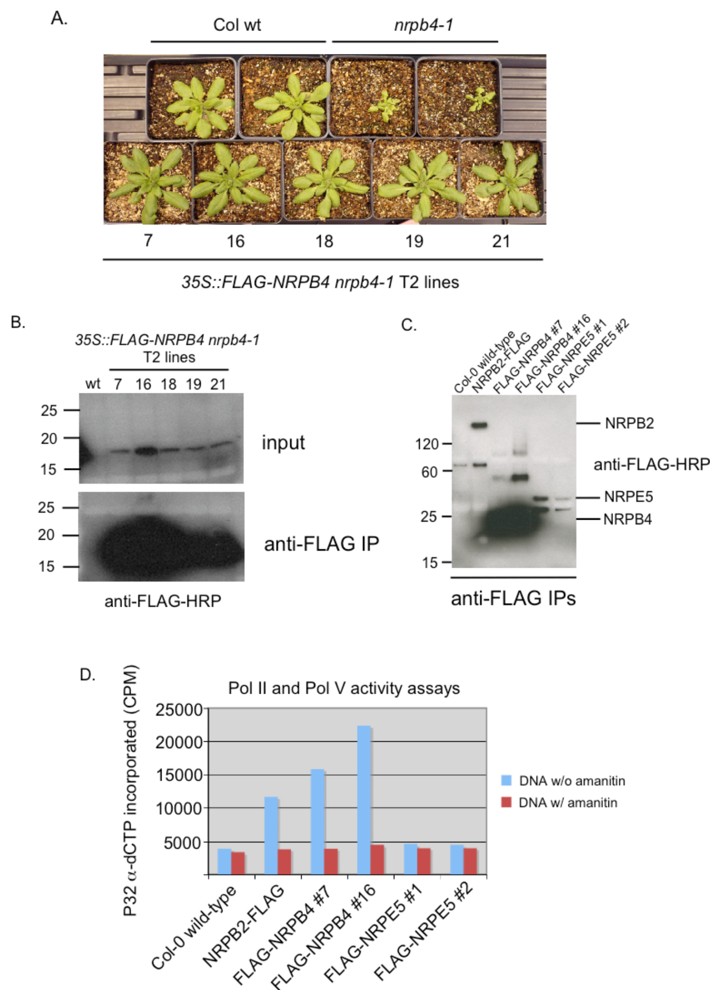


Figure 4. FLAG-NRPB4 rescues an *nrpb4* mutant and is active for promoter-independent transcription. A. Phenotypes of *nrpb4* mutants and rescue by *35S:FLAG-NRPB4* transgene. B. Immunoblot analysis of FLAG-NRPB4 protein levels in input and anti-FLAG IP samples, detected with anti-FLAG-HRP. C. Immunoblot analysis as in B., but with IP samples used for the activity assays shown in D. D. Promoter-independent transcription assay showing activity in IPs of NRPB2-FLAG and FLAG-NRPB4 transgenic lines (indicative of Pol II), but not in IPs of FLAG-NRPE5 transgenic plants (indicative of Pol V).

I hypothesize that the *nrpb4* phenotypes are pleiotropic in nature and are not directly caused by the *nrpb4* mutation, but instead are caused by general defects in transcription of developmentally important genes. To test this, microarray analysis could be performed to determine the changes in gene expression at a global level in *nrpb4* mutants relative to wild-type plants. A *35S::FLAG-NRPB4* transgene transformed into the *nrpb4-1* mutant rescues the sterility phenotype and FLAG-NRPB4 immunoprecipitates display promoter-independent activity that are inhibited by α-amanitin, indicative of Pol II (Figure 4B-D)(T.S. Ream, J.R. Haag and C.S. Pikaard, unpublished data). The transgene also rescues the aerial growth phenotype as compared to wild-type plants, but display stochastic defects in floral organ development (Figure 4A). The floral phenotypes have not been quantified and should be verified.

There are at least three possible explanations for the viability of the *nrpb4* mutants. First, *nrpb4-1* and *nrpb4-2* are hypomorphic alleles, with some NRPB4 activity that maintains viability, although the plants are sick. Second, NRPB4 may not be essential for the activity of Pol II and for any roles it has outside of Pol II transcription, similar to budding yeast *rpb4* knockouts, but complete loss of NRPB4 nonetheless results in pleiotropic, developmental phenotypes. Alternatively, NRPD4 might inefficiently substitute for NRPB4 in the *nrpb4* mutant, leading to the *nrpb4* phenotypes. In this latter case, probing Pol II immunoprecipitates in the *nrpb4* mutant for NRPD4 would test this question. Yeast Rpb4 and *Arabidopsis* NRPB4 proteins are 30% identical, whereas the yeast and *Arabidopsis* NRPB4 proteins are only 7% identical to NRPD4. The *Arabidopsis NRPB4* gene partially complements the yeast *rpb4* knockout, whereas *NRPD4* is unable to complement yeast *rpb4* knockouts (He et al. 2009). Based on this

evidence, one may argue that it is unlikely that NRPD4 significantly complements NRPB4 function.

To test if the function of yeast Rpb4 translates to plant NRPB4, it would be useful to test the promoter-dependent activity of immunoprecipitated Pol II in the *nrpb4* mutant compared to wild-type using a native antibody against a Pol II specific subunit, such as NRPB1 or NRPB7. It would also be interesting to test if NRPB7 still associates with Pol II in these *nrpb4* mutants because yeast Rpb7 is not detectable in purified Pol II in *rpb4* mutants, as mentioned previously (Edwards et al. 1991). To test if NRPB4 is important for stabilizing NRPB7 in Pol II, as in yeast, one can transform the *35S::FLAG-NRPB7* transgene into *nrpb4*/+ plants and assay for rescue of the *nrpb4* mutant phenotypes in the following generation.

Subunits with unknown functions

There are at least three RNA polymerase subunits for which no function has been determined that appear to be expressed in some tissues (see Chapter 4). These are the NRPB10-like protein (At1g61700), the NRPB12-like protein (At1g53690) and the NRPB7-like or NRPD7-like protein (At4g14520). These proteins were not identified in any of the RNA polymerase purifications and have not tested positive for interaction with any polymerase by co-IP (Chapter 4). (The *NRPB10*-like gene, At1g61700, has not been tested.) Insertion mutations within introns or exons are available for each of these genes, but it will be difficult to assign a function for these genes without knowing precisely what phenotype to analyze (Table 1). It is possible that these loci are expressed but are not functioning because they have significantly diverged amino acid sequences that preclude

their association with a given RNA polymerase. It is also possible that these subunits associate at such low levels with a given polymerase that they are not detected in our mass spectrometry datasets, or by co-IP. Another possibility is that these subunits represent a mid-point between becoming pseudogenes and being non-functional duplicates within their respective families. As one approach to understand their possible functions, one could test if over-expression of the alternative NRPB10-like or NRPB12-like subunits rescue a lethal mutant in NRPB10 or NRPB12. In the case of the NRPB12-like and NRPB10-like subunits, transgenic lines are available. However, only *nrpb12* mutants have been shown to be lethal and can therefore be tested for functional rescue (Appendix C)(Onodera et al. 2008). As a control, the *nrpb12* mutants can be rescued with a genomic or cDNA version of *NRPB12*. A functional, epitope-tagged NRPB12 transgenic line is currently lacking in the Pikaard lab and would be a useful tool to have. Insertions in the 5' and 3' UTR of the canonical *NRPB10* gene can be isolated as homozygous mutants, but it is likely that these insertions do not effect the transcription of *NRPB10*, although this has not been formally tested (Table 1). If the alternative subunits rescue their corresponding mutants (lethality for *nrpb12* mutants), one could argue that there are assembly mechanisms in the nucleus to selectively incorporate only one paralog into the polymerases, but that this regulation can be overcome by over-expression of the other subunit in the lethal mutant's background. In the case of the *NRPB7*-like or *NRPD7*-like gene, its possible, but unlikely, that this gene is only active when the *NRPD7* or *NRPB7* genes are mutated.

Several studies have examined the relative conservation of RNA polymerase subunit structure and function by testing for hetero-complementation. For example,

studies have used human, plant or fission yeast subunit genes to rescue corresponding knockout mutants in budding yeast (McKune et al. 1995; Imazawa et al. 2005; He et al. 2009). These studies have also been useful in determining if a putative homolog actually is the ortholog of a subunit for which little conservation is retained (Imazawa et al. 2005). To understand the relative conservation of subunit function from plants to fungi, one could take a comprehensive approach to testing which RNA polymerase subunits functionally complement their orthologs in yeast. This has been done in the case of *Arabidopsis* NRPB4 and NRPD4, as described previously. Hetero-complementation may also provide insight into the roles of the RNA polymerase subunits with no known function. For example, if both NRPB12 and the NRPB12-like complement a yeast *rpb12* mutant, then this would argue that the NRPB12-like protein is functional but is regulated in a way that it does not get incorporated into Pol II in a detectable manner.

Characterization of RNA polymerase subunit antibodies

From a technical perspective, it would be useful to determine if any of the polyclonal antibodies I have generated against various RNA polymerase subunits inhibit RNA polymerase transcription *in vitro* (Table 2). Several reports from yeast demonstrated that antibodies directed against various subunits inhibit transcription (Buhler et al. 1980; Breant et al. 1983; Huet et al. 1985). If any of the antibodies block transcription, they would provide valuable alternatives to α-amanitin to block transcription, potentially in a polymerase-specific manner. Likewise, it will be useful to determine which antibodies are capable of immunoprecipitating their respective

polymerases and which antibodies do not block transcription or interfere with subunit interactions.

Studies in yeast clearly demonstrate that each RNA polymerase subunit has a unique and often essential role in the function of the polymerase (see Introduction). To this end, it is reasonable to assume that Pol IV and Pol V non-catalytic subunits contribute important roles to Pol IV and Pol V, as reflected by mutations in *NRPE4*, *NRPE5* and now *NRPE7* (Figure 3)(Chapter 3)(He et al. 2009; Huang et al. 2009; Lahmy et al. 2009; Ream et al. 2009). Thus, there are many avenues of future research regarding the small subunits of RNA polymerases IV and V in *Arabidopsis*. In addition, viable *nrpb4* mutants provide a tool to study Pol II transcription in a higher eukaryote (Table 1, Figure 4). I have created many tools to begin answering some of questions proposed, and the experiments I have outlined above will be important for expanding our understanding of RNA polymerase functions across kingdoms.

ii.

TOWARD MECHANISMS OF POL IV AND POL V GENE SILENCING

Introduction

Genetic and cytogenetic assays have shed light on the roles of Pol IV and Pol V in gene silencing by determining the pathways in which they are involved, who they are interacting with and what types of targets they are regulating. My thesis work demonstrates that Pol IV and Pol V are large, multi-subunit complexes that establish gene silencing at heterochromatin. However, much remains to be revealed about how Pol IV

and Pol V function. In particular, the *in vitro* biochemical activities of Pol IV and Pol V are not published. Our lab has preliminary evidence for Pol IV activity, but much more needs to be done (see below). Pol V transcription has been demonstrated *in vivo* using genetic approaches in concert with RT-PCR, but there is no *in vitro* biochemical evidence for Pol V transcription (Wierzbicki et al. 2008). Therefore, nothing is known about how the twelve core subunits comprising Pol IV or Pol V function to initiate transcription, elongate, proofread and cleave their transcribed RNAs, or facilitate RNA processing in downstream steps.

Although some of the gene targets regulated by Pol IV or Pol V are known, very little is known about how Pol IV or Pol V are recruited specifically to their targets (Huettel et al. 2006; Huettel et al. 2007; Mosher et al. 2008). Transcripts that are produced by Pol IV or Pol V are part of a pathway involving an RNA-dependent RNA polymerase, RDR2, and a Dicer protein, DCL3. But do all Pol IV and Pol V transcripts channel through this pathway? Do noncoding Pol IV or Pol V RNAs have a role in heterochromatin formation? To this end, *nrpe1* and *nrpd2* mutants display dissociated chromocenters, coincident with decondensation of H3K9 methylation and dispersal of 5S rDNA clusters (Chapter 2)(Onodera et al. 2005)(Pontes et al., 2009). These phenotypes are indicative of disorganized higher-order chromatin structure, similar to observations in mutations of the CG maintenance methyltransferase, *met1*, and the chromatin remodelers *ddm1* and *drd1* (Soppe et al. 2002)(Pontes et al., 2009). What other proteins are involved in higher-order chromatin structure and how are these structures formed? This next section will speculate on the answers to some of these questions and discuss large-scale

genomics- or proteomics-based approaches that might be integrated to answer some of these questions.

Activities of Pol IV and Pol V

The biggest challenge in Pol IV and Pol V research is elucidating their *in vitro* biochemical activities. Initial tests of promoter-independent DNA-dependent RNA polymerase activity from DEAE-sepharose column fractions revealed that fractions enriched for Pol IV or Pol V did not display detectable levels of activity above background, whereas fractions enriched for Pol I, II or III did show appreciable levels of activity (Chapter 2) (Onodera et al. 2005). In the last few years, PhD student Jeremy Haag has worked hard to demonstrate activity from Pol IV and Pol V immunoprecipitated fractions using extraction protocols similar to those used for determining the subunit compositions of Pol IV and Pol V. Pol V has not shown activity in any assay attempted so far, which has included numerous types of DNA and RNA substrates in single-stranded or double-stranded forms, with or without primers. A common feature of all these substrates is that they are naked nucleic acid templates. It is possible that Pol V requires a very specific template that is only present *in vivo*. It may be that Pol V requires chromatin templates, possibly including modified histones and associated proteins that may recruit Pol V. These sites would presumably have *cis* regulatory sequences that Pol IV or Pol V initiation factors recognize.

In vivo, it is possible to use genetics to uncover the roles of the small subunits in Pol V transcription, as discussed in the previous section. Any viable mutants (i.e. *nrpe7, nrpe4, nrpe3b, nrpe9a nrpe9b*) should be tested for their requirement of Pol V-dependent

transcripts. If an *in vitro* assay can be devised, activity requirements for these subunits could be assayed by transforming *pNRPE1::NRPE1-FLAG* constructs into any of these mutant backgrounds and assaying for activity in anti-FLAG immunoprecipitates of Pol V. Alternatively, *pNRPE1::NRPE1-FLAG nrpe1-11* plants could be crossed into the aforementioned mutant backgrounds or antibodies to the NRPE1 subunit could be used for immunoprecipitation in the mutant backgrounds.

Recently, Jeremy Haag has shown that Pol IV has a small but reproducible activity on DNA templates that is dependent on the active site and is resistant to α-amanitin activity, which is predicted based on the divergence of amanitin-interacting amino acids in the Pol IV active site compared to Pol II (Haag and Pikaard, unpublished) (Bushnell et al. 2002; Erhard et al. 2009). These first hints of activity are promising and are currently being followed up. Ultimately, it will be necessary to know the precise buffer conditions and cofactor requirements for transcription, as well as the preferred templates. Over time, it is worth exploring other plant models as a source of tissue for Pol IV purification. Broccoli and maize are great sources of kilogram quantities of inflorescence tissue, but they are limited by laborious transformation protocols and their large size, making it impractical to grow a lot of individuals in a small space.

RDR2 was identified in the affinity purification of Pol IV and its interaction with Pol IV has been confirmed by co-IP (Haag, Ream and Pikaard, unpublished data). This association complicates the identification of Pol IV activity because background activity attributed to RDR2 is also found in Pol IV fractions (Haag and Pikaard, unpublished data). One caveat revealed by biochemical studies on RDR6 is that the enzyme has a strong preference to transcribe ssDNA in addition to ssRNA (Curaba and Chen 2007). If

this activity is present in RDR2, then this would complicate the interpretation of activity results obtained from affinity purified Pol IV. However, the RDR2-Pol IV interaction provides insight into the mechanism of Pol IV action and supports the genetic and cytogenetic evidence that Pol IV and RDR2 act in concert in the same pathway.

It is possible that the activities of Pol IV and RDR2 are coupled and depend on each other for stability, and if so this will make it hard to tease apart the activity attributed to each protein. Immunoblots of crude extracts in wild-type, *rdr2* and *nrpd1* mutants will test if the stability of Pol IV and RDR2 require each other. If RDR2 and Pol IV do not require each other for stability, then performing Pol IV activity assays in a *rdr2* mutant will eliminate background RDR2 activity and unambiguously allow Pol IV activity to be revealed. Alternatively, Pol IV and RDR2 may stimulate each other's activities, such that a mutation in either gene causes loss of activity of both proteins. A similar observation for Dicer-stimulation of RNA-dependent RNA polymerase activity has been observed in fission yeast (Colmenares et al. 2007). Some of these approaches are currently being pursued by Jeremy Haag in the Pikaard lab.

Once Pol IV activity assays are established, it will be interesting to test *nrpd7, nrpd4* and any other viable small subunit mutants for defects in *in vitro* transcription. Based on work in yeast, I would not expect there to be any obvious impairment to transcription in promoter-independent assays, but the subunits may be required for promoter-dependent transcription. However, this requires knowledge of Pol IV transcription units, which are currently unknown (see below).

RNA polymerases perform a variety of functions in addition to RNA catalysis. These include pausing at certain sites, backtracking along the template, RNA cleavage

and proofreading. All of these activities are essential for a polymerase to overcome natural barriers in the template and to correct mis-incorporated nucleotides. It is unknown if Pol IV and Pol V possess any of these activities *in vivo* or *in vitro*. Given that they share NRPB9 subunits with Pol II that mediate some of these activities, it is feasible that they would also perform these activities. Lack of a proofreading activity might result in the incorporation of errors into a Pol IV or V transcript, and this could have serious implications on the specificity of gene silencing, which uses RNA homology as the specificity determinant. TFIIS acts in concert with Rpb9 in yeast to mediate RNA cleavage at the Pol II active site. Does TFIIS also interact with Pol IV and Pol V to perform this function? Creation of TFIIS-tagged lines or antibodies are essential to answer this question.

Another aspect of Pol IV and Pol V biology that remains to be determined is how the core complexes are assembled. A long-term, ground-up approach to tackling this problem is to *in vitro* reconstitute the enzyme and determine the order in which subunits bind to each other. This is likely to be a very technically challenging process, since all twelve subunits must be expressed and soluble. However, this approach has worked to reconstitute the single, multi-subunit RNA polymerase in archaeabacteria (Werner and Weinzierl 2002). This approach may also yield enough Pol IV or Pol V for X-ray crystallography.

Understanding where Pol IV and Pol V act in the nucleus

If Pol IV is transcribing DNA, then this supports a model whereby Pol IV acts first on the DNA to generate a transcript that can subsequently be amplified by RDR2

into dsRNA, the substrate for DCL3. But how will Pol IV recognize where to transcribe in the genome? Data accumulated on a genome-wide scale will be useful in beginning to answer this question.

Genome-wide analyses are becoming more cost-effective and provide greater resolution and depth than ever before. By utilizing whole genome approaches, it should be possible to paint a chromatin landscape that is dependent on Pol IV or Pol V. Specifically, having maps of DNA methylation, siRNA production and transcript levels at extremely high resolution (<10 bp) in *nrpd1* and *nrpe1* vs. wild-type plants might reveal where Pol IV and Pol V are transcribing. In combination with ChIP (chromatin immunoprecipitation) of NRPD1-FLAG and NRPE1-FLAG lines, one should be able to demonstrate where Pol IV and Pol V reside on the chromatin. This type of approach has confirmed the association of Pol V with the chromatin region of *IGN5*, a Pol V-dependent locus (Wierzbicki, Ecker and Pikaard, unpublished data). If there are correlations with loss of transcripts at these specific loci in Pol IV or Pol V mutants, then this suggests Pol IV or Pol V transcribes those regions. These studies could help establish the direct targets of Pol IV and Pol V, as opposed to previous studies that analyzed changes in RNA expression in *nrpe1* and *nrpd1* mutants, which could be direct or indirect causes of Pol IV or Pol V mutations (Huettel et al. 2006; Mosher et al. 2008). The identification of Pol IV and Pol V target regions will allow for bioinformatics approaches to identify potential Pol IV and Pol V *cis* regulatory elements. If such regulatory elements are found, screens for transcription factors could be designed using oligos corresponding to these elements. A single transcription factor, KTF1/SPT5-like, has been isolated as a putative Pol V interacting protein that is required for RNA-directed

DNA methylation and siRNA production (Huang et al. 2009). It will be interesting to study the relationship between this protein and Pol V, and whether KTF1/SPT5-like is involved in recruiting Pol V. Comparing the localization of KTF1/SPT5-like on a genome-wide scale with Pol V will be informative towards this goal. In addition, it will be necessary to confirm whether KTF1/SPT5-like immunoprecipitates with Pol V.

These genome-wide approaches can be taken one step further and applied to the mutants in *nrpd7, nrpd4, nrpe7* and *nrpe5*, and any other subunit mutant combinations that are viable. Mapping where Pol IV and Pol V localize to DNA in these backgrounds, in combination with DNA methylation and siRNA profiles, will reveal if there are locus-specific requirements for any of the above subunits, as suggested for NRPE5, and based on evidence from *rpb5* mutants in yeast (Miyao and Woychik 1998; Soutourina et al. 2006; Lahmy et al. 2009). This can further be teased apart by looking at RNA levels by RNA-ChIP of NRPD1 and NRPE1 in these non-catalytic subunit mutants to see if loss of localization to certain sequences also results in loss of transcription, as expected.

Role of post-translational modifications in RNA polymerase-mediated silencing

In addition to DNA regulatory sequences or transcription factors, Pol IV and Pol V may be recruited by particular chromatin modifications, such as DNA methylation patterns or histone modifications. Histone hyper-acetylation and H3K4 methylation generally favor Pol II transcription, but nothing is known about histone modifications that correlate with Pol IV or Pol V action. To better understand the recruitment of Pol IV and Pol V to their templates, it will be necessary to know which chromatin environments they transcribe. With the advances in mass spectrometry, it is possible to crosslink Pol IV or

Pol V to chromatin and immunoprecipitated the chromatin associated with either protein. Mass spectrometry of the associated protein fraction will identify specific histone modifications that may correlate with Pol IV or Pol V binding or activity. Advances in mass spectrometry now allow for analysis of undigested histones, revealing the spectrum of histone modifications on a given histone subunit. This is very powerful because it determines which modifications are likely to associate with each other on a given histone, potentially in a combinatorial manner. To this end, isolation of histones associated with Pol IV or Pol V after chromatin immunoprecipitation using NRPE1-FLAG or NRPD1-FLAG lines will determine the spectrum of histone modifications associated with Pol IV or Pol V.

Functional characterization of the elaborated CTD of Pol V will be instrumental in understanding how Pol V is regulated. Studies by several groups, including our lab, have shown that the CTD is essential for silencing (Haag and Pikaard, unpublished data)(El-Shami et al. 2007). However, virtually nothing is known about the post-translation modifications that are regulating Pol V activity or interaction with transcription factors or RNA modifying proteins. Developing assays that will reveal the potentially dynamic post-translation modifications of Pol V *in vivo* are necessary.

In conclusion, these assays should provide a comprehensive picture of how and where Pol IV and Pol V transcribe, the subunits that are involved in the transcription of each locus, and how Pol IV and Pol V are recruited to their targets.

Pol IV and Pol V in the context of other gene silencing components

In addition to knowing the templates, products and regulatory mechanisms of Pol IV and Pol V transcription *in vivo* and *in vitro*, it will be important to understand the role of Pol IV and Pol V within the context of the pathways they regulate. Specifically, this includes understanding the network of protein-protein interactions that presumably occur among proteins involved in Pol IV- or Pol V-dependent silencing pathways. Identification of the Pol IV and Pol V core subunit structures using high resolution mass spectrometry has laid the foundation for these future studies. Traditional purification strategies have relied on multiple chromatographic steps to generate sufficiently pure quantities of protein to be analyzed. These procedures must be established for each protein complex and are very time-consuming. In contrast to traditional purification, single step or tandem affinity purifications are gaining widespread popularity. Indeed, I have utilized this approach to purify Pol IV and Pol V. The combination of high resolution mass spectrometry identifies hundreds of proteins and thousands of sequenced peptides in each sample (Table 3). This technique is very adaptable to medium throughput analysis where many transgenic lines can be made and the tagged proteins purified. Using appropriate controls such as wild-type plant tissue or vector-only transgenic lines as filters, one can begin to identify potential interacting proteins. We have initiated such "screens" in the Pikaard lab by affinity purifying a variety of epitope-tagged proteins involved in RNAi. In the context of Pol IV and Pol V, several interesting candidate genes have emerged (Table 4). Two interesting candidate genes are the homologs of the animal Reptin and Pontin proteins, which are Ruv-B-like DNA helicases involved in chromatin remodeling, stability of ribonucleoprotein complexes, interaction

with TATA-binding protein and silencing of heterochromatin (Ream and Pikaard, unpublished)(Qi et al. 2006; Gallant 2007; Venteicher et al. 2008; Huang et al. 2009; Ream et al. 2009). These proteins were among the most abundant proteins identified in Pol IV and Pol V IPs from callus tissue, but they are much less abundant, or absent, from Pol IV or Pol V IPs from leaf tissue, making it unclear if there is a tissue-specific association (Table 4). Although these proteins were identified in our control IPs, the total peptide counts and unique peptides corresponding to each protein were much lower in the controls compared to Pol V (Table 4). In addition, the Pontin homolog was identified in the affinity purification of Pol V from cauliflower (Huang et al. 2009). Clearly, these proteins are intriguing candidates that should be tested for association with Pol V (and Pol IV) by co-IP. Mutant analysis is complicated by the essential role of the Pontin homolog, but it is possible that knock-down in expression by RNAi may be informative (Holt et al. 2002).

Table 3. Total number of proteins identified in affinity purified Pol IV and Pol V lines

IP	Proteins with >2 peptide matches	Peptide count
callus		
NRPE1	623	3168
control	449	1662
NRPD1	381	1444
leaf		
NRPD1	186	1090
NRPE5	321	1508

Table 4. Peptides matching Reptin and Pontin homologs in Pol IV or Pol V affinity purifications

	At5g67630 Reptin-like		At5g22330 Pontin-like	
callus	peptides counted	total scans	peptides counted	total scans
NRPE1	31	131	41	206
NRPD1	18	35	18	24
control	5	6	9	15
leaf				
NRPE5	0	0	1	1
NRPD1	5	8	2	3

Initially, nearly 400 proteins were confidently identified in the Pol IV purification and over 600 in the Pol V purification (Table 3). However, it is likely that many of these proteins are contaminants, due to the limitations in purity of a single affinity purification step. By using a combination of computational filters, graduate student Sarah Tucker has helped me to identify proteins that are uniquely associated with Pol IV and/or Pol V, all five polymerases, or any other combination of the polymerases (Table 5). This approach eliminated proteins found in the control datasets or in other IPs of proteins that do not interact with Pol IV or Pol V under our purification conditions. After applying these filters, fewer than twenty-five proteins were identified to be common between Pol IV and Pol V, and many of these have only a single peptide hit, which is not above a two-peptide threshold for a confident protein identification. Of the proteins with >2 peptide hits, the majority are RNA polymerase subunits that have been confirmed partners of Pol IV or Pol V (Table 5).

Despite the progress in purifying Pol IV and Pol V, it is likely that the complexes that I have identified are incomplete. Yeast Pol II requires a series of initiation and elongation factors for activity *in vivo* and it is reasonable to assume that Pol IV or Pol V

require similar co-factors for their specificity or activity. As shown above, initial searches through the Pol IV and Pol V protein IP datasets have not revealed any obvious candidates.

Table 5. Proteins associated with Pol IV and Pol V after filters.

> 1 peptide?	Gene ID	description
x	At4g16265	NRPE9b/NRPD9b/NRPB9b
x	At3g52090	NRPE11/NRPD11/NRPB11
x	At2g04630	NRPE6b/NRPD6b/NRPB6b
x	At5g51940	NRPE6a/NRPD6a/NRPB6b
	At5g41010	NRPB12/NRPC12/NRPE12/NRPD12
x	At2g15400	NRPE3b
x	At1g11475	NRPB10/NRPE10/NRPD10/NRPC10/NRPA10
x	At3g23780	NRPD2/NRPE2
x	At4g15950	NRPE4/NRPD4
	At1g54410	dehydrin family protein
	At2g41100	ATCAL4, TCH3 | TCH3 (TOUCH 3)
	At4g08920	BLU1, HY4, OOP2, CRY1 | CRY1 (CRYPTOCHROME 1
	At5g49000	Identical to F-box/Kelch-repeat protein
	At5g26820	ferroportin-related
	At1g74470	geranylgeranyl reductase
	At1g07660	histone H4
	At1g12780	UGE1 | UGE1 (UDP-D-GLUCOSE/UDP-D-GALACTOSE 4-EPIMERASE 1)
x	At5g45390	NCLPP3, NCLPP4, CLPP4 | CLPP4 (Clp protease proteolytic subunit 4)
x	At1g02560	NCLPP1, NCLPP5, CLPP5 | CLPP5 (NUCLEAR ENCODED CLP PROTEASE 1)
x	At3g56340	40S ribosomal protein S26 (RPS26C)
x	At2g07698	ATP synthase alpha chain, mitochondrial, putative
	At5g25520	transcription elongation factor-related
x	AtMg01190	ATP1 | ATPase subunit 1
x	At5g46290	KAS I | KAS I (3-KETOACYL-ACYL CARRIER PROTEIN SYNTHASE I); fatty-acid synthase
x	At3g22110	PAC1 | PAC1 (20S proteasome alpha subunit C1); peptidase

One possibility for the lack of identification of such transcription factors is that they transiently interact with Pol IV or Pol V. A second possibility is that the complexes isolated were not tightly associated with chromatin and therefore were more soluble than a chromatin-associated fraction that might bind these factors on the template. To address the first possibility, it is possible to perform *in vivo* cross-linking followed by affinity purification. This would presumably stabilize any transient interactions and allow for their identification. Such studies have been utilized in the purification of chromatin-bound yeast RNA polymerase II (Tardiff et al. 2007). Proteins that depended on cross-

linking included Mediator, general transcription factors, elongation factors, ribonucleoprotein particle (RNP) proteins, and histones. Furthermore, some of these interactions with Pol II were demonstrated to be RNase-sensitive.

A more controlled, quantitative approach to identifying proteins that exist in an IP above the background level, compared to our current approach which involves subtraction of background proteins from samples independently processed in the mass spectrometer, is quantitative mass spectrometry coupled with SILAC (Stable Isotope Labeling of Amino acids in cell Culture). This approach allows for two samples to be combined before analysis by mass spectrometry and therefore allows for determining quantitative differences between proteins in the two samples (reviewed in (Vermeulen et al. 2008)). However, it may not reveal all the transient interactions owing to the limitations of the affinity purification protocol used.

Once a series of datasets from different purified RNAi components have been collected, it is possible to cross-reference the datasets for common proteins that are not found in the control datasets and to eliminate proteins with single peptide hits. This will result in a list of proteins common to any combination of samples. If, for example, datasets from proteins A and B are compared, any common proteins that pass the filters may be important for bridging the functions of A and B in a pathway, especially if A and B do not interact. Not only will these analyses tease out potential interacting proteins among hundreds of background proteins, but comparing multiple datasets from proteins with the same (or different) epitope tag will generate lists of proteins that are common contaminants due to their affinity for a particular epitope or bead matrix, or which are

specific to a certain tissue type. Such "bead-ome" lists have been compiled in other organisms and will be useful in plants as well (Trinkle-Mulcahy et al. 2008).

Additional candidate proteins for Pol V interaction are the Spt4-like proteins in *Arabidopsis*. As stated previously, an KTF1/SPT5-like protein was identified in a Pol V affinity purified complex and mutant *ktf1*/*spt5-like* plants show defects in DNA methylation and siRNA production (Bies-Etheve et al. 2009; Huang et al. 2009). Spt5 proteins in yeast and humans interact with Spt4 to form the DSIF elongation factor complex, which is implicated in stimulating elongation, mRNA capping and preventing premature termination of transcription (Hartzog et al. 1998; Wada et al. 1998; Sims et al. 2004). In addition, yeast *spt4* deletion mutants result in loss of silencing of a reporter gene integrated adjacent to the mating-type loci and the telomeres (Crotti and Basrai 2004). Therefore, it is plausible that the SPT5-like/KTF1 protein in *Arabidopsis*, which interacts with AGO4 and possibly Pol V (Bies-Etheve et al. 2009; Huang et al. 2009), may also form a complex with an Spt4-like protein. Interestingly, *Arabidopsis* has two homologs of Spt4 (87% identical), in contrast to yeast and humans, which each have a single Spt4 homolog. In *Arabidopsis*, the *SPT4* genes are differentially expressed in floral tissues, but are otherwise ubiquitously expressed, although at different levels (Atgenexpress website: http://jsp.weigelworld.org/expviz/expviz.jsp). I hypothesize that *Arabidopsis* may have functionally diversified forms of SPT4 that interact with Pol II, Pol IV or Pol V in a manner that is potentially analogous to the function of the yeast and human Spt4. To test this, one can isolate mutations in each *SPT4* gene. If the mutants are viable, then DNA methylation and siRNA production at Pol V-dependent loci should

be assayed. Creation of transgenic lines harboring epitope-tagged versions of each protein could be used to assay for interaction with SPT5-like, Pol II, Pol IV or Pol V.

A role for Pol V in heterochromatin stability

Our initial studies on the *nrpd2* mutant showed that heterochromatin is severely disrupted compared to wild-type plants (Chapter 2)(Onodera et al. 2005). This is manifested in dissociation of chromocenters (areas of condensed heterochromatin in the nucleus). Later studies demonstrated that Pol V, but not Pol IV, is critical for this activity in a novel pathway dependent on DDM1, MET1 and DRD1 (Pontes et al., 2009). One hypothesis proposed was that higher-order heterochromatin formation is dependent on RNA scaffolds, as observed in other organisms (Maison et al. 2002). One hypothesis is that the activity of Pol V is required to generate such RNAs that act as scaffolds to nucleate chromocenters and maintain organization of heterochromatin. This implies that longer RNAs are important and siRNAs may not be relevant, as suggested by intact chromocenters in *dcl3*, *rdr2* and *nrpd1* mutants (Pontes et al., 2009). If Pol V transcription maintains chromocenter integrity, one would expect to see chromocenter disorganization in NRPE1 active site mutants. The requirement for DMS3 (a structural maintenance of chromosomes hinge domain protein) in Pol V transcription (Wierzbicki et al. 2009) is intriguing and may provide clues for a mechanism, in combination with Pol V, for high-order maintenance of heterochromatin by the ability to nucleate different regions of heterochromatin. Currently, it is unclear how DMS3 interacts with Pol V in the context of chromatin. Therefore, assaying for chromocenter dissociation in this mutant would be informative.

iii.

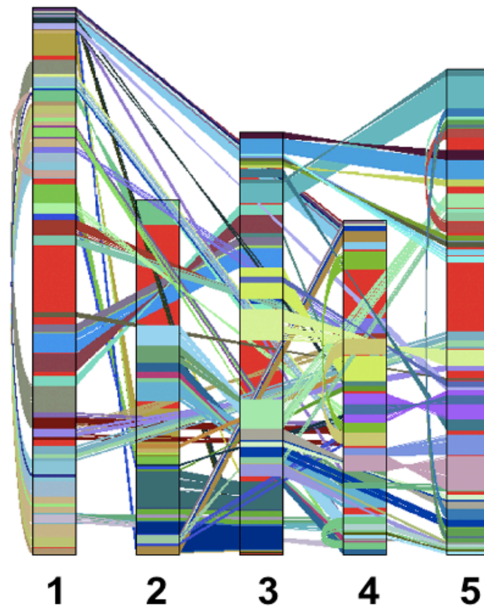
EVOLUTION OF PLANT RNA POLYMERASE SUBUNITS

The RNA polymerase subunit diversity in *Arabidopsis* is quite unique compared to most eukaryotes. The elucidation of the subunit compositions of Pol I, II, III, IV and V in this thesis has been a major step toward understanding how these subunits partition into different polymerases and now provides a framework for testing the functions of specific subunits within each polymerase. However, the identification of these subunits has posed more questions than answers at this point.

A major question that remains is how and why did *Arabidopsis* evolve so many subunits? This question undoubtedly has multiple answers. First, the evolution of Pol IV and Pol V has contributed to the functional diversification of these subunits. Second, duplication of the *Arabidopsis* genome has also been a factor. Many parts of the *Arabidopsis* genome have undergone duplication over time (Figure 5A) (http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml). Sequencing of the genome has allowed researchers to identify regions of synteny between chromosomes and on the same chromosome. Analysis of all 50+ confirmed and predicted *Arabidopsis* RNA polymerase subunits reveals that at least four new subunits were created during duplications of large regions of the genome, namely either NRPD7 or NRPE7, NRPB10 or At1g61700, NRPE9a or NRPE9b and NRPE5 or At2g41340

(Figure 5B). Given that At1g61700 nor At2g41340 have been assigned a function, I hypothesize that these subunits were created upon a duplication. Because NRPD7 and NRPE7 as well as NRPE9a and NRPE9b associate into at least one polymerase, it is harder to determine which subunit came first.

A. Duplicated segments of the *Arabidopsis* genome



B. RNA polymerase subunits that evolved by large-scale gene duplications

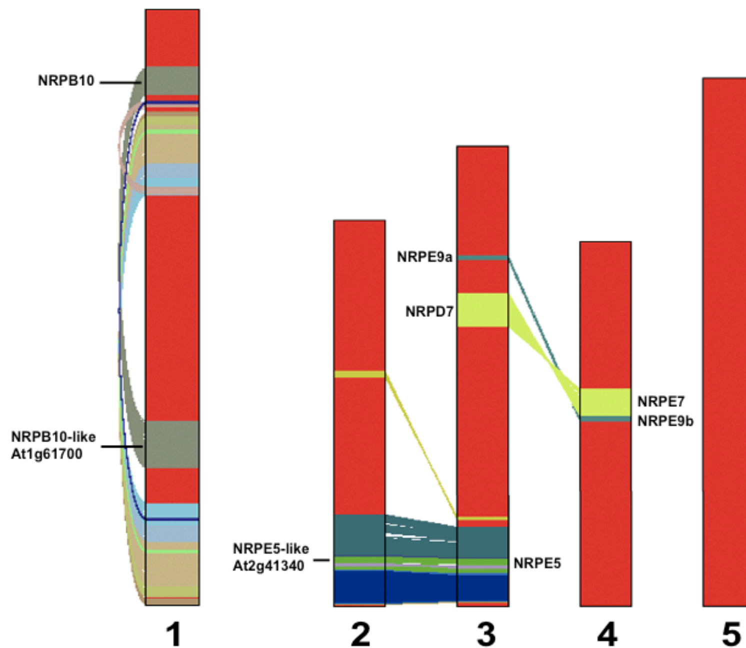


Figure 5. RNA polymerase subunits arose by duplication of the *Arabidopsis* genome. A. Regions of synteny in the *Arabidopsis* genome. B. RNA polymerase subunits that are located in regions of synteny. Figures adapted from The *Arabidopsis* Information Resource website: (http://www.tigr.org/tdb/e2k1/ath1/Arabidopsis_genome_duplication.shtml).

In contrast to large segmental duplications, several subunits might have evolved by local duplication of the neighboring gene, resulting in two genes in close proximity to each other. At2g15400 and At2g15430 encode NRPE3b and NRPE3a, respectively, and are separated by only 20 kb on chromosome 2. At1g60620 and At1g60850 encode NRPC3 and NRPA3, respectively, and are separated by 75 kb on chromosome 1. In both cases, these subunits correspond to the α subunit in *E. coli*. The NRPB7-like family member, At4g14520, is more similar to NRPE7 and NRPD7 than NRPB7. Interestingly, this gene is only 65 kb from NRPE7 on chromosome 4.

In summary, this evidence suggests that duplication of subunits by different mechanisms could have contributed to subunit diversification in *Arabidopsis*. This partially explains why the NRPB5-like family of *Arabidopsis* is so large and why some subunits that are normally encoded by single genes in yeast are encoded by two genes in *Arabidopsis* (i.e. Rpac40-related, NRPB3, NRPB10 and NRPB9 families). It will be interesting to mine the data of other plant genomes as they become available to see whether duplications in other plant genomes led to diversification of the same or different sets of RNA polymerase subunit families, and if such duplications have contributed to functional diversification.

While the duplication mechanisms above explain some of the diversification of RNA polymerase subunits, they do not obviously account for why there are two classes of NRPB5-like proteins, the NRPE5 sub-family and the NRPB5 sub-family. To understand this diversification, data mining of other plant genomes is necessary to determine when the NRPE5 family evolved from the NRPB5 family. Similar logic can be used for the evolution of the NRPB7-like family, which includes NRPE7 and NRPD7, as well as other RNA polymerase gene families. This approach will be fruitful towards identifying the full complement of RNA polymerase subunits in plants and will contribute to our understanding of the evolution of each RNA polymerase subunit family. In the end, it will allow for modeling of the number and types of steps involved in the evolution of Pol IV and Pol V complexes in different plants. This approach will also reveal if *Arabidopsis* represents a "typical" plant in regards to the numbers of genes that have evolved within each subunit family.

A final remaining question is: Why have Pol IV/V evolved? While there is no simple or clear answer, I speculate on different possibilities. Because Pol IV and Pol V are plant-specific, initially it was thought that they might be involved in a plant-specific process. While this is not excluded, it is clear that Pol IV and Pol V have roles in gene silencing pathways that are conserved in non-plant eukaryotes. Non-plant eukaryotes use Pol II as a silencing polymerase and a housekeeping polymerase, whereas plants have apparently separated these functions among Pol II, Pol IV and Pol V, which are structurally related.

In yeast, transcription of silenced heterochromatin occurs during S phase, when chromatin is more conducive toward transcription. It is possible that Pol IV and Pol V

evolved so that transcription of heterochromatin could take place at any stage of the cell cycle. This might have particular implications for plants where quick changes in gene expression or gene silencing must take place in order to cope with a stress. This is particularly relevant in mature parts of the plant where cell division is limited but cells must still respond to changing environmental stimuli.

Over time, the fast evolution rates of Pol IV and Pol V subunits relative to Pol II subunits may be important to keep pace with the fast-evolution rates of genomic parasites, which are constantly on the offensive in evading host silencing mechanisms (Luo and Hall 2007). This faster evolution may reflect a "work in progress" scenario, where Pol IV and Pol V are still fine-tuning their structures and their mechanisms of transcription-mediated silencing, whereas Pol II has had more evolutionary time to become efficient at its role.

Plants are sessile organisms that must cope with the microenvironment in which they grow, in contrast to animals which are able to move, and therefore plants may have evolved a more elaborate and functionally diversified means of gene silencing by utilizing a distinct class of RNA polymerases devoted to silencing.

References cited

Armache, K.J., Mitterweger, S., Meinhart, A., and Cramer, P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* **280**(8): 7131-7134.

Bies-Etheve, N., Pontier, D., Lahmy, S., Picart, C., Vega, D., Cooke, R., and Lagrange, T. 2009. RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family. *EMBO Rep*.

Breant, B., Huet, J., Sentenac, A., and Fromageot, P. 1983. Analysis of yeast RNA polymerases with subunit-specific antibodies. *J Biol Chem* **258**(19): 11968-11973.

Brosnan, C.A., Mitter, N., Christie, M., Smith, N.A., Waterhouse, P.M., and Carroll, B.J. 2007. Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis. *Proc Natl Acad Sci U S A* **104**(37): 14741-14746.

Buhler, J.M., Huet, J., Davies, K.E., Sentenac, A., and Fromageot, P. 1980. Immunological studies of yeast nuclear RNA polymerases at the subunit level. *J Biol Chem* **255**(20): 9949-9954.

Bushnell, D.A., Cramer, P., and Kornberg, R.D. 2002. Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution. *Proc Natl Acad Sci U S A* **99**(3): 1218-1222.

Chen, C.Y., Chang, C.C., Yen, C.F., Chiu, M.T., and Chang, W.H. 2009. Mapping RNA exit channel on transcribing RNA polymerase II by FRET analysis. *Proc Natl Acad Sci U S A* **106**(1): 127-132.

Choder, M. 2004. Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem Sci* **29**(12): 674-681.

Choder, M. and Young, R.A. 1993. A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol Cell Biol* **13**(11): 6984-6991.

Colmenares, S.U., Buker, S.M., Buhler, M., Dlakic, M., and Moazed, D. 2007. Coupling of double-stranded RNA synthesis and siRNA generation in fission yeast RNAi. *Mol Cell* **27**(3): 449-461.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**(5523): 1863-1876.

Crotti, L.B. and Basrai, M.A. 2004. Functional roles for evolutionarily conserved Spt4p at centromeres and heterochromatin in Saccharomyces cerevisiae. *EMBO J* **23**(8): 1804-1814.

Curaba, J. and Chen, X. 2007. Biochemical activities of Arabidopsis RNA-dependent RNA polymerase 6. *J Biol Chem*.

Davis, J.A., Takagi, Y., Kornberg, R.D., and Asturias, F.A. 2002. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Mol Cell* **10**(2): 409-415.

Dequard-Chablat, M., Riva, M., Carles, C., and Sentenac, A. 1991. RPC19, the gene for a subunit common to yeast RNA polymerases A (I) and C (III). *J Biol Chem* **266**(23): 15300-15307.

Devaux, S., Kelly, S., Lecordier, L., Wickstead, B., Perez-Morga, D., Pays, E., Vanhamme, L., and Gull, K. 2007. Diversification of function by different isoforms of conventionally shared RNA polymerase subunits. *Mol Biol Cell* **18**(4): 1293-1301.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C., and Ekwall, K. 2005. RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev* **19**(19): 2301-2306.

Dunoyer, P., Himber, C., Ruiz-Ferrer, V., Alioua, A., and Voinnet, O. 2007. Intra- and intercellular RNA interference in Arabidopsis thaliana requires components of the microRNA and heterochromatic silencing pathways. *Nat Genet* **39**(7): 848-856.

Edwards, A.M., Kane, C.M., Young, R.A., and Kornberg, R.D. 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *J Biol Chem* **266**(1): 71-75.

El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.A., Jacobsen, S.E., Cooke, R., and Lagrange, T. 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**(20): 2539-2544.

Erhard, K.F., Jr., Stonaker, J.L., Parkinson, S.E., Lim, J.P., Hale, C.J., and Hollick, J.B. 2009. RNA polymerase IV functions in paramutation in Zea mays. *Science* **323**(5918): 1201-1205.

Fujioka, Y., Utsumi, M., Ohba, Y., and Watanabe, Y. 2007. Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol* **48**(9): 1243-1253.

Gallant, P. 2007. Control of transcription by Pontin and Reptin. *Trends Cell Biol* **17**(4): 187-192.

Goler-Baron, V., Selitrennik, M., Barkai, O., Haimovich, G., Lotan, R., and Choder, M. 2008. Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes. *Genes Dev* **22**(15): 2022-2027.

Hartzog, G.A., Wada, T., Handa, H., and Winston, F. 1998. Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev* **12**(3): 357-369.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. 2009. NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. *Genes Dev* **23**(3): 318-330.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* **308**(5718): 118-120.

Holt, B.F., 3rd, Boyes, D.C., Ellerstrom, M., Siefers, N., Wiig, A., Kauffman, S., Grant, M.R., and Dangl, J.L. 2002. An evolutionarily conserved mediator of plant disease resistance gene function is required for normal Arabidopsis development. *Dev Cell* **2**(6): 807-817.

Huang, L., Jones, A.M., Searle, I., Patel, K., Vogler, H., Hubner, N.C., and Baulcombe, D.C. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* **16**(1): 91-93.
Huet, J., Riva, M., Sentenac, A., and Fromageot, P. 1985. Yeast RNA polymerase C and its subunits. Specific antibodies as structural and functional probes. *J Biol Chem* **260**(28): 15304-15310.
Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A.J., and Matzke, M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *Embo J*.
Huettel, B., Kanno, T., Daxinger, L., Bucher, E., van der Winden, J., Matzke, A.J., and Matzke, M. 2007. RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim Biophys Acta* **1769**(5-6): 358-374.
Imazawa, Y., Hisatake, K., Mitsuzawa, H., Matsumoto, M., Tsukui, T., Nakagawa, K., Nakadai, T., Shimada, M., Ishihama, A., and Nogi, Y. 2005. The fission yeast protein Ker1p is an ortholog of RNA polymerase I subunit A14 in Saccharomyces cerevisiae and is required for stable association of Rrn3p and RPA21 in RNA polymerase I. *J Biol Chem* **280**(12): 11467-11474.
Jasiak, A.J., Armache, K.J., Martens, B., Jansen, R.P., and Cramer, P. 2006. Structural biology of RNA polymerase III: subcomplex C17/25 X-ray structure and 11 subunit enzyme model. *Mol Cell* **23**(1): 71-81.
Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**(7): 761-765.
Kolodziej, P. and Young, R.A. 1989. RNA polymerase II subunit RPB3 is an essential component of the mRNA transcription apparatus. *Mol Cell Biol* **9**(12): 5387-5394.
Kuhn, C.D., Geiger, S.R., Baumli, S., Gartmann, M., Gerber, J., Jennebach, S., Mielke, T., Tschochner, H., Beckmann, R., and Cramer, P. 2007. Functional architecture of RNA polymerase I. *Cell* **131**(7): 1260-1272.
Lahmy, S., Pontier, D., Cavel, E., Vega, D., El-Shami, M., Kanno, T., and Lagrange, T. 2009. PolV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *Proc Natl Acad Sci U S A* **106**(3): 941-946.
Larkin, R.M. and Guilfoyle, T.J. 1998. Two small subunits in Arabidopsis RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J Biol Chem* **273**(10): 5631-5637.
Larkin, R.M., Hagen, G., and Guilfoyle, T.J. 1999. Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5. *Gene* **231**(1-2): 41-47.
Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. 2006. An ARGONAUTE4-Containing Nuclear Processing Center Colocalized with Cajal Bodies in Arabidopsis thaliana. *Cell* **126**(1): 93-106.
Lin, Y., Nomura, T., Cheong, J., Dorjsuren, D., Iida, K., and Murakami, S. 1997. Hepatitis B virus X protein is a transcriptional modulator that communicates with transcription factor IIB and the RNA polymerase II subunit 5. *J Biol Chem* **272**(11): 7132-7139.

Lotan, R., Bar-On, V.G., Harel-Sharvit, L., Duek, L., Melamed, D., and Choder, M. 2005. The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes Dev* **19**(24): 3004-3016.
Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., and Choder, M. 2007. The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *J Cell Biol* **178**(7): 1133-1143.
Luo, J. and Hall, B.D. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* **64**(1): 101-112.
Maison, C., Bailly, D., Peters, A.H., Quivy, J.P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., and Almouzni, G. 2002. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat Genet* **30**(3): 329-334.
Makino, Y., Yogosawa, S., Kayukawa, K., Coin, F., Egly, J.M., Wang, Z., Roeder, R.G., Yamamoto, K., Muramatsu, M., and Tamura, T. 1999. TATA-Binding protein-interacting protein 120, TIP120, stimulates three classes of eukaryotic transcription via a unique mechanism. *Mol Cell Biol* **19**(12): 7951-7960.
Mann, C., Buhler, J.M., Treich, I., and Sentenac, A. 1987. RPC40, a unique gene for a subunit shared between yeast RNA polymerases A and C. *Cell* **48**(4): 627-637.
McKune, K., Moore, P.A., Hull, M.W., and Woychik, N.A. 1995. Six human RNA polymerase subunits functionally substitute for their yeast counterparts. *Mol Cell Biol* **15**(12): 6895-6900.
Miyao, T. and Woychik, N.A. 1998. RNA polymerase subunit RPB5 plays a role in transcriptional activation. *Proc Natl Acad Sci U S A* **95**(26): 15281-15286.
Mosher, R.A., Schwach, F., Studholme, D., and Baulcombe, D.C. 2008. PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc Natl Acad Sci U S A* **105**(8): 3145-3150.
Onodera, Y., Haag, J.R., Ream, T., Nunes, P.C., Pontes, O., and Pikaard, C.S. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**(5): 613-622.
Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. 2008. Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* **180**(1): 207-218.
Orlicky, S.M., Tran, P.T., Sayre, M.H., and Edwards, A.M. 2001. Dissociable Rpb4-Rpb7 subassembly of rna polymerase II binds to single-strand nucleic acid and mediates a post-recruitment step in transcription initiation. *J Biol Chem* **276**(13): 10097-10102.
Peyroche, G., Milkereit, P., Bischler, N., Tschochner, H., Schultz, P., Sentenac, A., Carles, C., and Riva, M. 2000. The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J* **19**(20): 5473-5482.
Pontes, O., Li, C.F., Nunes, P.C., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. 2006. The Arabidopsis Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center. *Cell* **126**(1): 79-92.
Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. 2005. Reinforcement of silencing at

transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. *Genes Dev* **19**(17): 2030-2040.

Qi, D., Jin, H., Lilja, T., and Mannervik, M. 2006. Drosophila Reptin and other TIP60 complex components promote generation of silent chromatin. *Genetics* **174**(1): 241-251.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A.D., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. 2009. Subunit compositions of the RNA-silencing enzymes Pol IV and Pol V reveal their origins as specialized forms of RNA polymerase II. *Mol Cell* **33**(2): 192-203.

Rosenheck, S. and Choder, M. 1998. Rpb4, a subunit of RNA polymerase II, enables the enzyme to transcribe at temperature extremes in vitro. *J Bacteriol* **180**(23): 6187-6192.

Runner, V.M., Podolny, V., and Buratowski, S. 2008. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol Cell Biol* **28**(6): 1883-1891.

Saez-Vasquez, J. and Pikaard, C.S. 1997. Extensive purification of a putative RNA polymerase I holoenzyme from plants that accurately initiates rRNA gene transcription in vitro. *Proc Natl Acad Sci U S A* **94**(22): 11869-11874.

Sakurai, H., Mitsuzawa, H., Kimura, M., and Ishihama, A. 1999. The Rpb4 subunit of fission yeast Schizosaccharomyces pombe RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol Cell Biol* **19**(11): 7511-7518.

Sampath, V. and Sadhale, P. 2005. Rpb4 and Rpb7: a sub-complex integral to multi-subunit RNA polymerases performs a multitude of functions. *IUBMB Life* **57**(2): 93-102.

Selitrennik, M., Duek, L., Lotan, R., and Choder, M. 2006. Nucleocytoplasmic shuttling of the Rpb4p and Rpb7p subunits of Saccharomyces cerevisiae RNA polymerase II by two pathways. *Eukaryot Cell* **5**(12): 2092-2103.

Sheffer, A., Varon, M., and Choder, M. 1999. Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. *Mol Cell Biol* **19**(4): 2672-2680.

Sheth, U. and Parker, R. 2003. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**(5620): 805-808.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. 2004. Elongation by RNA polymerase II: the short and long of it. *Genes Dev* **18**(20): 2437-2468.

Smith, L.M., Pontes, O., Searle, I., Yelina, N., Yousafzai, F.K., Herr, A.J., Pikaard, C.S., and Baulcombe, D.C. 2007. An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in Arabidopsis. *Plant Cell* **19**(5): 1507-1521.

Song, L., Han, M.H., Lesicka, J., and Fedoroff, N. 2007. Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A* **104**(13): 5437-5442.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. 2002. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. *EMBO J* **21**(23): 6549-6559.

Soutourina, J., Bordas-Le Floch, V., Gendrel, G., Flores, A., Ducrot, C., Dumay-Odelot, H., Soularue, P., Navarro, F., Cairns, B.R., Lefebvre, O., and Werner, M. 2006. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol Cell Biol* **26**(13): 4920-4933.

Tardiff, D.F., Abruzzi, K.C., and Rosbash, M. 2007. Protein characterization of Saccharomyces cerevisiae RNA polymerase II after in vivo cross-linking. *Proc Natl Acad Sci U S A* **104**(50): 19948-19953.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H., and Lamond, A. 2008. Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J Cell Biol* **183**(2): 223-239.

Ujvari, A. and Luse, D.S. 2006. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* **13**(1): 49-54.

Venteicher, A.S., Meng, Z., Mason, P.J., Veenstra, T.D., and Artandi, S.E. 2008. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* **132**(6): 945-957.

Vermeulen, M., Hubner, N.C., and Mann, M. 2008. High confidence determination of specific protein-protein interactions using quantitative mass spectrometry. *Curr Opin Biotechnol* **19**(4): 331-337.

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G.A., Winston, F., Buratowski, S., and Handa, H. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12**(3): 343-356.

Werner, F. and Weinzierl, R.O. 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* **10**(3): 635-646.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**(4): 635-648.

Wierzbicki, A.T., Ream, T.S., Haag, J.R., and Pikaard, C.S. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* **41**(5): 630-634.

Woychik, N.A., Lane, W.S., and Young, R.A. 1991. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* **266**(28): 19053-19055.

Woychik, N.A., Liao, S.M., Kolodziej, P.A., and Young, R.A. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* **4**(3): 313-323.

Woychik, N.A., McKune, K., Lane, W.S., and Young, R.A. 1993. Yeast RNA polymerase II subunit RPB11 is related to a subunit shared by RNA polymerase I and III. *Gene Expr* **3**(1): 77-82.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. 2004. Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr Biol* **14**(13): 1214-1220.

APPENDIX A

ROLES OF RNA POLYMERASE IV IN GENE SILENCING

A review published in *Trends in Plant Science* vol. 13;7 390-397

My contributions to this work:

I wrote the section on “The role of Pol IV in abiotic and biotic stress-inducible siRNA production”. I also helped with discussing the scope of the review and in editing the manuscript.

Review

Special Issue: Noncoding and small RNAs

Roles of RNA polymerase IV in gene silencing

Craig S. Pikaard, Jeremy R. Haag, Thomas Ream and Andrzej T. Wierzbicki

Department of Biology, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA

Eukaryotes typically have three multi-subunit enzymes that decode the nuclear genome into RNA: DNA-dependent RNA polymerases I, II and III (Pol I, II and III). Remarkably, higher plants have five multi-subunit nuclear RNA polymerases: the ubiquitous Pol I, II and III, which are essential for viability; plus two non-essential polymerases, Pol IVa and Pol IVb, which specialize in small RNA-mediated gene silencing pathways. There are numerous examples of phenomena that require Pol IVa and/or Pol IVb, including RNA-directed DNA methylation of endogenous repetitive elements, silencing of transgenes, regulation of flowering-time genes, inducible regulation of adjacent gene pairs, and spreading of mobile silencing signals. Although biochemical details concerning Pol IV enzymatic activities are lacking, genetic evidence suggests several alternative models for how Pol IV might function.

RNA polymerases IVa and IVb: non-essential polymerases devoted to gene silencing

In all eukaryotes, DNA-dependent RNA polymerases (Pol) I, II and III transcribe essential genes, including rRNAs, mRNAs and tRNAs (see Glossary for abbreviations used in the article). Pol I, II and III are complicated enzymes with 12–17 subunits, which include structural and functional homologs of the five bacterial RNAP subunits [1]. The largest and second-largest Pol subunits, the homologs of bacterial β' and β, interact to form the DNA entry and RNA exit channels in addition to the catalytic center of RNA synthesis (Figure 1a) [2].

At present, the catalytic subunits homologous to those depicted in Figure 1a are the only known Pol IVa and Pol IVb subunits in *Arabidopsis*, a species discussed throughout this review. These subunits were initially identified by C.S. Pikaard, who examined the newly sequenced *Arabidopsis* genome and found two genes comprising an atypical fourth class of polymerase largest subunits, and two genes for an atypical class of second-largest subunits. His collaborator J. Eisen (Institute for Genomic Research, Rockville, MD) confirmed that these putative subunits are founding members of novel plant-specific clades [3] (see also [4–6]). As with the Pol I, II and III subunits, the atypical subunits are nuclear proteins [4,7,8], representing a new class of polymerase that has been designated nuclear RNA polymerase IV (Pol IV) [4,5].

Glossary

AGO: ARGONAUTE, proteins in this family bind to small RNAs, including siRNAs and miRNAs, and are capable of cleaving RNAs complementary to the small RNAs, a process known as slicing.
CLSY1: CLASSY1, a putative chromatin remodeling protein involved in RNA-directed DNA methylation.
CTD: C-terminal domain.
DCL1: *Arabidopsis* DICER-LIKE 1, involved primarily in miRNA biogenesis.
DCL2: *Arabidopsis* DICER-LIKE 2, generates 22-nt siRNAs.
DCL3: *Arabidopsis* DICER-LIKE 3, involved in 24-nt siRNA biogenesis.
DCL4: *Arabidopsis* DICER-LIKE 4, generates 21-nt siRNAs.
DeCL: Defective chloroplasts and leaves. Also known as DCL in the literature, which can cause confusion with Dicer-like proteins.
DRD1: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1, a putative chromatin remodeling protein involved in RNA-directed DNA methylation.
DRM2: DOMAINS REARRANGED METHYLYTRANSFERASE 2, the primary *Arabidopsis de novo* DNA methyltransferase.
dsRNA: double-stranded RNA.
GFP: Green fluorescent protein, initially derived from jellyfish.
HEN1: HUA ENHANCER 1; methylates the 2′ hydroxyl groups of siRNA and miRNA 3′-terminal nucleotides.
HST1: HASTY1, an exportin 5 homolog implicated in nuclear export of miRNAs.
HYL1: HYPONASTIC LEAVES 1, a dsRNA-binding protein that interacts with DCL1.
l-siRNA: long siRNA of ~40 nt, as opposed to the predominant 21–24-nt size range.
miRNA: microRNA, small RNAs transcribed from dedicated genes, mediate mRNA cleavage or translational arrest.
nat-siRNA: siRNA derived from natural antisense transcripts derived from adjacent genes.
Pol I: DNA-DEPENDENT RNA POLYMERASE I, synthesizes the precursor for the three largest rRNAs.
Pol II: DNA-DEPENDENT RNA POLYMERASE II, transcribes most genes, including mRNAs and miRNAs.
Pol III: DNA-DEPENDENT RNA POLYMERASE III, mostly transcribes 5S rRNA genes and tRNA genes.
Pol IVa: nuclear RNA polymerase IVa, includes the NRPD1a and NRPD2a subunits.
Pol IVb: nuclear RNA polymerase IVb, includes the NRPD1b and NRPD2a subunits.
RdDM: RNA-directed DNA methylation, one of several gene silencing pathways in the nucleus.
RDR2: RNA-DEPENDENT RNA POLYMERASE 2, required for the biogenesis of 24-nt siRNAs in *Arabidopsis* in the RNA-directed DNA methylation pathway.
RDR6: RNA-DEPENDENT RNA POLYMERASE 6, involved in the ta-siRNA, nat-siRNA, l-siRNA, transgene and viral silencing, and long-distance silencing pathways.
RISC: RNA-induced silencing complex, includes an ARGONAUTE protein and siRNA (siRISC) or miRNA (miRISC).
RNA: Ribonucleic acid.
RNA-FISH: RNA fluorescent *in situ* hybridization, a means for locating specific RNAs.
RNAP: DNA-dependent RNA polymerase.
RNP: ribonucleoprotein, a complex of RNA and proteins.
rRNA: ribososomal RNA, four rRNAs are present in ribosomes.
SDE3: SILENCING DEFECTIVE 3, a putative RNA helicase.
SGS3: SUPPRESSOR OF GENE SILENCING 3, a putative coiled-coil protein.
siRNA: small interfering RNA.

Corresponding author: Pikaard, C.S. (pikaard@biology2.wustl.edu).

1360-1385/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tplants.2008.04.008

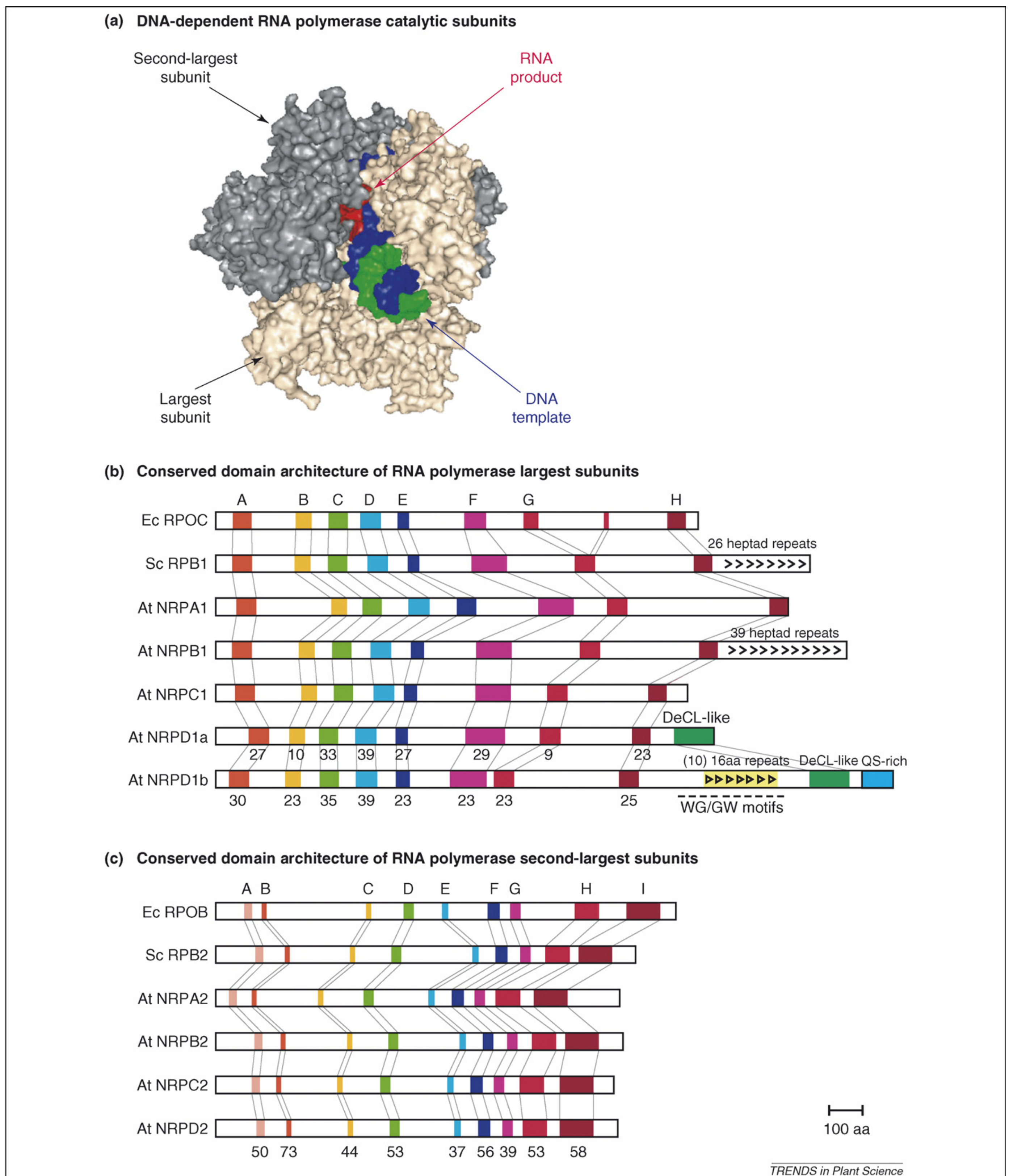


Figure 1. Catalytic subunits of DNA-dependent RNA polymerases. **(a)** The largest and second-largest subunits form the catalytic center. The image is a surface rendering generated using the crystal coordinates for a yeast Pol II elongation complex determined by K. Westover, D. Bushnell and R. Kornberg [PDB: (Protein Data Bank [http://www.rcsb.org/pdb/home/home.do) 1R9T]. Only the two largest Pol II subunits are shown. The DNA template strand is shown in blue, the non-template strand in green, and the nascent RNA in red. **(b)** Domain structures of the largest subunits of RNAP. The largest subunits of *E. coli* (*Ec* RPOC) and yeast Pol II (*Sc* RPB1) are compared with the largest subunits of *Arabidopsis* Pol I (*At* NRPA1), Pol II (*At* NRPB1), Pol III (*At* NRPC1), Pol IVa (*At* NRPD1a) and Pol IVb (*At* NRPD1b). Positions of conserved domains A–H are highlighted. Numbers below Pol IV domains indicate the percentage identities to corresponding *Arabidopsis* Pol II subunit domains. CTDs of the largest subunits of yeast and *Arabidopsis* Pol II have 26 or 39 copies, respectively, of a seven amino acid (heptad) repeat. The domain with similarity to the DEFECTIVE CHLOROPLASTS AND LEAVES protein (DeCL domain), present in the CTDs of the largest subunits of Pol IVa and Pol IVb, is shown in green. The CTD of NRPD1b also includes a region rich in WG–GW motifs, overlapping ten, imperfect, 16-amino-acid repeats, and a domain composed of alternating glutamines and serines (QS-rich domain). **(c)** Domain structures of the second-largest subunits of RNAP. *E. coli* (*Ec* RPOB) and yeast Pol II subunits (*Sc* RPB2) are compared with the second-largest subunits of *Arabidopsis* Pol I (*At* NRPA2), Pol II (*At* NRPB2), Pol III (*At* NRPC2) and Pol IV (*At* NRPD2). Positions of conserved domains A–I are highlighted. Numbers below Pol IV domains are percentage identities to the corresponding *Arabidopsis* Pol II subunit domains.

Box 1. Pol IV subunit nomenclature

Nomenclature for Pol IV subunit genes derives from naming systems used in other eukaryotic model systems (e.g. budding yeast [*Saccharomyces cerevisiae*], in which RNA polymerase I, II and III are designated *RPA*, *RPB* and *RPC*, respectively). In *Arabidopsis*, an N, for 'nuclear', was added (e.g. *NRPA*, *NRPB* etc.) to polymerase subunit gene names to circumvent nomenclature conflicts with unrelated genes. The resulting gene names were registered with The *Arabidopsis* Information Resource by joint request of the David Baulcombe and Craig Pikaard laboratories. Largest subunits that are homologs of bacterial β′ are designated, by convention, with the number 1, such that the unique *Arabidopsis* genes *NRPA1*, *NRPB1* and *NRPC1* encode the largest subunits of Pol I, II and III, respectively. Likewise, the genes encoding the second-largest subunits of *Arabidopsis* Pol I, II and III are designated *NRPA2*, *NRPB2* and *NRPC2*, respectively. On the basis of this naming scheme, the two related, but distinct, Pol IV largest subunits were designated *NRPD1a* and *NRPD1b*. Likewise, the two Pol IV second-largest subunit genes are designated *NRPD2a* and *NRPD2b*. Only *NRPD2a* is functional in the Col-O ecotype of *Arabidopsis* that has been studied to date [4,5,9,10]. Therefore, *NRPD2a* can be referred to simply as *NRPD2*. In other plant species, there are numerous functional genes for both the largest and second-largest subunits of Pol IV.

NRPD1a is the largest subunit of Pol IVa [4,5], whereas NRPD1b is the largest subunit of Pol IVb [9,10] (subunit nomenclature is discussed in Box 1). The largest subunits in both Pol IVa and Pol IVb have C-terminal domains (CTDs) that share similarity with the DEFECTIVE CHLOROPLASTS AND LEAVES protein (abbreviated DeCL in this article), which is required for 4.5S rRNA processing in chloroplasts (Figure 1b) [11]. The CTD of NRPD1b also includes ten imperfect 16-amino-acid repeats within a tryptophan and glycine (WG–GW)-rich region. A glutamine and serine (Q–S)-rich domain is present at the distal end of the CTD (Figure 1b). The WG–GW motifs are proposed to mediate Argonaute protein interactions [8,12], but the significance of the DeCL and Q–S domains is unknown. However, the DeCL and Q–S domains might facilitate additional molecular interactions in a manner analogous to the function of the CTD of the largest subunit of Pol II. This CTD mediates numerous interactions that govern processes such as transcriptional activation by enhancers, transcription elongation, and several mRNA processing steps [13–15]. Both Pol IVa and Pol IVb have an NRPD2 subunit that is encoded by the same gene, *NRPD2a* [4,5,9,10]. NRPD1a and NRPD1b each co-immunoprecipitate and co-localize with NRPD2 [7], but the alternative largest subunits do not immunoprecipitate with one another, indicating that Pol IVa and Pol IVb are distinct physical entities.

The full subunit compositions of Pol IVa and Pol IVb are not known, nor are their templates or enzymatic products. However, a flurry of studies in the past three years has shown that Pol IVa and, to a lesser extent, Pol IVb are crucial for several RNA-mediated gene silencing phenomena. These pathways, and the roles of Pol IV in them, are the focus of our review.

Roles of Pol IVa and Pol IVb in the RNA-directed DNA methylation pathway

Arabidopsis has four Dicer endonucleases (DCLs), six single-subunit RNA-dependent RNA polymerases (RDRs) and ten Argonaute proteins (AGOs) that participate in microRNA (miRNA)- and small interfering (siRNA)-mediated transcriptional or post-transcriptional silencing [16–19]. In the RNA-directed DNA methylation (RdDM) pathway of transcriptional gene silencing [20–23], double-stranded RNAs generated with the involvement of RDR2 are cleaved by DCL3, and the resulting siRNAs are loaded into AGO4–RISC and/or AGO6–RISC complexes that mediate the *de novo* methylation of cytosines within DNA sequences complementary to the siRNAs [22,24–28]. The realization that Pol IVa and Pol IVb are players in the RdDM pathway came from a combination of genetic screens [5,10] and reverse-genetic analyses [4,9]. Silencing-defective (*sde*) mutants were identified in screens for the de-repression of a silenced transgene locus, and analysis of these mutants led to the identification of *sde4* as an allele of *NRPD1a* [5]. A subsequent test to determine if one of the atypical second-largest subunit (*NRPD2*) genes might partner with *NRPD1a* revealed that insertional mutants of *NRPD2a* also disrupted the silencing pathway. Coinciding with this disruption was the disappearance of 24-nt siRNAs and the loss of cytosine methylation at corresponding loci [5]. Our laboratory initially focused on *NRPD2*, showing that its activity was not redundant with that of the equivalent Pol I, II or III subunits and that it did not co-purify with Pol I, II or III [4]. However, NRPD2 was found to localize within the nucleus and to affect the coalescence of heterochromatic sequences into chromocenters [4]. Heterochromatic DNA is typically heavily methylated, and loss of cytosine methylation occurred at a subset of heterochromatic loci in *nrpd2* mutants as well as in *nrpd1a* mutants [4]. Collectively, the initial studies of NRPD1a and NRPD2 pointed to the existence of Pol IVa.

Kanno *et al*. [29] carried out a genetic screen for mutations causing the de-repression of a reporter gene silenced by RdDM. This led to the identification of *DRD1*, a member of the SWI2–SNF2 chromatin remodeling protein family, in addition to *DRD2* and *DRD3,* which turned out to be *NRPD2a* and *NRPD1b*, respectively [10]. The realization that the *NRPD1b* gene had been mistakenly annotated as two genes [4,5,10] also led to a reverse-genetic examination of cytosine methylation and siRNA phenotypes in *nrpd1b* insertional mutants [9]. Collectively, these independent studies revealed the existence of Pol IVb and showed that siRNAs eliminated in Pol IVa mutants [4,5] are not abolished in Pol IVb mutants [9,10], despite similar losses of cytosine methylation [9,10]. These observations, based on a small number of loci, indicated that Pol IVa and Pol IVb act at different steps in the RdDM pathway, with Pol IVa acting upstream of siRNA production, and Pol IVb functioning at a later step in the pathway, mostly downstream of siRNA production [10]. Recent genome-wide analyses of small RNA populations have shown that there are at least 4600 *Arabidopsis* loci that give rise to small RNAs, with 94% of them being dependent on Pol IVa [30]. Pol IVb plays little, if any, role in siRNA abundance at approximately one-third of these loci; it has intermediate effects at another one-third of the loci; and it is absolutely required for siRNA production at one-third of the Pol IVa-dependent loci [30]. However, there are no definitive examples of siRNAs that

are dependent on Pol IVb only, and which do not require Pol IVa. These results are consistent with the hypothesis that Pol IVa acts upstream of siRNA production. The role of Pol IVb in siRNA production is less clear, and it could be indirect. A positive feedback relationship exists between the formation of heterochromatin and the continued production of siRNA. As such, the role of Pol IVb in facilitating RdDM might explain the influence of Pol IVb on siRNA abundance, as has been depicted in circular models for the RdDM pathway [7,8].

The localization of proteins involved in RdDM has provided insight into the RdDM pathway [7,8,31,32]. Pol IVa, Pol IVb and DRD1 co-localize with chromosomal loci that are both sources and targets of abundant siRNAs, suggesting that these proteins are involved in the generation of siRNA precursors or the targeting of siRNA-directed chromatin modifications [7]. AGO4 and DRM2, the primary *de novo* DNA methyltransferase, also co-localize at source/target loci in some nuclei [32]. RNA-FISH combined with protein immunolocalization has shown that siRNAs co-localize with RDR2, DCL3, AGO4 and NRPD1b within a nucleolar compartment interpreted to be an siRNA processing center [7]. This processing center includes several molecular markers of Cajal bodies [8], which are dynamic compartments important for assembling ribonucleoprotein complexes involved in pre-mRNA splicing, pre-rRNA processing, RNA methylation and pseudo-uridylation, telomerase assembly and histone mRNA 3′ end formation [33,34]. Formation of siRNA–RISC complexes is consistent with the overall theme of assembling ribonucleoprotein complexes within Cajal bodies [8,33–35]. Recent evidence suggests that miRNA processing in plants also occurs within nucleolus-associated Cajal body-like entities that include the spliceosomal proteins SmB and SmD3 – both found in Cajal bodies and spliceosomes – but which lack the canonical Cajal body protein coilin [36]. Other groups have suggested that these miRNA processing centers are not Cajal bodies, because they lack coilin [37,38]. However, *Drosophila* lacks coilin yet has functional Cajal bodies [39]. These observations can be reconciled by the hypothesis that there are numerous sub-classes of Cajal bodies, some of which have coilin and some of which do not [34,35,39].

Because Pol IVa co-localizes with loci that give rise to abundant 24-nt siRNAs and because loss of *NRPD1a* function causes all other known components of the RdDM pathway to mislocalize, Pol IV is thought to act at an initial step of the pathway, upstream of RDR2 [7]. CLSY1, which like DRD1 is an SWI–SNF family protein, co-localizes with RDR2 at the inner perimeter of the nucleolus; and, in *clsy1* mutants, RDR2 localization is severely disrupted [40]. Pol IVa localization is also affected, albeit to a lesser degree [40], suggesting that CLSY1 functions at the interface between Pol IVa and RDR2, presumably facilitating the generation of dsRNAs that are diced by DCL3 and loaded into AGO4 effector complexes [16,17,26,41] within the nucleolar siRNA processing center [7,8]. NRPD1b co-localizes with AGO4 both within the processing center [7,8] and at target loci [32], interacting with AGO4 through the CTD [8,12]. Current models suggest that siRNA–AGO4–Pol IVb effector complexes then locate their targets by virtue of siRNA-target base-pairing interactions [7,8]. Pol IVb, DRD1 and DRM2 are then thought to collaborate in the siRISC-directed DNA methylation process through an as yet unknown mechanism [21]. DNA methylation then appears to feed back on the production of siRNAs, such that siRNAs are depleted in *drm* mutants at some loci [4,7,41] and in *ddm1* (*decrease in DNA methylation 1*) or *met1* (*cytosine methyltransferase 1*) mutants that are required for maintaining DNA methylation patterns at other loci [42]. Therefore, it is possible that Pol IVa preferentially transcribes methylated DNA [4] or aberrant RNAs generated from methylated loci [7,43,44] as a means of perpetuating the repression cycle.

A role for Pol IV in flowering

Although they are non-essential in terms of viability, Pol IVa and Pol IVb nonetheless play roles in development, affecting flowering time in the context of the RdDM pathway. Under short-day conditions, flowering in *nrpd1a* and *nrpd1b* mutants is significantly delayed, as is also the case in *rdr2*, *dcl3*, *ago4* and *drm* mutants [9,45]. The flowering-time regulators *FCA* and *FPA* were identified in screens for mutants that disrupt RNA-directed gene silencing, and they appear to be players in the RdDM pathway, wherein they act at some, but not all, loci [46]. At least two flowering genes, *FWA* and *FLC*, appear to be targets of silencing through Pol IV-dependent siRNA pathways [45,47,48].

The role of Pol IV in abiotic and biotic stress-inducible siRNA production

Pol IV plays an important role in the production of natural antisense transcript siRNAs (nat-siRNAs) [49–53]. These siRNAs are generated from dsRNAs derived from the overlapping 3′ ends of convergently transcribed gene pairs. Expression of one member of the gene pair is constitutive, but expression of the other is inducible, as in the case of the *P5CDH* and *SRO5* gene pair, respectively. Salt stress induces *SRO5* expression such that its transcript can anneal with the *P5CDH* mRNA to form a region of dsRNA. In a process involving Pol IVa, RDR6, SGS3 and DCL2, a 24-nt nat-siRNA is produced, and this is thought to guide the cleavage of *P5CDH* transcripts, setting the stage for generation of additional DCL1-dependent 21-nt siRNAs [49]. The resulting downregulation of *P5CDH* results in increased proline synthesis, a physiological response that helps to confer salt tolerance.

Pathogen-inducible siRNAs provide two examples of additional means for generating nat-siRNAs [54,55]. In the first, infection of *Arabidopsis* with *Pseudomonas syringae* generates a 22-nt nat-siRNA in a pathway that requires Pol IVa, RDR6 and SGS3. This pathway is similar to that which generates the salt stress-induced nat-siRNA, except that DCL2 is not involved; instead, DCL1, HYL1 and HEN1 – which are typically involved in miRNA biogenesis – are required for siRNA production in the pathogen response. The end result is the downregulation of *PPRL*, a negative regulator of pathogen resistance. More recently, investigators demonstrated that *Pseudomonas syringae* infection induces expression of a 39–41-nt RNA [54]. This so-called long siRNA (l-siRNA) matches the overlapping region of the *SRRLK* and *AtRAP* gene pair,

and it specifically downregulates *AtRAP*, another negative regulator of the pathogen defense response, in a pathway requiring Pol IVa and Pol IVb, DCL1, HYL1, HEN1, HST1 (HASTY1), RDR6, DCL4, AGO7 and SDE3. Most of these proteins (i.e. DCL1, HYL1, HEN1, HST1, RDR6, DCL4 and AGO7) are also players in the so-called *trans*-acting siRNA (ta-siRNA) pathway, in which miRNA-mediated cleavage of a specific target mRNA initiates the subsequent production of siRNAs from the cleaved mRNA [56–59]. Resulting siRNAs then target additional mRNAs for cleavage, thereby amplifying the signal in a regulatory cascade. It is not yet clear whether a similar regulatory cascade occurs upon bacterial infection and, if so, where Pol IVa and Pol IVb fit within such a pathway.

Roles of Pol IV in the spreading of silencing

Pol IVa is required for both short-range spreading of RNA silencing cell-to-cell through plasmadesmata and long-range silencing through the phloem [60,61]. Two independent screens revealed a requirement for Pol IVa and RDR2 in the short-range spreading of silencing [40,62], and DCL4 [40,63], DCL1, HEN1 and AGO1 [62] are also required. By contrast, HYL1, DCL3, AGO4, RDR6 [40,62,63], Pol IVb (NRPD1b) and DRD1 [40] are all dispensable. Although both 24-nt and 21-nt transgene-specific siRNAs are produced, the DCL4-dependent 21-nt siRNAs are believed to be the primary short-range mobile signals [40,62,63]. However, longer siRNAs can suffice when overproduced in mutants of *DRB4,* which encodes a dsRNA binding protein that partners with DCL4 in the production of 21-nt siRNAs [62].

In Pol IVa mutants, silencing is impaired even in the phloem cells where the silencing signal is initiated, suggesting that Pol IVa acts at an initiating step in the process that ultimately gives rise to the mobile silencing signal(s). Interestingly, the spreading of silencing can be dramatically enhanced in *dcl3* and *ago4* mutants [40], coincident with increased 21-nt siRNA production and loss of 24-nt siRNAs. A possibility is that Pol IVa/RDR2-dependent dsRNA substrates can be channeled into either 24-nt or 21-nt siRNA production, with the 21-nt siRNAs acting as the primary short-range mobile signals.

An ability to distinguish between production and perception of silencing signals has come from a study in which wild type or mutant rootstocks or scions (shoots) were grafted onto one another and monitored for long-distance silencing of a green fluorescent protein (GFP) transgene [64]. Pol IVa (NRPD1a), RDR2, DCL3, AGO4 and RDR6 are all required for the scion to respond to a silencing signal derived from a dsRNA hairpin expressed in the rootstock [65]. However, none of these proteins are required to generate the mobile signal. Interestingly, RDR6 is required for the perception of the long-distance signal [65] but is dispensable for short-range silencing [40,62]. Pol IVb (NRPD1b) is dispensable for both short and long-distance silencing, consistent with the hypothesis that Pol IVb functions in chromatin modification rather than in RNA production.

The nature of the long-distance silencing signal is unknown, but *dcl1–8* hypomorphs and *dcl2*;*dcl3*;*dcl4* triple mutants defective for miRNA or siRNA production, respectively, continue to produce the mobile signal in roots, as do mutants for Pol IVa, Pol IVb, *RDR2* and *RDR6* [65]. Therefore, it seems unlikely that Dicer-generated small RNAs are the long-distance signaling molecules. Instead, larger RNAs might serve as the mobile signal(s). An intriguing observation is that siRNAs produced in the scion upon reception of the silencing signal do not correspond to the approximately two-thirds of the GFP gene that was used as the hairpin trigger sequence; instead, the siRNAs neatly correspond to the third of the GFP transgene located downstream (3′) of the trigger sequences [65]. It is not clear why this should be the case if siRNAs are the mobile signal. Antisense siRNAs could anneal anywhere throughout the first two-thirds of the target mRNA and might be expected to prime RDR activity in the upstream direction. Likewise, siRNA-directed cleavage of target mRNAs, which would render the 3′ target fragment uncapped, 'aberrant' and a potential substrate for RDR6 [66], would generate a diverse set of cleaved fragments throughout the first two-thirds of the GFP target. Therefore, a possibility is that the dsRNA trigger molecule itself, or its component strands, is the mobile signal(s), which is plausible given the evidence that intact mRNAs can traffic through phloem [67]. If the antisense strand of the dsRNA trigger were to anneal to the intact mRNA in the shoot such that only the 3′ portion of the GFP mRNA were to remain single-stranded, the resulting structure might somehow direct RDR6- and Pol IVa-dependent amplification of the single-stranded sequences 3′ of the trigger sequence.

Unsolved mysteries and future directions

Pol IVa is integral to numerous RNA silencing pathways, including the RdDM pathway, the nat-siRNA and l-siRNA pathways, the short-range spreading of silencing pathway, and the pathway for the perception of long-distance silencing signals (Figure 2). Pol IVb is apparently less gregarious, acting primarily in the RdDM pathway [30], but also playing an undefined role in the l-siRNA pathway [54]. It seems probable that both Pol IVa and Pol IVb possess enzymatic activity, given that the NRPD1a, NRPD1b and

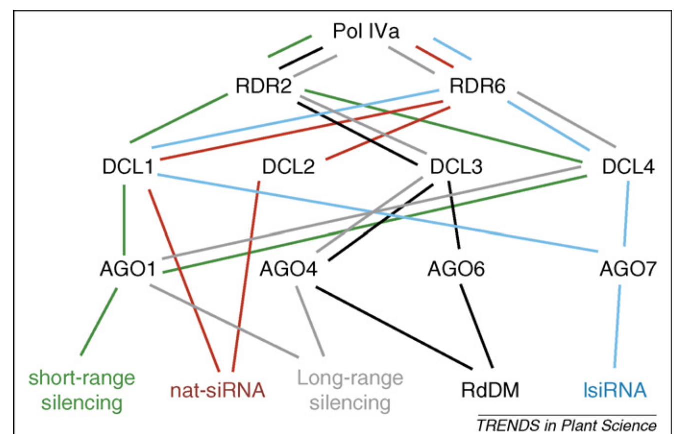


Figure 2. A variety of proteins participate in Pol IVa-dependent silencing pathways. The figure shows a subset of the proteins that are involved in RdDM, nat-siRNA, l-siRNA, short-range silencing, and long-distance silencing pathways. Proteins involved in the various pathways are linked by color-coded lines. The diagram does not imply the order of events, but illustrates the diversity of functional collaborations that are possible. Not all mutants have been tested in every pathway; therefore, other potential connections might exist. However, the figure reflects the models provided by the authors of the studies discussed in the text.

NRPD2 subunits possess the key conserved amino acids of the metal A and metal B sites found within the catalytic centers of other multi-subunit RNA polymerases [68,69]. But what do Pol IVa and Pol IVb transcribe, and what are their products? At present, we have no answer. In fact, our only biochemical clue is a negative result: a conventional, promoter-independent transcription assay [70] using sheared double-stranded template DNA revealed that chromatographic fractions enriched for Pol IV lack DNA-dependent RNA polymerase activity, unlike fractions enriched for Pol I, II and III [4]. Based on this result, it seems likely that Pol IVa and Pol IVb use very specific templates.

A distinct possibility is that Pol IVa transcribes RNA [7,43,44]. Pol IVa is mislocalized by RNase treatment of nuclei, but not by DNase treatment, whereas Pol II shows the opposite nuclease sensitivities [7]. Moreover, there is precedent for DNA-dependent RNA polymerases transcribing RNA. Hepatitis Delta Virus (HDV) and plant viroid RNAs are replicated by Pol II transcription [71,72]. Likewise, *Escherichia coli* RNAP is regulated by binding to 6S RNA, which is transcribed in order to be released [73].

Previous models for the RdDM pathway have suggested that Pol IVa transcribes methylated DNA or transcripts of methylated loci, with resulting Pol IVa transcripts being amplified or made double-stranded by RDR2 (Figures 3ab). However, in the nat-siRNA and l-siRNA pathways, regions of dsRNA are apparently generated by Pol II transcription of overlapping gene pairs, and these transcripts persist in *nrpd1a* mutants, suggesting that there is no need for Pol IVa in the initial formation of dsRNA. Likewise, Pol IVa plays roles in short-range spreading of silencing triggered by dsRNA hairpin trigger sequences, and in long-distance silencing likely to involve annealing of a mobile RNA to target mRNAs, thereby forming dsRNA. In each of these cases, there is no obvious need for Pol IVa in the initial generation of dsRNAs.

Pol IVa might use initial dsRNAs as templates, generating transcripts that are then made double-stranded by RDR2 or RDR6, one or both of which are involved in all known Pol IVa-dependent pathways (Figure 3c). Subsequent dicing, siRNA-mediated target slicing *in trans*, and RDR transcription of sliced templates might then amplify the initial signal and generate small RNAs beyond the region of initial transcript overlap. Alternatively,

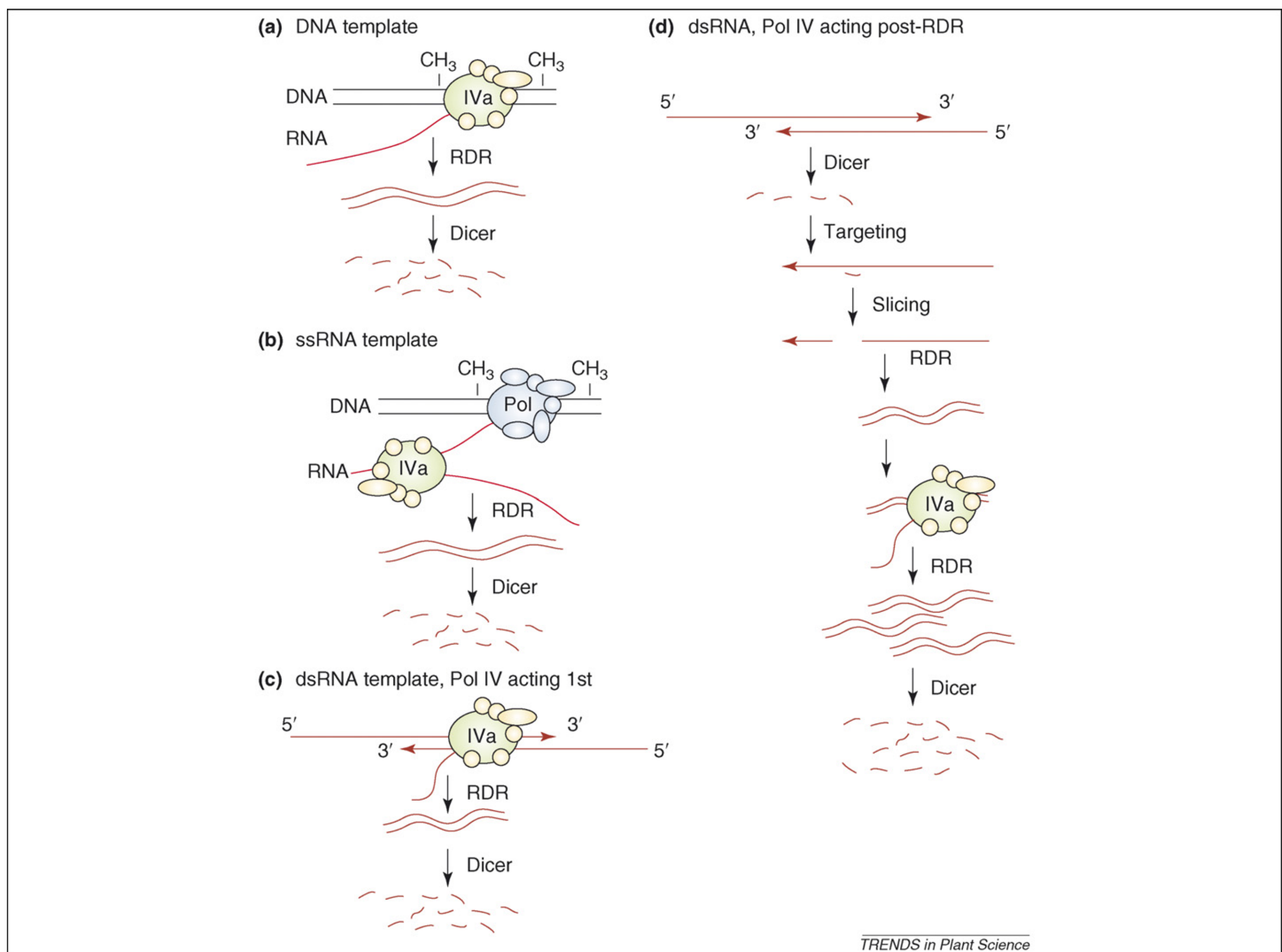


Figure 3. Possible modes of Pol IVa function. Pol IVa might transcribe a specialized DNA template, such as methylated DNA **(a)** or single-stranded RNA transcripts derived from methylated DNA loci **(b)**. Alternatively, Pol IVa might transcribe dsRNA generated from bidirectional transcripts, including transcripts of natural antisense gene pairs, or dsRNAs resulting from the annealing of long-distance mobile RNAs with target mRNAs **(c)** and **(d)**. The model shown in (d) might account for the involvement of numerous Dicer proteins and numerous RDR inputs in the nat-siRNA and long-distance silencing pathways.

dicing of initial dsRNA regions might lead to the production of siRNAs that prime RDR on sliced or unsliced target RNAs, resulting in secondary dsRNAs that are then transcribed by Pol IVa and amplified by further RDR activity (Figure 3d). The model in Figure 3d would account for the involvement of more than one Dicer and more than one RDR-requiring step in the nat-siRNA and long-distance silencing pathways.

Pol IVa appears to be dispensable in some dsRNA-initiated phenomena. For instance, one group [10] screened for methylation-defective mutants by using a dsRNA hairpin to trigger RNA-directed DNA methylation. They recovered nine alleles of *NRPD1b*, and twelve alleles of *NRPD2a*, but no alleles of *NRPD1a* or *RDR2* were identified [10], suggesting that the production of dsRNA hairpins had bypassed a need for Pol IVa or RDR2. Similarly, deep sequencing of small RNA libraries has shown that more than 90% of all siRNAs are Pol IVa-dependent and are mostly derived from transposable elements and tandem repeats [30,74]. Inverted repeats, however, can contribute to the siRNA pool by a Pol IVa-independent mechanism [74]. Because transcription of inverted repeats can produce hairpin dsRNAs on their own, their Pol IVa-independence fits with the idea that Pol IVa functions at other loci in the production of dsRNAs that then feed into siRNA production. Why some dsRNA hairpin-initiated silencing phenomena require Pol IVa, but others do not, is not clear. The strength of the promoters driving hairpin formation might be an important variable.

Pol IVb is even more of a mystery than Pol IVa. NRPD1b mostly appears to reinforce Pol IVa-dependent siRNA production [9,30] yet is required, in addition to Pol IVa, for RdDM [9,10,75]. One possibility is that Pol IVb binds to DNA and interacts with AGO4 through its CTD [8,12], facilitating siRNA–DNA base-pairing, which in turn enables the recruitment of DRM2. Alternatively, siRNA–AGO4 complexes might anneal to Pol IVb transcripts, thereby recruiting DRM2 and/or histone modifying enzymes to the vicinity of the corresponding DNA, as in models for siRNA-mediated silencing in fission yeast (*Schizosaccharomyces pombe*) [76,77]. AGO4 can slice RNAs in an siRNA-guided process, providing evidence that AGO4–siRNA–RISC complexes can interact with RNA transcripts [78]. Nonetheless, direct siRNA interactions with DNA cannot be ruled out.

Clearly, there is much that needs to be learned concerning the templates, products, subunit structures, and interacting partners of Pol IVa and Pol IVb. Development of *in vitro* assays will be invaluable for deciphering the functions of these enigmatic polymerases and is a major challenge for the future.

Acknowledgements

Our work is supported by National Institutes of Health grants R01GM60380 and R01GM077590, and by the Monsanto Company-Washington University Plant Biology Research Agreement.

References

1 Werner, F. (2007) Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* 65, 1395–1404
2 Kettenberger, H. *et al.* (2004) Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* 16, 955–965
3 The-Arabidopsis-Genome-Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
4 Onodera, Y. *et al.* (2005) Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120, 613–622
5 Herr, A.J. *et al.* (2005) RNA polymerase IV directs silencing of endogenous DNA. *Science* 308, 118–120
6 Luo, J. and Hall, B.D. (2007) A multistep process gave rise to RNA polymerase IV of land plants. *J. Mol. Evol.* 64, 101–112
7 Pontes, O. *et al.* (2006) The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* 126, 79–92
8 Li, C.F. *et al.* (2006) An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* 126, 93–106
9 Pontier, D. *et al.* (2005) Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* 19, 2030–2040
10 Kanno, T. *et al.* (2005) Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* 37, 761–765
11 Bellaoui, M. *et al.* (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. *Plant Mol. Biol.* 53, 531–543
12 El-Shami, M. *et al.* (2007) Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* 21, 2539–2544
13 Phatnani, H.P. and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* 20, 2922–2936
14 Hahn, S. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat. Struct. Mol. Biol.* 11, 394–403
15 Shilatifard, A. (2004) Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim. Biophys. Acta* 1677, 79–86
16 Brodersen, P. and Voinnet, O. (2006) The diversity of RNA silencing pathways in plants. *Trends Genet.* 22, 268–280
17 Baulcombe, D. (2004) RNA silencing in plants. *Nature* 431, 356–363
18 Herr, A.J. and Baulcombe, D.C. (2004) RNA silencing pathways in plants. *Cold Spring Harb. Symp. Quant. Biol.* 69, 363–370
19 Chapman, E.J. and Carrington, J.C. (2007) Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* 8, 884–896
20 Wassenegger, M. (2000) RNA-directed DNA methylation. *Plant Mol. Biol.* 43, 203–220
21 Huettel, B. *et al.* (2007) RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochim. Biophys. Acta* 1769, 358–374
22 Wassenegger, M. (2005) The role of the RNAi machinery in heterochromatin formation. *Cell* 122, 13–16
23 Bender, J. (2004) DNA methylation and epigenetics. *Annu. Rev. Plant Biol.* 55, 41–68
24 Cao, X. *et al.* (2003) Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* 13, 2212–2217
25 Zilberman, D. *et al.* (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719
26 Xie, Z. *et al.* (2004) Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* 2, E104
27 Matzke, M. *et al.* (2006) RNA-directed DNA methylation and Pol IVb in *Arabidopsis*. *Cold Spring Harb. Symp. Quant. Biol.* 71, 449–459
28 Zheng, X. *et al.* (2007) Role of *Arabidopsis* AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. *EMBO J.* 26, 1691–1701
29 Kanno, T. *et al.* (2004) Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr. Biol.* 14, 801–805
30 Mosher, R.A. *et al.* (2008) PolIVb influences RNA-directed DNA methylation independently of its role in siRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3145–3150
31 Pikaard, C.S. (2006) Cell biology of the *Arabidopsis* nuclear siRNA pathway for RNA-directed chromatin modification. *Cold Spring Harb. Symp. Quant. Biol.* 71, 473–480
32 Li, C.F. *et al.* (2008) Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in *Arabidopsis thaliana*. *PLoS Genet.* 4, e27

33 Cioce, M. and Lamond, A.I. (2005) Cajal bodies: a long history of discovery. *Annu. Rev. Cell Dev. Biol.* 21, 105–131
34 Handwerger, K.E. and Gall, J.G. (2006) Subnuclear organelles: new insights into form and function. *Trends Cell Biol.* 16, 19–26
35 Pontes, O. and Pikaard, C.S. siRNA and miRNA processing: new roles for Cajal bodies. *Curr. Opin. Gen. Dev*. (in press)
36 Fujioka, Y. *et al.* (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in *Arabidopsis*. *Plant Cell Physiol.* 48, 1243–1253
37 Fang, Y. and Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living *Arabidopsis* plants. *Curr. Biol.* 17, 818–823
38 Song, L. *et al.* (2007) *Arabidopsis* primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5437–5442
39 Liu, J.L. *et al.* (2006) The *Drosophila melanogaster* Cajal body. *J. Cell Biol.* 172, 875–884
40 Smith, L.M. *et al.* (2007) An SNF2 protein associated with nuclear RNA silencing and the spread of a silencing signal between cells in *Arabidopsis*. *Plant Cell* 19, 1507–1521
41 Zilberman, D. *et al.* (2004) Role of *Arabidopsis* ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr. Biol.* 14, 1214–1220
42 Lippman, Z. *et al.* (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* 1, E67
43 Vaughn, M.W. and Martienssen, R.A. (2005) Finding the right template: RNA Pol IV, a plant-specific RNA polymerase. *Mol. Cell* 17, 754–756
44 Vaucheret, H. (2005) RNA polymerase IV and transcriptional silencing. *Nat. Genet.* 37, 659–660
45 Chan, S.W. *et al.* (2004) RNA silencing genes control *de novo* DNA methylation. *Science* 303, 1336
46 Baurle, I. *et al.* (2007) Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. *Science* 318, 109–112
47 Swiezewski, S. *et al.* (2007) Small RNA-mediated chromatin silencing directed to the 3′ region of the *Arabidopsis* gene encoding the developmental regulator, FLC. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3633–3638
48 Liu, J. *et al.* (2004) siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* 18, 2873–2878
49 Borsani, O. *et al.* (2005) Endogenous siRNAs derived from a pair of natural *cis*-antisense transcripts regulate salt tolerance in *Arabidopsis*. *Cell* 123, 1279–1291
50 Henz, S.R. *et al.* (2007) Distinct expression patterns of natural antisense transcripts in *Arabidopsis*. *Plant Physiol.* 144, 1247–1255
51 Jen, C.H. *et al.* (2005) Natural antisense transcripts with coding capacity in *Arabidopsis* may have a regulatory role that is not linked to double-stranded RNA degradation. *Genome Biol.* 6, R51
52 Wang, X.J. *et al.* (2005) Genome-wide prediction and identification of *cis*-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* 6, R30
53 Jin, H. *et al.* (2008) Small RNAs and the regulation of *cis*-natural antisense transcripts in *Arabidopsis*. *BMC Mol. Biol.* 9, 6
54 Katiyar-Agarwal, S. *et al.* (2007) A novel class of bacteria-induced small RNAs in *Arabidopsis*. *Genes Dev.* 21, 3123–3134
55 Katiyar-Agarwal, S. *et al.* (2006) A pathogen-inducible endogenous siRNA in plant immunity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18002–18007
56 Vazquez, F. *et al.* (2004) Endogenous *trans*-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. *Mol. Cell* 16, 69–79
57 Allen, E. *et al.* (2005) microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121, 207–221
58 Peragine, A. *et al.* (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* 18, 2368–2379
59 Yoshikawa, M. *et al.* (2005) A pathway for the biogenesis of *trans*-acting siRNAs in *Arabidopsis*. *Genes Dev.* 19, 2164–2175
60 Voinnet, O. (2005) Non-cell autonomous RNA silencing. *FEBS Lett.* 579, 5858–5871
61 Himber, C. *et al.* (2003) Transitivity-dependent and -independent cell-to-cell movement of RNA silencing. *EMBO J.* 22, 4523–4533
62 Dunoyer, P. *et al.* (2007) Intra- and intercellular RNA interference in *Arabidopsis thaliana* requires components of the microRNA and heterochromatic silencing pathways. *Nat. Genet.* 39, 848–856
63 Dunoyer, P. *et al.* (2005) DICER-LIKE 4 is required for RNA interference and produces the 21-nucleotide small interfering RNA component of the plant cell-to-cell silencing signal. *Nat. Genet.* 37, 1356–1360
64 Palauqui, J.C. *et al.* (1997) Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* 16, 4738–4745
65 Brosnan, C.A. *et al.* (2007) Nuclear gene silencing directs reception of long-distance mRNA silencing in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14741–14746
66 Axtell, M.J. *et al.* (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* 127, 565–577
67 Kim, M. *et al.* (2001) Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* 293, 287–289
68 Lehmann, E. *et al.* (2007) Molecular basis of RNA-dependent RNA polymerase II activity. *Nature* 450, 445–449
69 Cramer, P. (2006) Recent structural studies of RNA polymerases II and III. *Biochem. Soc. Trans.* 34, 1058–1061
70 Schwartz, L.B. *et al.* (1974) Isolation and partial characterization of the multiple forms of deoxyribonucleic acid-dependent ribonucleic acid polymerase in the mouse myeloma, MOPC 315. *J. Biol. Chem.* 249, 5889–5897
71 Greco-Stewart, V.S. *et al.* (2007) The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology* 357, 68–78
72 Ding, B. and Itaya, A. (2007) Viroid: a useful model for studying the basic principles of infection and RNA biology. *Mol. Plant Microbe Interact.* 20, 7–20
73 Wassarman, K.M. and Saecker, R.M. (2006) Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* 314, 1601–1603
74 Zhang, X. *et al.* (2007) Role of RNA polymerase IV in plant small RNA metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4536–4541
75 Huettel, B. *et al.* (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. *EMBO J.* 25, 2828–2836
76 Buhler, M. *et al.* (2007) RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell* 129, 707–721
77 Buhler, M. *et al.* (2006) Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* 125, 873–886
78 Qi, Y. *et al.* (2006) Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 443, 1008–1012

APPENDIX B

THE ARABIDOPSIS CHROMATIN-MODIFYING NUCLEAR siRNA PATHWAY INVOLVES A NUCLEOLAR RNA PROCESSING CENTER

Published in Cell vol. 126;1 79-92

My contributions to this work:

I performed the DNA methylation assays by using Southern blotting in Figure 1. I also created the *nrpd1a nrpd2* double mutant used in Figure 1 to demonstrate that the *nrpd1a nrpd2* mutants are indistinguishable from the single mutants in terms of DNA methylation, arguing genetically that NRPD1a and NRPD2 are part of the same complex. I designed several of the anti-peptide antibodies used in the immunolocalization analyses that were instrumental to the results and interpretation of the paper. I offered comments on the manuscript and participated in discussion about the interpretation of the results.

Cell

The *Arabidopsis* Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center

Olga Pontes,[1] Carey Fei Li,[2] Pedro Costa Nunes,[1,4] Jeremy Haag,[1] Thomas Ream,[1] Alexa Vitins,[1]
Steven E. Jacobsen,[2,3] and Craig S. Pikaard[1,*]
[1] Biology Department, Washington University, 1 Brookings Drive, St. Louis, MO 63130, USA
[2] Department of Molecular, Cell and Developmental Biology
[3] Howard Hughes Medical Institute
University of California, Los Angeles, Los Angeles, CA 90095, USA
[4] Secção de Genética, Centro de Botânica Aplicada à Agricultura, Instituto Superior de Agronomia, Tapada da Ajuda, 1349-017 Lisboa, Portugal
*Contact: pikaard@biology.wustl.edu
DOI 10.1016/j.cell.2006.05.031

SUMMARY

In *Arabidopsis thaliana*, small interfering RNAs (siRNAs) direct cytosine methylation at endogenous DNA repeats in a pathway involving two forms of nuclear RNA polymerase IV (Pol IVa and Pol IVb), RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), DICER-LIKE 3 (DCL3), ARGONAUTE4 (AGO4), the chromatin remodeler DRD1, and the de novo cytosine methyltransferase DRM2. We show that RDR2, DCL3, AGO4, and NRPD1b (the largest subunit of Pol IVb) colocalize with siRNAs within the nucleolus. By contrast, Pol IVa and DRD1 are external to the nucleolus and colocalize with endogenous repeat loci. Mutation-induced loss of pathway proteins causes downstream proteins to mislocalize, revealing their order of action. Pol IVa acts first, and its localization is RNA dependent, suggesting an RNA template. We hypothesize that maintenance of the heterochromatic state involves locus-specific Pol IVa transcription followed by siRNA production and assembly of AGO4- and NRPD1b-containing silencing complexes within nucleolar processing centers.

INTRODUCTION

In diverse eukaryotes, small interfering RNAs (siRNAs) regulate processes that include mRNA degradation, viral suppression, centromere function, and silencing of retrotransposons and endogenous DNA repeats (Almeida and Allshire, 2005; Baulcombe, 2004; Grewal and Rice, 2004; Tomari and Zamore, 2005). siRNAs are generated by Dicer endonuclease cleavage of double-stranded RNAs (dsRNAs), whose production in *Neurospora*, *C. elegans*, *S. pombe*, and plants involves one or more RNA-dependent RNA polymerases (RdRPs) (Baulcombe, 2004; Wassenegger and Krczal, 2006). Following dicing of dsRNAs into ~20–25 bp duplexes (Bernstein et al., 2001; Hannon, 2002), one RNA strand is loaded into effector complexes that carry out the silencing functions. A defining feature of these effector complexes is the inclusion of an Argonaute (AGO) family protein (Carmell et al., 2002; Sontheimer and Carthew, 2004). In RNA-slicing effector complexes, the AGO-associated siRNA base pairs with its target, thereby positioning the target RNA for endonucleolytic cleavage (Song et al., 2004). Within effector complexes that direct chromatin modifications (Grewal and Rice, 2004; Verdel et al., 2004; Volpe et al., 2002; Wassenegger, 2005), the mechanisms by which siRNAs guide target modifications are not yet understood.

In *Arabidopsis thaliana,* silencing at endogenous repeat loci involves histone H3K9 methylation and RNA-directed DNA methylation that is correlated with the production of homologous siRNAs (Cao et al., 2003; Lippman et al., 2003; Xie et al., 2004; Zilberman et al., 2004). Key players in this chromatin-modifying nuclear siRNA pathway include DICER-LIKE 3 (DCL3), ARGONAUTE4 (AGO4), RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), and two forms of nuclear RNA polymerase IV (Pol IV). The largest and second largest subunits of Pol IV are similar to the catalytic β and β′ subunits of *E. coli* DNA-dependent RNA polymerase and to the corresponding subunits of eukaryotic nuclear RNA polymerases I, II, and III (see Onodera et al., 2005 and references therein). Two genes encode distinct Pol IV largest subunits, and two genes encode Pol IV second largest subunits. Both of the largest-subunit genes (*NRPD1a* and *NRPD1b*) are expressed, but only one of the second-largest-subunit genes (*NRPD2a*) is functional (Herr et al., 2005; Onodera et al., 2005; Pontier et al., 2005). As a result, there are two genetically nonredundant forms of Pol IV, namely Pol IVa and Pol IVb,

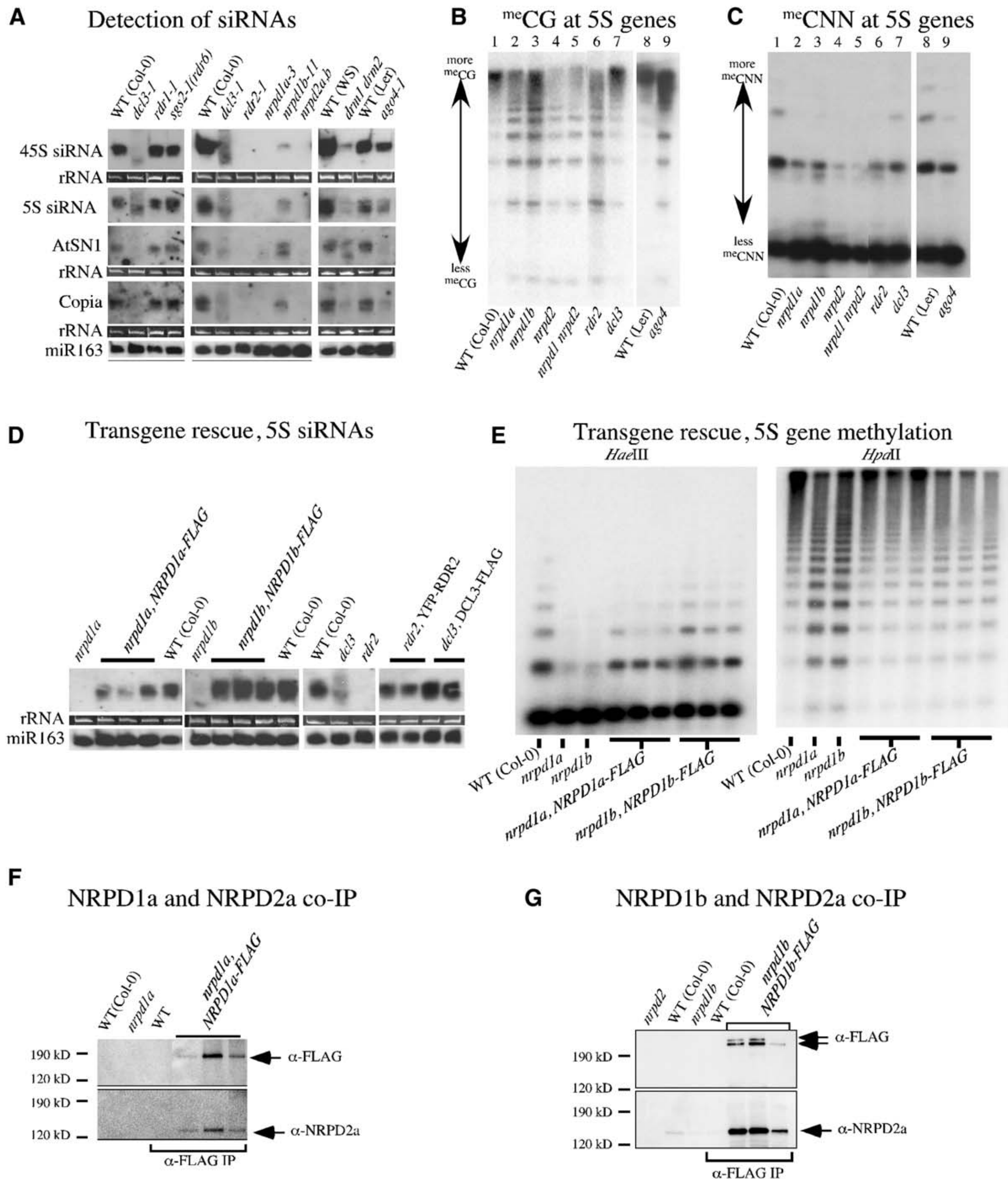


Figure 1. Loss of siRNAs and Cytosine Methylation at Repeated DNA Sequences in Mutants of the Nuclear siRNA Pathway

(A) siRNAs of wild-type (WT) and mutant plants. RNA blots were hybridized to probes corresponding to the 45S rRNA gene intergenic spacer (45S siRNA), the 5S rRNA gene siRNA siR1003, the *AtSN1* family of retroelements, the *Copia* transposable element family, or the microRNA miR163.

(B and C) Loss of CG or CNN methylation at 5S gene repeats. Genomic DNA digested with HpaII or HaeIII was hybridized to a 5S gene probe. *nrpd1a*, *nrpd1b*, *nrpd2*, *rdr2*, and *dcl3* mutants are in the Col-0 genetic background. *ago4* is in the Ler background.

(D) siRNA production in *nrpd1a*, *nrpd1b*, *rdr2*, and *dcl3* mutants is rescued by corresponding transgenes. Genomic clones under the control of their own promoters and encoding C-terminal FLAG-tagged proteins rescued the *nrpd1a*, *nrpd1b*, and *dcl3* mutants (three, three, and two independent transformants, respectively), whereas a YFP-RDR2 cDNA fusion under the control of the cauliflower mosaic virus 35S promoter rescued *rdr2* (two independent transformants shown).

(E) Transgene rescue of 5S rDNA methylation in *nrpd1a* and *nrpd1b* mutants. Southern blot analysis of HaeIII- and HpaII-digested genomic DNA with a 5S gene probe shows that the loss of methylation in *nrpd1a* and *nrpd1b* mutants, relative to wild-type (WT), is restored in each of three independent *NRPD1a-FLAG* or *NRPD1b-FLAG* transgenic lines.

designated according to which largest subunit is used. Disruption of Pol IV, RDR2, DCL3, or AGO4 genes causes decreased cytosine methylation and siRNA accumulation at endogenous repeats, including 5S ribosomal RNA genes and transposable elements (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005; Xie et al., 2004). However, the order in which these proteins act in the biogenesis of nuclear siRNAs is unclear.

Using RNA fluorescence in situ hybridization (RNA-FISH) together with protein immunolocalization, we present evidence for siRNA processing centers associated with the nucleolus. Within these centers, siRNAs colocalize with a significant portion of the RDR2, DCL3, AGO4, and NRPD1b protein pools. The two subunits of Pol IVa, however, do not localize to the processing centers but colocalize with chromosomal loci that are both sources and targets of siRNAs. A portion of the NRPD1b pool also colocalizes with target loci, as does the SWI2/SNF2 chromatin-remodeling ATPase family member DRD1, a protein required for RNA-directed DNA methylation that acts downstream of siRNA production (Kanno et al., 2004). Based on cytological, biochemical, and genetic evidence, we present a spatial and temporal model for nuclear siRNA biogenesis.

RESULTS

Loss of siRNAs and Cytosine Methylation in Nuclear siRNA Pathway Mutants

In *A. thaliana*, siRNAs homologous to repeated gene families are readily detected on RNA blots, as shown for siRNAs corresponding to the intergenic spacers of 45S or 5S rRNA genes or siRNAs corresponding to *AtSN1* or *Copia* transposable-element families (Figure 1A). Collectively, these endogenous repeats represent genes transcribed by RNA polymerase I (45S rRNA genes), RNA polymerase II (*Copia* elements), and RNA polymerase III (5S genes, *AtSN1* elements). The siRNAs are essentially eliminated upon mutation of the Pol IVa largest subunit, *NRPD1a*, or upon mutation of the second subunit of both Pol IVa and Pol IVb, *NRPD2* (note that the *nrpd2a-2 nrpd2b-1* double mutant [Onodera et al., 2005] is abbreviated as *nrpd2* throughout this paper). siRNAs are also eliminated in *rdr2* mutants. By contrast, siRNAs are reduced in abundance, but not eliminated, in *nrpd1b* or *ago4* mutants. A smear of alternatively sized small RNAs is generated in a *dcl3* mutant (Figure 1A) and is probably explained by the action of alternative Dicers (Gasciolli et al., 2005). The abundance of siRNAs is also greatly reduced in the *drm1 drm2* mutant, indicating that de novo cytosine methylation plays a role in nuclear siRNA accumulation.

Loss of endogenous siRNAs correlates with loss of cytosine methylation at corresponding DNA sequences. For instance, 5S gene repeats are heavily methylated at CG motifs, making them resistant to digestion by the methylation-sensitive restriction endonuclease HpaII in wild-type *A. thaliana* (Figure 1B, lanes 1 and 8). CG methylation at HpaII sites is decreased to a similar extent in *rdr2*, *ago4*, *nrpd1a*, *nrpd1b*, and *nrpd2* mutants, resulting in more hybridization signal in digested bands nearer the bottom of Southern blots (Figure 1B). Methylation is least affected in a *dcl3* mutant, presumably because other Dicers partially compensate (Gasciolli et al., 2005).

CNN methylation is a hallmark of RNA-directed DNA methylation, which is accomplished by the de novo cytosine methyltransferase DRM2 (Cao et al., 2003). At 5S gene loci, sensitivity to digestion by HaeIII reports on CNN methylation. 5S genes are more sensitive to HaeIII digestion in *rdr2*, *nrpd1a*, *nrpd1b*, and *nrpd2* mutants compared to wild-type plants (Figure 1C). Mutation of *DCL3* has a lesser effect on CNN methylation, again suggesting partial compensation by other Dicers. Collectively, the data of Figures 1A–1C indicate that the loss of endogenous repeat siRNAs correlates with the loss of both CG and CNN methylation, implicating RNA-directed DNA methylation (Aufsatz et al., 2002; Cao et al., 2003).

To facilitate cytological and biochemical studies, we developed transgenic lines that express functional, epitope-tagged versions of the proteins involved in the nuclear siRNA pathway. Genomic-clone transgenes expressing NRPD1a, NRPD1b, or DCL3 bearing C-terminal FLAG epitope tags all rescued their corresponding mutations and restored siRNA production, as did a YFP-RDR2 fusion engineered using a full-length *RDR2* cDNA (Figure 1D). The *NRPD1a* and *NRPD1b* transgenes also restored cytosine methylation at 5S gene repeats (Figure 1E). Collectively, these results indicate that the recombinant proteins retain their biological functions.

The Alternative Pol IV Largest Subunits, NRPD1a and NRPD1b, Physically Interact with NRPD2

Genetic evidence suggests that the Pol IV second largest subunit NRPD2 interacts with NRPD1a or NRPD1b within Pol IVa or Pol IVb, respectively (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). To obtain biochemical evidence for such interactions, we exploited transgenic plants expressing FLAG-tagged NRPD1a or NRPD1b and an anti-NRPD2 antibody (Onodera et al., 2005) to ask whether NRPD2 associates with the alternative largest subunits in vivo. Indeed, NRPD2 coimmunoprecipitates with both NRPD1a-FLAG and NRPD1b-FLAG in multiple independent transgenic plants (Figures 1F and 1G). The quantity of

(F) Physical interaction between Pol IVa subunits NRPD1a and NRPD2 detected by coimmunoprecipitation. Proteins from multiple independent NRPD1a-FLAG transgenic lines were immunoprecipitated using anti-FLAG antibody, then subjected to SDS-PAGE and electroblotting. Membranes were sequentially analyzed to detect the FLAG epitope (top) and NRPD2 (bottom).

(G) Physical interaction between NRPD1b and NRPD2. The experiment was performed as for (F) using multiple independent NRPD1b-FLAG transgenic lines.

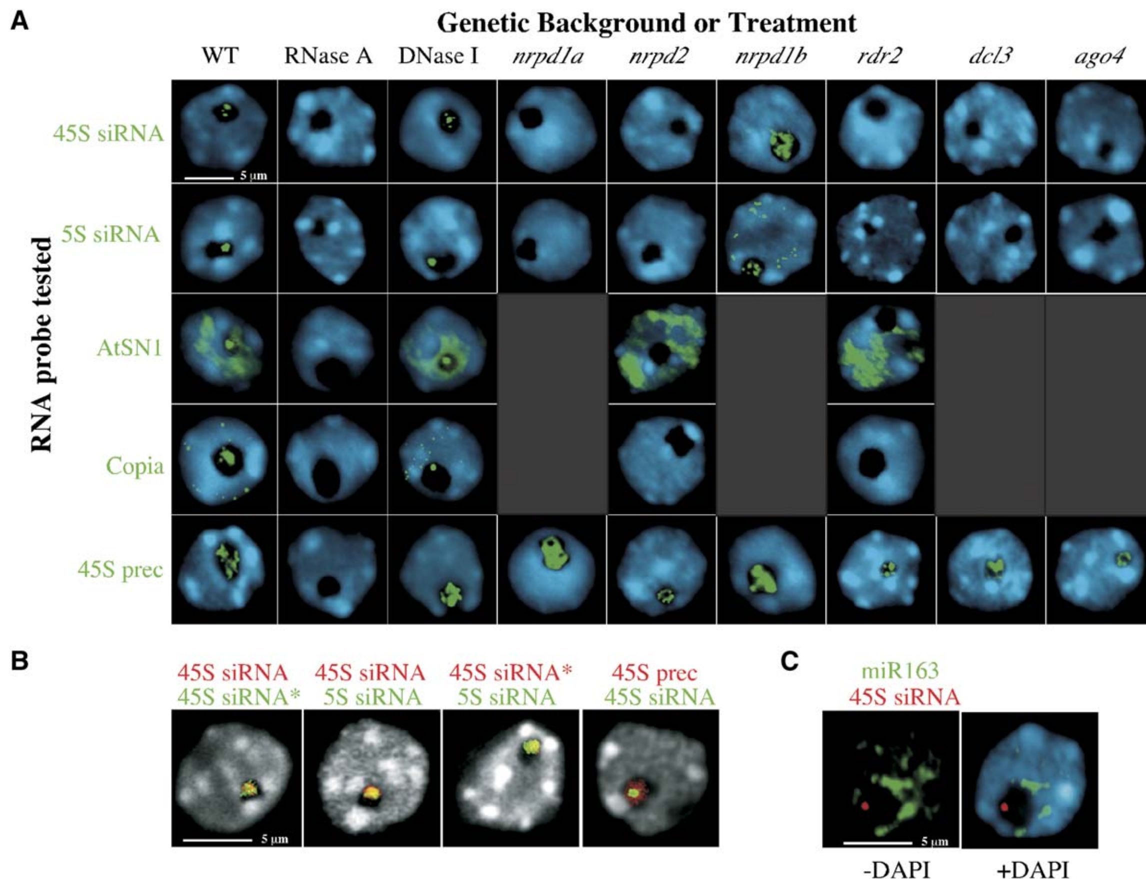


Figure 2. Nuclear Localization of siRNAs

(A) RNA-FISH using the same probe sequences used for the RNA blots of Figure 1A was performed in wild-type, nuclease-treated, or mutant nuclei as indicated. As a control, a probe that detects the 45S rRNA precursor transcripts was also used. Nuclei were counterstained with DAPI (blue). Size bars represent 5 μm in all panels.

(B) Different siRNAs colocalize within the nucleolus. Simultaneous detection of RNA target pairs was performed using two-color FISH. Three-dimensional projections of five to seven optical sections obtained by multiphoton microscopy are shown. The red or green color of the lettering corresponds to the color of the signal for the indicated probes. Nuclei were counterstained with DAPI (false colored gray in these images). Thirty-five nuclei were observed for each probe combination. In all nuclei examined, at least 50% of the green and red pixels overlapped in the digital images to yield yellow signals.

(C) Two-color FISH using the 45S siRNA probe (red) and miR163 probe (green). Nuclei were counterstained with DAPI (blue). A localization pattern like that shown was observed in all 155 nuclei examined.

coimmunoprecipitated NRPD2 is proportional to the abundance of NRPD1a or NRPD1b in the different lines, as expected of subunits with fixed stoichiometries.

siRNAs Are Concentrated within the Nucleolus

It is not known where endogenous siRNAs are generated or processed within the cell. So, to detect siRNAs or their precursors, we employed RNA fluorescence in situ hybridization (RNA-FISH) with digoxigenin- or biotin-labeled probes (Figure 2A) identical in sequence to those used for siRNA blot hybridization (see Figure 1A). With all siRNA probes, an intense hybridization signal was observed within the nucleolus, which is the region of the nucleus not stained appreciably by the fluorescent DNA binding dye DAPI. This was true of leaf mesophyll cells at interphase, as shown throughout this paper, and in root meristem cells (O.P., unpublished data). In the case of the *AtSN1* probe, a diffuse signal was also observed throughout the nucleoplasm. The nucleolar dots detected with siRNA probes occupy a small portion of the nucleolus when compared to the 45S pre-rRNA precursor transcripts that are generated by RNA polymerase I and processed in the nucleolus (Figure 2A, bottom row).

Hybridization signals detected using different siRNA probes colocalized, as shown using two-color RNA-FISH with probes specific for 45S siRNAs corresponding to opposite DNA strands (45S siRNA and 45S siRNA*) or 5S siRNAs (Figure 2B). These siRNA probe signals are spatially distinct from the signals obtained using a miRNA probe (Figure 2C). Collectively, these data indicate that nuclear siRNA hybridization signals localize within a discrete compartment of the nucleolus, smaller than the

volume occupied by 45S pre-rRNA and distinct from sites where miRNA or their precursors are concentrated.

As shown in Figure 2A, siRNA and pre-rRNA hybridization signals are eliminated if nuclei are treated with ribonuclease A (RNase A) prior to extensive washing and probe hybridization but are not affected by DNase I treatment. These tests suggest that the hybridization signals result from the RNA probes' annealing to RNA targets. Importantly, the nucleolar dot signals are absent in *nrpd2*, *nrpd1a*, *rdr2*, *dcl3*, or *ago4* mutants, and, typically, no signal is observed elsewhere (although low-intensity, dispersed signals occurred infrequently; see Table S1 in the Supplemental Data available with this article online for quantitative data). The exception is *nrpd1b*, for which dispersal of the nucleolar dot (as shown in Figure 2A) is more common than complete loss of signal (see Table S1). In general, these observations are consistent with the RNA blot hybridization data (Figure 1A). Importantly, 45S pre-rRNAs are unaffected by the siRNA pathway mutations, as expected.

The loss of hybridization signals in the mutants, including *dcl3* and *ago4*, which should act downstream of siRNA precursor formation, suggests that we are detecting siRNAs in the nucleolar dots rather than precursors. Perhaps the latter escape detection because they are dispersed throughout the nucleus and not concentrated in one location. However, the *AtSN1* signals, external to the nucleolus, that persist in the mutants might be precursor RNAs.

Nucleolar siRNA Processing Centers

The detection of nuclear siRNAs prompted us to ask where the proteins of the nuclear siRNA pathway are located. NRPD1a, NRPD1b, RDR2, DCL3, and AGO4 were immunolocalized in transgenic nuclei by virtue of their epitope or YFP tags, whereas native NRPD2 was localized using an anti-peptide antibody (Figure 3A, top row). NRPD1a and NRPD2, the known subunits of Pol IVa, showed similar, punctate localization patterns; significantly, neither protein associates with the nucleolus. By contrast, FLAG-tagged NRPD1b, the largest subunit of Pol IVb, localizes within a nucleolar dot in addition to puncta external to the nucleolus (see also Li et al., 2006 [this issue of *Cell*] and Table S2). RDR2, DCL3, and AGO4 also display prominent nucleolar dot signals in addition to puncta or diffuse signals outside the nucleolus. RDR2 signals are distinctive in that a ring or crescent at the perimeter of the nucleolus is typically observed in addition to the nucleolar dot, and this is true for both epitope-tagged and native RDR2. Control experiments showed that no immunolocalization signals were detected in transgenic nuclei if primary antibodies were omitted; likewise, no signals were detected in wild-type nuclei using anti-FLAG, anti-Myc, or anti-YFP antibodies (see Figure S1).

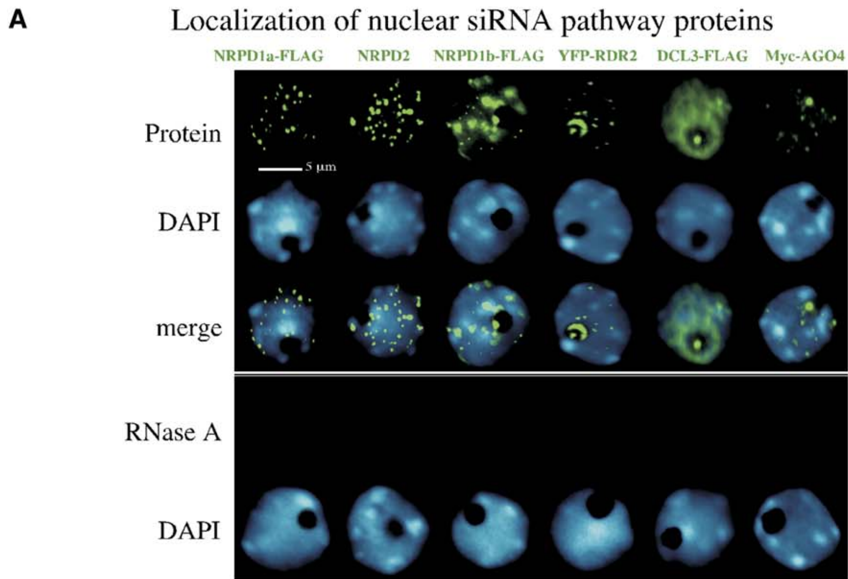
Nucleolar dot signals can be observed at the center or the periphery of the nucleolus, consistent with data of Li et al. (2006) showing that AGO4 colocalizes with markers of nucleolar accessory bodies, or Cajal bodies (Cioce and Lamond, 2005). Cajal bodies are dynamic nuclear organelles that can move in and out of nucleoli (Boudonck et al., 1999) and are implicated in the assembly of RNA-protein complexes, including snRNPs and snoRNPs (Cioce and Lamond, 2005). Therefore, what we call nucleolar dots throughout this paper are likely to be Cajal bodies or related entities (see Li et al., 2006).

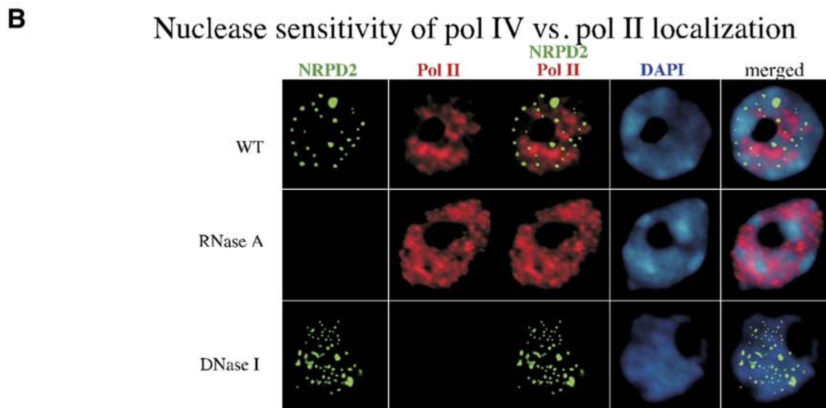
Treating nuclei with RNase A prior to antibody incubation caused a complete loss of signal for all of the proteins in the majority of nuclei examined, suggesting that the proteins are not retained in RNA-depleted nuclei (Figure 3A). However, a minority of the nuclei continued to show wild-type protein localization patterns, albeit at reduced intensity, suggesting that not all nuclei are equally accessible to RNase treatment (see Table S2). Further analysis showed that, whereas NRPD2, NRPD1a, and NRPD1b signals are lost from RNase A-treated nuclei, the proteins are not lost from DNase I-treated nuclei, although NRPD1b and NRPD2 are partially mislocalized (Figure 3B and Figure S2, green signals). Conversely, the signals for the second largest subunit of DNA-dependent RNA polymerase II are lost upon DNase, but not RNase, treatment (Figure 3B, red signals). Collectively, these observations suggest that Pol IV interacts with RNA rather than DNA templates, unlike Pol II.

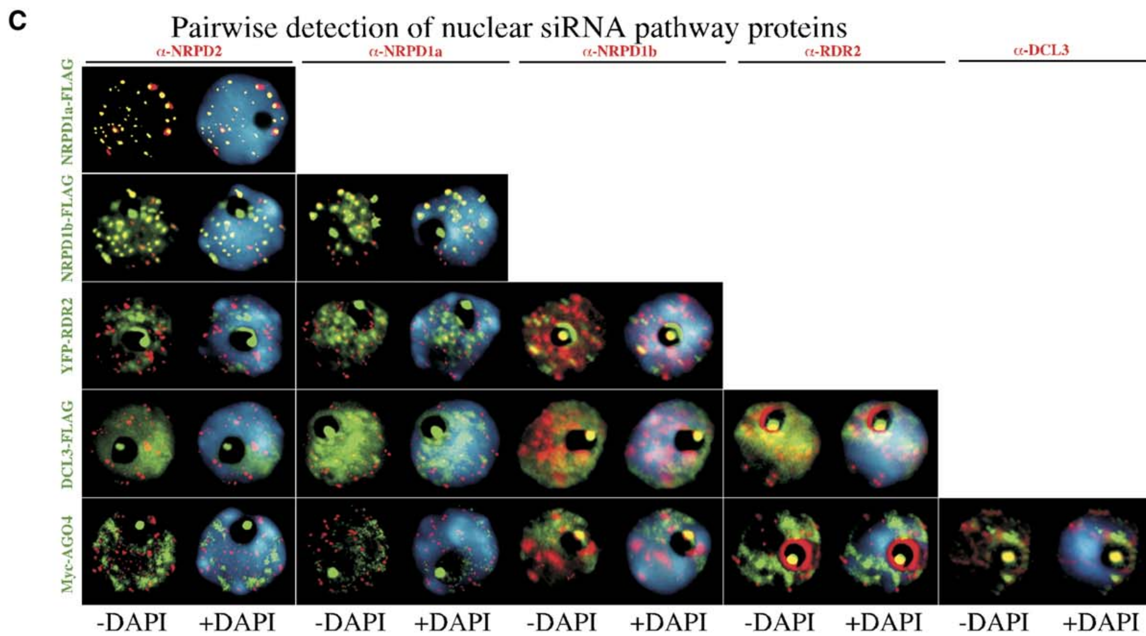
Using anti-epitope antibodies that detect transgene-encoded recombinant proteins, in combination with anti-peptide antibodies recognizing the native proteins, we simultaneously localized pairs of proteins using two-color immunofluorescence (Figure 3C; Table S3). The native proteins and the recombinant proteins were found to display the same localization patterns, indicating that the anti-peptide antibodies are specific for their targets and that the epitope tags do not disrupt recombinant protein localization. NRPD1a and NRPD2, the subunits of Pol IVa, colocalize precisely, resulting in yellow signals (Figure 3C, top row; note that differences in intensity of the green and red signals influence the apparent extent of overlap). Slightly more than half of the NRPD1b foci external to the nucleolus colocalize with the NRPD1a/NRPD2 foci (Figure 3C, second row from top), suggesting that Pol IVb occurs at approximately half of the Pol IVa foci. However, the remaining NRPD1b foci are spatially distinct from NRPD2 (and NRPD1a). A conclusion from the latter observation is that the Pol IVb largest subunit can exist apart from the second largest subunit, both external to the nucleolus and within the nucleolus, where no NRPD2 is detectable.

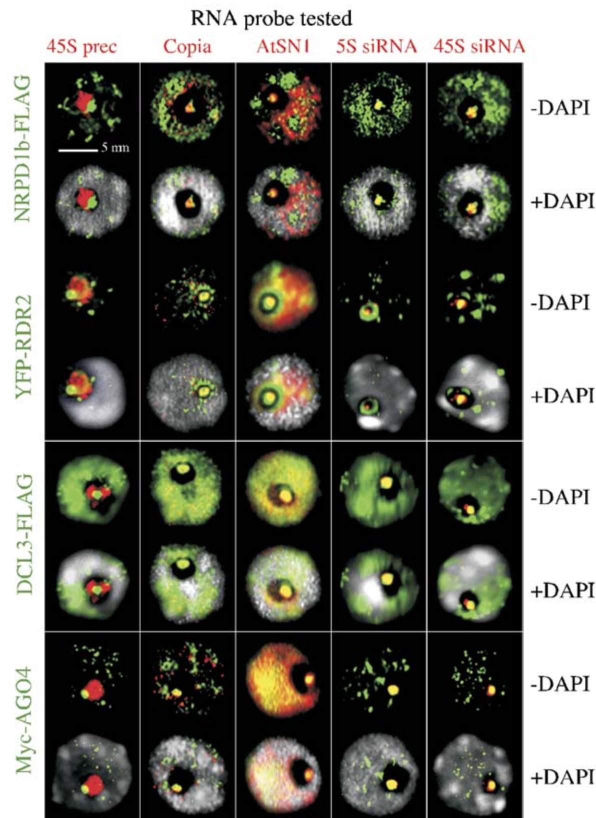
External to the nucleolus, NRPD1a, NRPD2, and NRPD1b do not colocalize with RDR2, DCL3, or AGO4. However, the portion of the NRPD1b pool that is nucleolus associated colocalizes with RDR2, DCL3, and AGO4 within the nucleolar dot (Figure 3C).

We next asked whether the nucleolar dots previously detected by RNA-FISH (Figure 2) correspond to the same nucleolar dots where NRPD1b, RDR2, DCL3, and AGO4 colocalize (Figure 3). To address this question, we performed protein immunolocalization followed by


A
Localization of nuclear siRNA pathway proteins
NRPD1a-FLAG
NRPD2
NRPD1b-FLAG
YFP-RDR2
DCL3-FLAG
Myc-AGO4
Protein
5 μm
DAPI
merge
RNase A
DAPI


B
Nuclease sensitivity of pol IV vs. pol II localization
NRPD2
Pol II
NRPD2
Pol II
DAPI
merged
WT
RNase A
DNase I


C
Pairwise detection of nuclear siRNA pathway proteins
α-NRPD2
α-NRPD1a
α-NRPD1b
α-RDR2
α-DCL3
NRPD1a-FLAG
NRPD1b-FLAG
YFP-RDR2
DCL3-FLAG
Myc-AGO4
-DAPI
+DAPI
-DAPI
+DAPI
-DAPI
+DAPI
-DAPI
+DAPI
-DAPI
+DAPI



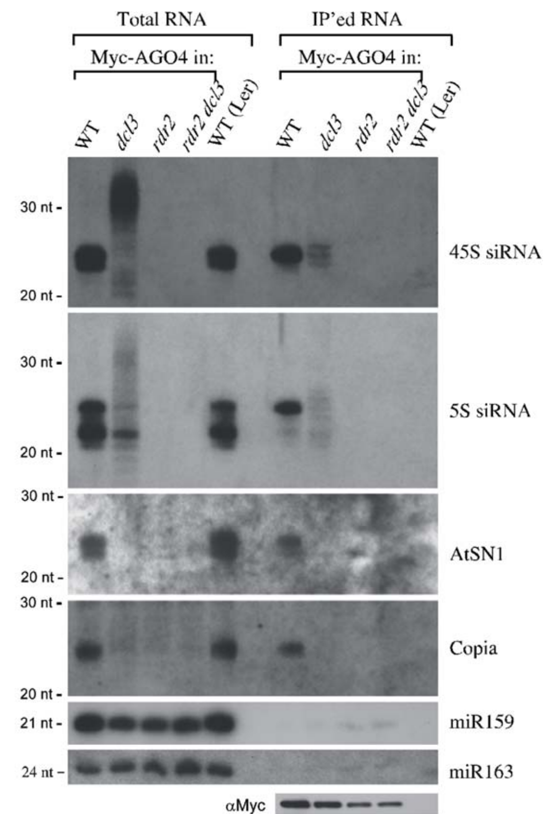


Figure 4. siRNAs Colocalize with NRPD1b, RDR2, DCL3, and AGO4

(A) Nuclei were hybridized with 45S rRNA precursor, *Copia*, *AtSN1*, 5S siRNA, or 45S siRNA probes (red signals). NRPD1b-FLAG, YFP-RDR2, DCL3-FLAG, or Myc-AGO4 was immunolocalized using anti-FLAG, anti-YFP, or anti-Myc antibodies (green signals). Images shown are three-dimensional projections of five to seven optical sections obtained by multiphoton microscopy. Pairs of images are presented for each protein localized, the lowermost image including the DAPI signal (false colored gray) to help reveal the nucleolus.

(B) siRNAs physically associate with AGO4. Total RNA or RNA immunoprecipitated (IP) using anti-Myc antibodies from transgenic plants expressing Myc-AGO4 in wild-type, *dcl3*, *rdr2*, or *dcl3 rdr2* backgrounds was subjected to RNA blot hybridization using 45S siRNA, 5S siRNA, *AtSN1*, *Copia*, and miR159 probes. RNA of nontransgenic wild-type plants (ecotype Ler) served as a control. The presence of AGO4 in immunoprecipitates was confirmed by immunoblotting using anti-Myc antibody.

RNA-FISH (Figure 4A). As is evident by the yellow signals resulting from siRNA probe and protein signal overlap, NRPD1b, RDR2, DCL3, and AGO4 typically colocalize with 45S, 5S, *AtSN1*, and *Copia* siRNAs within the nucleolar dots but do not colocalize precisely with 45S rRNA precursor transcripts (Figure 4A; see also Table S4). We interpret the colocalization of NRPD1b, RDR2, DCL3, AGO4, and siRNAs as evidence of siRNA processing centers in which dsRNAs generated by RDR2 are diced by DCL3 to generate siRNAs that are loaded into RISC effector complexes that contain AGO4 and NRPD1b.

Consistent with the interpretation that siRNAs are stably associated with AGO4, immunoprecipitation of Myc-AGO4 pulls down 45S, 5S, *AtSN1*, and *Copia* siRNAs (Figure 4B). Moreover, in *rdr2* or *rdr2 dcl3* double mutants, siRNAs are no longer found in the Myc-AGO4 immunoprecipitates. In *dcl3* mutants, siRNAs associated with AGO4 are greatly reduced in abundance and variable in size, consistent with the hypothesis that AGO4 is capable of binding siRNAs generated by other Dicers that partially compensate for the loss of DCL3.

Figure 3. Immunolocalization of Nuclear siRNA Pathway Proteins

(A) Epitope-tagged NRPD1a, NRPD1b, DCL3, and AGO4 recombinant proteins that rescue corresponding mutations were immunolocalized (green signals) using anti-FLAG or anti-Myc antibodies. Native NRPD2 was detected using anti-peptide antisera. RDR2-YFP was localized using anti-YFP. Nuclei were counterstained with DAPI.

(B) Immunolocalization of NRPD2 and the Pol II second largest subunit in wild-type untreated, RNase A-, or DNase I-treated nuclei.

(C) Anti-peptide antibodies recognizing native proteins (red signals) were used in combination with antibodies recognizing FLAG-, Myc-, or YFP-tagged recombinant proteins (green signals) in nuclei of transgenic plants. Colocalizing proteins generate yellow signals.

Pol IV and the Putative Chromatin Remodeler DRD1 Colocalize with Endogenous Repeats

To determine where the endogenous DNA repeats are located relative to the nucleolar dots, we used DNA-FISH to localize the 45S rRNA gene loci (i.e., the nucleolus organizer regions; NORs) and 5S rRNA gene clusters. The FISH signals for the highly condensed portions of 45S and 5S rRNA gene loci are not detected within the nucleolus (Figure 5, red signals), indicating that the bulk of the target gene loci, composed mostly of inactive repeats, are distant from the nucleolar dots.

By combining protein immunolocalization (green signals) with DNA-FISH (red signals), we asked whether the Pol IV foci external to the nucleolus correspond to endogenous repeat loci. Indeed, NORs and 5S gene loci were found to colocalize with NRPD1a, NRPD1b, and NRPD2, yielding yellow signals at most, though not all, of the loci (see Table S5 for quantitative data). Some overlap between 5S gene loci and RDR2 or DCL3 signals was also observed, although the diffuse distribution of DCL3 may make the apparent overlap coincidental. We also examined the localization of DRD1, a SWI2/SNF2-related protein that is involved in RNA-directed DNA methylation via a Pol IVb-dependent pathway (Kanno et al., 2005; Kanno et al., 2004). DRD1 is distributed throughout the nucleus, with the exception of the nucleolus, and is concentrated at chromocenters that include NORs and 5S gene loci (Figure 5, bottom row). Collectively, these observations suggest that Pol IVa, Pol IVb, and DRD1 are present at the endogenous repeat loci, presumably acting in the generation of siRNA precursors or in the downstream functioning of siRNA-containing effector complexes.

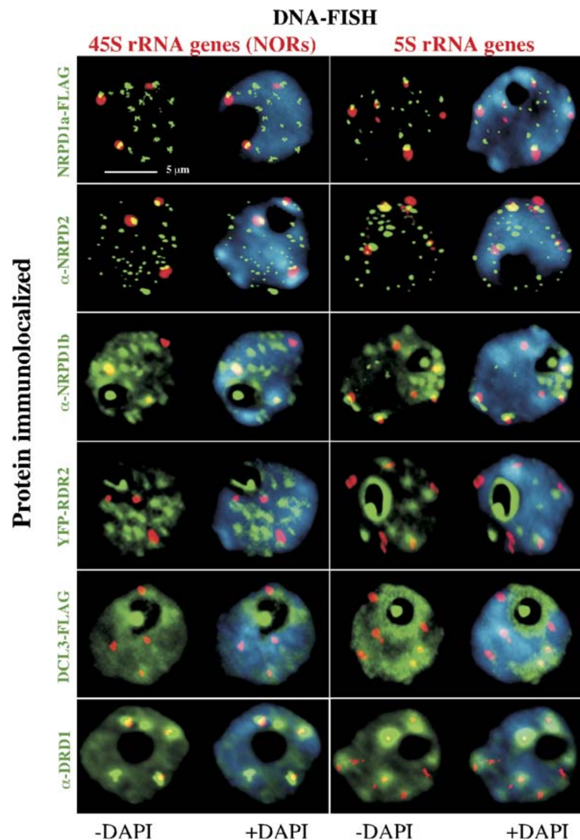


Figure 5. Pol IV Colocalizes with Endogenous Repeat Loci
45S rRNA gene loci (nucleolus organizer regions; NORs) or 5S gene chromosomal loci were visualized using DNA-FISH (red signals), and the indicated proteins were immunolocalized (green signals). Yellow indicates overlapping DNA and protein signals. NRPD1a-FLAG and DCL3-FLAG recombinant proteins were detected in nuclei of transgenic plants using anti-FLAG antibodies; NRPD2, NRPD1b, and DRD1 were detected in nuclei of nontransgenic plants using anti-peptide antibodies recognizing the native proteins; and recombinant YFP-RDR2 was detected using anti-YFP (green signals). Nuclei were counterstained with DAPI (blue). Note that *A. thaliana* has four NORs and six 5S gene loci in the Col-0 ecotype. The NORs tend to coalesce such that only three NORs are observed in most of the images shown.

Mutation-Induced Mislocalization of Nuclear siRNA Pathway Proteins

To deduce the order in which proteins of the nuclear siRNA pathway act, we examined the effect of mutations on each protein's localization, resulting in the matrix of images shown in Figure 6 (see Table S6 for quantitative data). Protein signals were absent upon mutation of the genes that encode the corresponding proteins, as expected, indicating that all of the mutants are protein nulls and that the antibodies are specific for their intended targets. NRPD1a localization is unaffected in *rdr2*, *dcl3*, or *ago4* mutants, as is NRPD2 localization, consistent with Pol IVa acting upstream of RDR2, DCL3, and AGO4. RDR2 localization is dependent on Pol IVa (NRPD1a and NRPD2), but not on NRPD1b, DCL3, or AGO4, indicating that RDR2 acts downstream of Pol IVa, but upstream of Pol IVb, dicing and effector complex assembly.

DCL3 localization is dependent on both Pol IVa and RDR2 but is independent of AGO4 and NRPD1b, suggesting that dicing occurs following double-stranded RNA formation, mediated by RDR2, and upstream of effector complex assembly and Pol IVb function. Consistent with this interpretation, the NRPD1b nucleolar dot is absent in *nrpd1a*, *rdr2*, *dcl3*, and *ago4* mutants but is still present in a *drd1* mutant (see Figure S3), indicating that the nucleolar NRPD1b signal is dependent on siRNA processing and effector complex assembly but is formed upstream of steps that involve chromatin remodeling by DRD1. The NRPD1b signals that are outside the nucleolus are unaffected in *rdr2* or *dcl3* mutants but are less punctate and therefore appear more diffuse in the *drd1* mutant, suggesting that DRD1 influences NRPD1b localization at target loci.

DISCUSSION

A Spatial and Temporal Model for the Nuclear siRNA Pathway

RNA-directed DNA methylation requires de novo methyltransferase activity, suggesting that DRM-class cytosine

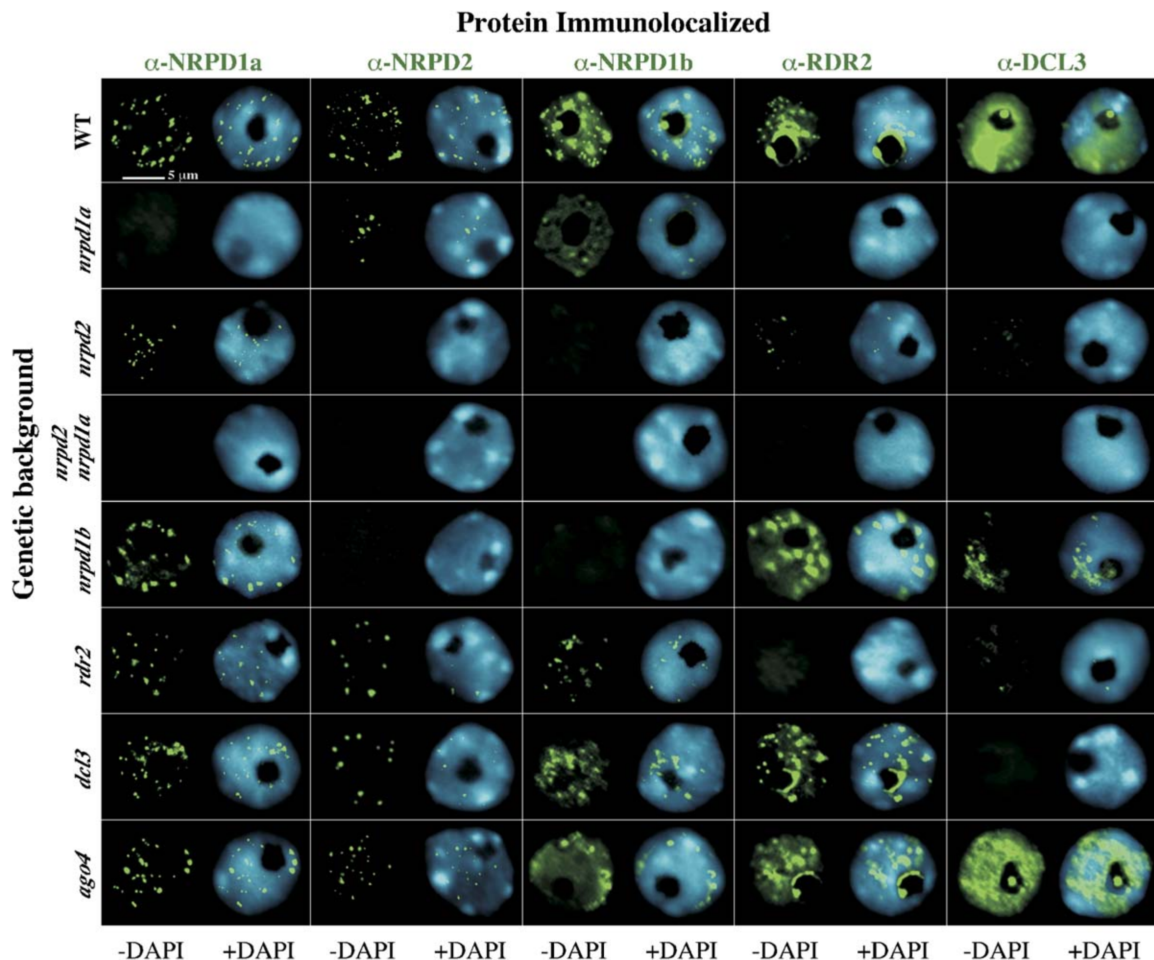


Figure 6. Effects of Mutations on the Localization of Proteins Involved in Nuclear siRNA Biogenesis
The figure shows a matrix of images in which NRPD1a, NRPD2, NRPD1b, RDR2, and DCL3 were immunolocalized using anti-peptide antibodies recognizing the native proteins (green signals) in multiple genetic backgrounds as indicated along the vertical axis. Nuclei were counterstained with DAPI (blue).

methyltransferases (probably *DRM2* only, because *DRM1* is not expressed appreciably) act downstream of siRNA production (Cao et al., 2003). However, endogenous nuclear siRNAs fail to accumulate in *drm* mutants (Xie et al., 2004; Zilberman et al., 2004), suggesting that DRM2 also acts upstream of siRNA production (see also Figure 1A). Our model attempts to address this apparent paradox (Figure 7). Based on a study in *Neurospora* suggesting that methylation impedes RNA polymerase elongation (Rountree and Selker, 1997), we propose that transcripts trailing from polymerases that are stalled or slowed by DRM-mediated methylation (Figure 7, upper left) are sensed as aberrant and, directly or indirectly, become templates for Pol IVa. In this model, Pol IVa is spatially tethered to the DNA by virtue of the RNA template. This aspect of the model accounts for the colocalization of Pol IVa subunits with endogenous repeat loci and their loss in RNase A-treated nuclei. We place Pol IVa first in the pathway because Pol IVa is located directly at the endogenous repeat loci and because mutation of either Pol IVa subunit (NRPD1a or NRPD2) eliminates siRNA production. By contrast, mutation of NRPD1b, the largest subunit of Pol IVb, which also colocalizes with the endogenous repeat loci, does not eliminate siRNA production but does affect RNA-directed cytosine methylation, suggesting that Pol IVb acts late in the pathway (Kanno et al., 2005; Pontier et al., 2005; Vaucheret, 2005; see also Figures 1A–1C). The fact that siRNA accumulation is reduced in *nrpd1b* mutants (see Figure 1A) may be due to the destabilization of the NRPD2 pool upon loss of NRPD1b (see Figure 1G, Figure 6 and Pontier et al., 2005). Loss of NRPD2 would indirectly deplete Pol IVa activity by depriving NRPD1a of its partner catalytic subunit. Alternatively, decreased Pol IVb-dependent cytosine methylation might decrease the incidence of aberrant transcript production at endogenous repeat loci, thereby depleting the pool of Pol IVa templates. These alternative explanations are not mutually exclusive.

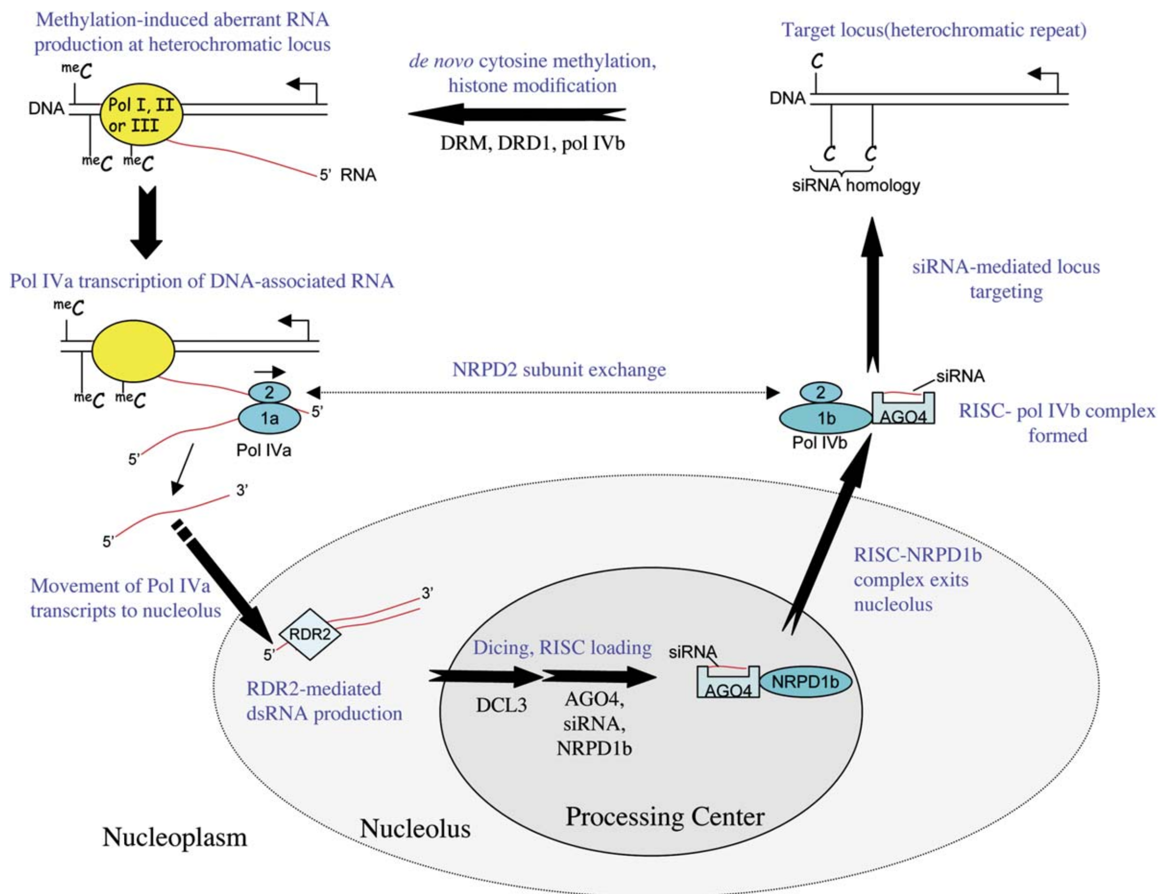


Figure 7. A Spatial and Temporal Model for Nuclear siRNA Biogenesis

Subunits of Pol IVa (abbreviated 1a and 2) colocalize with endogenous repeat loci but are mislocalized upon RNase A treatment, suggesting that Pol IVa transcribes RNA templates whose spatial distribution is influenced by DNA. We propose that cytosine methylation by DRM induces the production of aberrant RNAs, possibly by impeding polymerase elongation, which Pol IVa then uses as templates. Pol IVa transcripts then move, by an unknown mechanism, to the nucleolus, where RDR2, DCL3, and AGO4 are located. In the siRNA processing center, the largest subunit of Pol IVb, NRPD1b, joins the AGO4-containing RISC complex and acquires the NRPD2 subunit to become functional Pol IVb only upon leaving the nucleolus. Formation of Pol IVb is required for the stability of the NRPD2 pool despite the fact that NRPD2 colocalizes more precisely with NRPD1a than with NRPD1b, suggesting that NRPD2 subunits exchange between Pol IVa and b. AGO4, Pol IVb, and DRD1 then play unspecified roles in guiding heterochromatic modifications at the endogenous repeats, including de novo cytosine methylation by DRM. Methylation-dependent production of aberrant RNAs results in a positive feedback loop for maintaining heterochromatin at the DNA repeats.

Like Pol IVa, RDR2 is required for endogenous siRNA production. RDR2 is mislocalized in an *nrpd1a* mutant, whereas the converse is not true (see Figure 6), indicating that RDR2 acts downstream of Pol IVa. RDR2 is not abundant at the endogenous repeats but is concentrated in the nucleolus. Collectively, these observations suggest that Pol IVa generates precursor RNAs at the endogenous repeats and that these transcripts then move to the nucleolus, where their complements are generated by RDR2 transcription. Annealing of these RNAs would produce dsRNAs that are then diced by DCL3 and loaded into an AGO4-containing effector complex, or RISC (RNA-induced silencing complex), within the siRNA processing center. The observation that Pol IVa subunits and RDR2 are not mislocalized in *dcl3* or *ago4* mutants is consistent with Pol IVa and RDR2 acting upstream of DCL3 and AGO4. Likewise, the absence of siRNAs associated with AGO4 in *rdr2* mutants, the atypical sizes of siRNAs associated with AGO4 in *dcl3* mutants, and the mislocalization of AGO4 in *rdr2* or *dcl3* mutants (see also Li et al., 2006) indicate that AGO4 acts downstream of RDR2 and DCL3.

Two observations suggest that Pol IVb acts downstream of AGO4-RISC assembly. First, the largest subunit of Pol IVb, NRPD1b, colocalizes with the nucleolar dot, but only if siRNAs are being produced and assembled into effector complexes; the nucleolar NRPD1b signal is absent in *nrpd1a*, *rdr2*, *dcl3*, or *ago4* mutants. Second, the NRPD2 subunit is never observed within the nucleolus yet is presumably essential for Pol IVb function based on the genetic screen of Kanno et al. that recovered nine mutant alleles of *NRPD1b* and 12 alleles of *NRPD2a* but no alleles of *NRPD1a* (Kanno et al., 2005). The genetic evidence

strongly predicts that NRPD1b is nonfunctional in the absence of the second largest subunit. We propose that NRPD1b associates with AGO4-RISC, which is supported by our immunolocalization data and the finding that NRPD1b can be coimmunoprecipitated in association with AGO4 (Li et al., 2006). Upon leaving the nucleolus as a subunit of AGO4-RISC, we deduce that NRPD1b can then associate with NRPD2, forming functional Pol IVb. Consistent with this hypothesis, NRPD2 coimmunoprecipitates with AGO4 (J.H. and C.S.P., unpublished data) as well as with NRPD1b (see Figure 1G).

How AGO4-RISC-Pol IVb complexes mediate their effects on chromatin modification at target loci is unclear. One possibility is that AGO4-RISC directs Pol IVb to its target sites. Alternatively, AGO4 might transfer the siRNA to Pol IVb when the NRPD2 subunit joins the NRPD1b subunit, after the AGO4-RISC-NRPD1b complex leaves the nucleolus. The siRNA, or a Pol IVb transcript primed by the siRNA, might then be used to conduct a homology search for target sequences, aided by DRD1 (Kanno et al., 2004), a member of the SWI2/SNF2-related family of chromatin-remodeling ATPases that is within a subfamily most closely related to yeast RAD54. In double-strand DNA break repair, RAD54 is required for helping broken DNA ends conduct a homology search and invade homologous duplex DNA of a sister chromosome, thereby facilitating repair by homologous recombination (Krogh and Symington, 2004). A partnership between Pol IVb and DRD1 could account for their presence at the target loci, the observation that NRPD1b and DRD1 are both essential for cytosine methylation but not siRNA production (Kanno et al., 2004, 2005), and the partial mislocalization of NRPD1b in a *drd1-6* mutant (see Figure S2). Moreover, RNA polymerases and chromatin-remodeling ATPases are nucleotide triphosphate-hydrolyzing molecular motors that can be envisioned working together, with processive movement of the polymerase possibly providing directionality to subsequent chromatin modifications. Resulting de novo DNA methylation by DRM2, which is predicted to contribute to aberrant RNA production, would provide for positive feedback in our model (Figure 7).

As touched upon previously, our observation that NRPD2 signals are severely reduced in *nrpd1b*, more so than in the *nrpd1a* mutant (see Figure 1G and Figure 6), is consistent with previously published immunoblot data (Pontier et al., 2005). Nonetheless, it is surprising given the nearly perfect colocalization of NRPD2 with NRPD1a, as opposed to only ~50% overlap of NRPD2 with NRPD1b (see Figure 3C). Based on these data, one might expect NRPD1a to be most important for NRPD2 stability. To reconcile these findings, we propose that NRPD2 must be able to exchange between Pol IVb and Pol IVa (Figure 7), with NRPD1b interactions somehow more important for the overall stability of the NRPD2 pool.

The idea that incomplete, or otherwise aberrant, transcripts can induce transcriptional silencing at endogenous repeats may have parallels with the silencing of nonproductive human immunoglobulin genes. In this phenomenon, genes whose transcripts contain premature stop codons following V-D-J recombination are transcriptionally silenced (Buhler et al., 2005), indicating a link between nonsense-mediated decay (NMD) and chromatin modification. In *Arabidopsis*, proteins of the exon-joining complex and NMD pathways were identified within the nucleolar proteome, and some were shown to localize as nucleolar dots (Pendle et al., 2005). Whether these proteins colocalize with the siRNA processing centers is unclear at present.

The nucleolus is best known as the site of 45S pre-rRNA transcription and ribosome assembly. However, small-RNA-directed pre-rRNA cleavage, methylation, and pseudouridylation; biogenesis of signal-recognition particle and telomerase small RNAs; tRNA processing by RNase P; and some pre-mRNA processing also take place within the nucleolus (Bertrand et al., 1998; Filipowicz and Pogacic, 2002; Kiss, 2002; Pederson, 1998). Our findings suggest that processing of endogenous nuclear siRNAs, and possibly RISC storage or sequestration, are additional nucleolar functions to be explored.

EXPERIMENTAL PROCEDURES

Mutant Plant Strains

Arabidopsis rdr2-1 and *dcl3-1* were provided by Jim Carrington, *sgs2-1* (alias *sde1*; *rdr6*) was provided by Herve Vaucheret, and *drd1-6* was provided by Tatsuo Kanno and Marjori Matzke. *drm2-1*, *ago4-1*, and *nrpd1b-11* (SALK_029919) were obtained from the Arabidopsis Biological Resource Center. *nrpd1a* and *nrpd2* mutants were described previously (Onodera et al., 2005).

Generation of Transgenic Lines

Full-length genomic sequences including promoters were amplified by PCR from *A. thaliana* Col-0 DNA using Pfu polymerase (Stratagene) and cloned into pENTR/D-TOPO (Invitrogen). NRPD1a primers were 5′-CACCGGTGTCTCACATTCCAAAGTCCCC-3′ (forward) and 5′-CGGGTTTTCGGAGAAACCACC-3′ (reverse). *NRPD1b* primers were 5′-CACCGCGTACTACAAACGGAAACGGTCA-3′ and 5′-TGTCTGCG TCTGGGACGG-3′. Genomic DCL3 was amplified from BAC clone T15B3 using 5′-CACCCCGACCGAAATCCTCATGACCTAA-3′ and 5′-CTTTTGTATTATGACGATCTTGCGGCGC-3′; the CACC added to forward primers allowed directional cloning into the entry vector. Reverse primers eliminated stop codons to allow epitope-tag fusion. Genes were recombined into pEarleyGate 302 (Earley et al., 2006) to add C-terminal FLAG epitopes. RDR2 coding sequences were amplified by RT-PCR using Pfx Platinum DNA polymerase (Invitrogen) and primers 5′-CACCATGGTGTCAGAGACGACGAC-3′ and 5′-GGGCAATCAAAT GGATACAAGTCC-3′. PCR products captured in pENTR/D-TOPO were recombined into pEarleyGate 104 (Earley et al., 2006), fusing RDR2 sequences C-terminal to YFP expressed from a CaMV 35S promoter. Transformation of constructs into corresponding homozygous mutants was by the floral dip method (Clough and Bent, 1998).

Southern Blotting and Small-RNA Blot Hybridization

Genomic DNA (250 ng) digested with HaeIII or HpaII was subjected to agarose gel electrophoresis, blotted to nylon membranes, and hybridized to a 5S gene probe as described previously (Onodera et al., 2005). Generation of RNA probes labeled with [α-^{32}P]CTP and small-RNA blot hybridization were also as described previously (Onodera et al., 2005). Specific oligodeoxynucleotides used in T7 polymerase reactions (CCTGTCTC hybridized to the T7 promoter adaptor) were as follows: 45S siRNA: 5′-CAATGTCTGTTGGTGCCAAGAGGGAAAAG

GGCCCTGTCTC-3′; 45S prec: 5′-AGTCCGTGGGGAACCCCCTTTT TCGGTTCGCCCCTGTCTC-3′; 5S siRNA: 5′-AGACCGTGAGGCCAA ACTTGGCATCCTGTCTC-3′; *Copia*: 5′-TTATTGGAACCCGGTTAGG ACCTGTCTC-3′, and miR163: 5′-TTGAAGAGGACTTGGAACTTCG ATCCTGTCTC-3′.

Antibodies

Rabbit antibodies raised against NRPD2 and Pol II second-largest-subunit peptides were described previously (Onodera et al., 2005). Chicken antibodies recognizing DCL3, NRPD1a, NRPD1b, or RDR2 were generated against peptides conjugated to keyhole limpet hemocyanin. Peptides were as follows: DCL3: SLEPEKMEEGGGSNC; NRPD1a: EELQVPVGTLTSIGC; NRPD1b: MEEESTSEILDGEIC; RDR2: ETTTNRSTVKISNVC; DRD1: NKNVHKRKQNQVDDGC. Immunolocalization was performed using 1:200 dilutions of antisera, except that NRPD1b antiserum was diluted 1:500. FLAG-tagged proteins were detected using mouse monoclonal anti-FLAG antibody (Sigma-Aldrich) diluted 1:400. RDR2-YFP was detected using mouse anti-GFP/YFP (BD Biosciences) diluted 1:500.

Immunolocalization

Leaves from 28-day-old plants were harvested and nuclei were extracted as described previously (Onodera et al., 2005). After postfixation in 4% paraformaldehyde/PBS (phosphate-buffered saline), washes in PBS, and blocking at 37ºC, slides were exposed overnight to primary antisera in PBS and 0.5% blocking reagent (Roche). After washes in PBS, slides were incubated at 37ºC with anti-mouse-FITC diluted 1:100 (Sigma), goat anti-chicken Alexa 488 diluted 1:300 (Molecular Probes), or goat anti-chicken Alexa 543 diluted 1:400 (Molecular Probes). Nuclei were counterstained with 1 μg/ml DAPI (Sigma) in Vectashield (Vector Laboratories).

Immunoprecipitation and Immunoblotting of Epitope-Tagged Proteins

Pol IV immunoprecipitation was performed using protein extracted from 2.0 g of tissue according to Baumberger and Baulcombe (2005), except that homogenates were filtered through two layers of Miracloth and subjected to centrifugation at 16,000 × g for 15 min at 4ºC prior to incubation with anti-FLAG M2 affinity gel (Sigma). Proteins eluted in 2× SDS-PAGE loading buffer at 100ºC for 2 min were fractionated on 7.5% Tris-glycine SDS-polyacrylamide gels (Cambrex) and electroblotted to PVDF membranes (Millipore). Membranes incubated with peroxidase-linked anti-FLAG M2 antibody diluted 1:2000 (Sigma) were visualized using chemiluminescence detection (Amersham). Membranes were then stripped using 25 mM glycine-HCl (pH 2.0), 1% (w/v) SDS for 30 min with agitation, followed by two 10 min washes in Tris-buffered saline, 0.05% (v/v) Tween 20. NRPD2 immunoblotting was as described in Onodera et al. (2005).

For coimmunoprecipitation of AGO4 and siRNAs, flowers (0.7 g) frozen in liquid nitrogen were homogenized in 2 ml of IP buffer (50 mM Tris-Cl [pH 7.5], 150 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40) containing fresh DTT (2 mM), PMSF (1 mM), pepstatin (0.7 μg/ml), MG132 (10 μg/ml), and Complete protease inhibitor cocktail (Roche). Following centrifugation, lysates precleared with Protein G-agarose beads (Pierce) for 1 hr at 4ºC were incubated with anti-Myc (Upstate) diluted 1:250 for 3 hr at 4ºC. Antibody-antigen complexes were captured on Protein G-agarose (60 μl) at 4ºC for 2 hr and washed four times with IP buffer. For siRNA detection, beads were treated with Proteinase K and extracted sequentially with TE containing 1.5%, 0.5%, or 0.1% SDS. Pooled supernatants extracted with phenol: chloroform (1:1) followed by chloroform were ethanol precipitated. Total siRNAs and RNA blots were prepared and hybridized as previously described (Mette et al., 2000; Zilberman et al., 2003). DNA probes were used to detect 5S siRNAs, 45S siRNAs, miR157, and miR163; RNA probes were used to detect *AtSN1* and *Copia* siRNAs. Probe sequences were as follows: 5S siRNA: 5′-ATGCCAAGTTTGGCCTC ACGGTCT-3′; 45S siRNA: 5′-GTCTGTTGGTGCCAAGAGGGAAAAG GGCTAAT-3′; *AtSN1*: 5′-ACCAACGTGTTGTTGGCCCAGTGGTAAA TCTCTCAGATAGAGG-3′; *Copia*: 5′-TTATTGGAACCCGGTTAGGA-3′; miR159: 5′-TAGAGCTCCCTTCAATCCAAA-3′; miR163: 5′-ATCGA AGTTGGAAGTCCTCTTCAA-3′.

RNA and DNA In Situ Hybridization

RNA probes were labeled by in vitro T7 polymerase (Ambion) transcription with digoxigenin-11-UTP or biotin-16-UTP RNA labeling mix (Roche). RNA in situ hybridization was carried out at 42ºC overnight using a probe solution containing 1 μg RNA probe, 5 μg yeast tRNA (Roche), 50% dextran sulfate, 100 mM PIPES [pH 8.0], 10 mM EDTA, and 3 M NaCl as described previously (Highett et al., 1993). Slides were washed sequentially in 2× SSC, 50% formamide, 50ºC followed by 1× SSC, 50% formamide, 50ºC, then 1× SSC 20ºC, and finally TBS at 20ºC. Where applicable, nuclei were incubated at 37ºC for 30 min in a solution of RNase-free DNase I (0.015 U/μl) or in a solution of RNase A (100 μg/ml, Roche). Nuclease reactions were stopped in 10 mM EDTA (pH 7.5) for 2 min followed by three washes in 0.1× SSC.

DNA-FISH using 5S or 45S rRNA gene probes labeled with biotin-dUTP or digoxigenin-dUTP was performed as described (Pontes et al., 2003). Digoxigenin-labeled probes were detected using mouse anti-digoxigenin antibody (1:250, Roche) followed by rabbit anti-mouse antibody conjugated to Alexa 488 (Molecular Probes). Biotin-labeled probes were detected using goat anti-biotin conjugated with avidin (1:200, Vector Laboratories) followed by streptavidin-Alexa 543 (Molecular Probes). DNA was counterstained with DAPI (1 μg/ml) in Vectashield (Vector Laboratories). For dual protein/nucleic acid localization experiments, slides were first subjected to immunofluorescence, then postfixed in 4% formaldehyde/PBS followed by RNA- or DNA-FISH.

Microscopy

Nuclei were routinely examined using a Nikon Eclipse E800i epifluorescence microscope, with images collected using a Photometrics Coolsnap ES Mono digital camera. The images were pseudocolored, merged, and processed using Adobe Photoshop (Adobe Systems). Multiphoton optical-section stacks were collected using a Zeiss LSM 510 Meta microscope. Single optical sections using 40× averaging were acquired by simultaneous scanning to avoid artifactual shift between two optical channels. The 488 nm line of an argon laser was used for detection of FITC FLAG-tagged proteins, and the 543 nm line of a helium-neon laser was used for detection of Alexa 543 siRNA signals. For the detection of DAPI, either a 715 or 750 nm multiphoton tuned titanium-sapphire laser was used. Projections of 3D data stacks were composed using Imaris 4.1 software from Bitplane (http://www.bitplane.com).

Supplemental Data

Supplemental Data include three figures and six tables and can be found with this article online at http://www.cell.com/cgi/content/full/126/1/79/DC1/.

ACKNOWLEDGMENTS

O.P. performed all microscopy, P.C.N. generated siRNA blots, T.R. performed DNA methylation assays, and J.H. generated epitope-tagged Pol IV lines and coIP data. YFP-RDR2 and FLAG-DCL3 lines were generated by O.P. and A.V., respectively; C.S.P. wrote the manuscript. C.F.L. and S.E.J. generated the Myc-AGO4 transgenic line and contributed Figure 4B. O.P. and P.C.N. were supported by fellowships SFRH/BPD/17508/2004 and SFRH/BD/6520/2001, respectively, from the Fundação para a Ciência e Tecnologia (Portugal). Pikaard lab work was supported by NIH grants R01GM60380 and R01GM077590 and the Monsanto Company/Washington University Biology Research Agreement. C.F.L. was supported by Ruth L. Kirschstein National Research Service Award GM07185. S.E.J. is a Howard Hughes Medical

Institute Investigator. Work in the Jacobsen laboratory was supported by NIH grant GM60398. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of NIH, HHMI, or the Monsanto Company. We thank Howard Berg (Donald Danforth Plant Science Center) for multiphoton microscopy training, Tom Juehne and Keming Song (Sigma-Aldrich Company) for Pol IV antibody production, Dr. Wanda Viegas (Instituto Superior de Agronomia, Portugal) for comentoring P.C.N., and Eric Richards (Washington University) for suggestions to improve the manuscript.

Received: March 1, 2006
Revised: March 31, 2006
Accepted: May 17, 2006
Published: July 13, 2006

REFERENCES

Almeida, R., and Allshire, R.C. (2005). RNA silencing and genome regulation. Trends Cell Biol. *15*, 251–258.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, A.J., and Matzke, M. (2002). RNA-directed DNA methylation in Arabidopsis. Proc. Natl. Acad. Sci. USA *99*, 16499–16506.

Baulcombe, D. (2004). RNA silencing in plants. Nature *431*, 356–363.

Baumberger, N., and Baulcombe, D.C. (2005). Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. Proc. Natl. Acad. Sci. USA *102*, 11928–11933.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.

Bertrand, E., Houser-Scott, F., Kendall, A., Singer, R.H., and Engelke, D.R. (1998). Nucleolar localization of early tRNA processing. Genes Dev. *12*, 2463–2468.

Boudonck, K., Dolan, L., and Shaw, P.J. (1999). The movement of coiled bodies visualized in living plant cells by the green fluorescent protein. Mol. Biol. Cell *10*, 2297–2307.

Buhler, M., Mohn, F., Stalder, L., and Muhlemann, O. (2005). Transcriptional silencing of nonsense codon-containing immunoglobulin minigenes. Mol. Cell *18*, 307–317.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr. Biol. *13*, 2212–2217.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Cioce, M., and Lamond, A.I. (2005). Cajal bodies: a long history of discovery. Annu. Rev. Cell Dev. Biol. *21*, 105–131.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. *16*, 735–743.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K., and Pikaard, C.S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. *45*, 616–629.

Filipowicz, W., and Pogacic, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. Curr. Opin. Cell Biol. *14*, 319–327.

Gasciolli, V., Mallory, A.C., Bartel, D.P., and Vaucheret, H. (2005). Partially redundant functions of Arabidopsis DICER-like enzymes and a role for DCL4 in producing trans-acting siRNAs. Curr. Biol. *15*, 1494–1500.

Grewal, S.I., and Rice, J.C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. Curr. Opin. Cell Biol. *16*, 230–238.

Hannon, G.J. (2002). RNA interference. Nature *418*, 244–251.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118–120.

Highett, M.I., Beven, A.F., and Shaw, P.J. (1993). Localization of 5 S genes and transcripts in Pisum sativum nuclei. J. Cell Sci. *105*, 1151–1158.

Kanno, T., Mette, M.F., Kreil, D.P., Aufsatz, W., Matzke, M., and Matzke, A.J. (2004). Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. Curr. Biol. *14*, 801–805.

Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. Nat. Genet. *37*, 761–765.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell *109*, 145–148.

Krogh, B.O., and Symington, L.S. (2004). Recombination proteins in yeast. Annu. Rev. Genet. *38*, 233–271.

Li, C.F., Pontes, O., El-Shami, M., Henderson, I.R., Bernatavichute, Y.V., Chan, S.W.-L., Lagrange, T., Pikaard, C.S., and Jacobsen, S.E. (2006). An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. Cell *126*, this issue, 93–106.

Lippman, Z., May, B., Yordan, C., Singer, T., and Martienssen, R. (2003). Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol. *1*, E67.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. *19*, 5194–5201.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613–622.

Pederson, T. (1998). The plurifunctional nucleolus. Nucleic Acids Res. *26*, 3871–3876.

Pendle, A.F., Clark, G.P., Boon, R., Lewandowska, D., Lam, Y.W., Andersen, J., Mann, M., Lamond, A.I., Brown, J.W., and Shaw, P.J. (2005). Proteomic analysis of the Arabidopsis nucleolus suggests novel nucleolar functions. Mol. Biol. Cell *16*, 260–269.

Pontes, O., Lawrence, R.J., Neves, N., Silva, M., Lee, J.H., Chen, Z.J., Viegas, W., and Pikaard, C.S. (2003). Natural variation in nucleolar dominance reveals the relationship between nucleolus organizer chromatin topology and rRNA gene transcription in Arabidopsis. Proc. Natl. Acad. Sci. USA *100*, 11418–11423.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev. *19*, 2030–2040.

Rountree, M.R., and Selker, E.U. (1997). DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. Genes Dev. *11*, 2383–2395.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434–1437.

Sontheimer, E.J., and Carthew, R.W. (2004). Molecular biology. Argonaute journeys into the heart of RISC. Science *305*, 1409–1410.

Tomari, Y., and Zamore, P.D. (2005). Perspective: machines for RNAi. Genes Dev. *19*, 517–529.

Vaucheret, H. (2005). RNA polymerase IV and transcriptional silencing. Nat. Genet. *37*, 659–660.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672–676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833–1837.

Wassenegger, M. (2005). The role of the RNAi machinery in heterochromatin formation. Cell *122*, 13–16.

Wassenegger, M., and Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. Trends Plant Sci. *11*, 142–151.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. *2*, E104.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr. Biol. *14*, 1214–1220.

Supplemental Data

The *Arabidopsis* Chromatin-Modifying Nuclear siRNA Pathway Involves a Nucleolar RNA Processing Center

Olga Pontes, Carey Fei Li, Pedro Costa Nunes, Jeremy Haag, Thomas Ream, Alexa Vitins, Steven E. Jacobsen, and Craig S. Pikaard

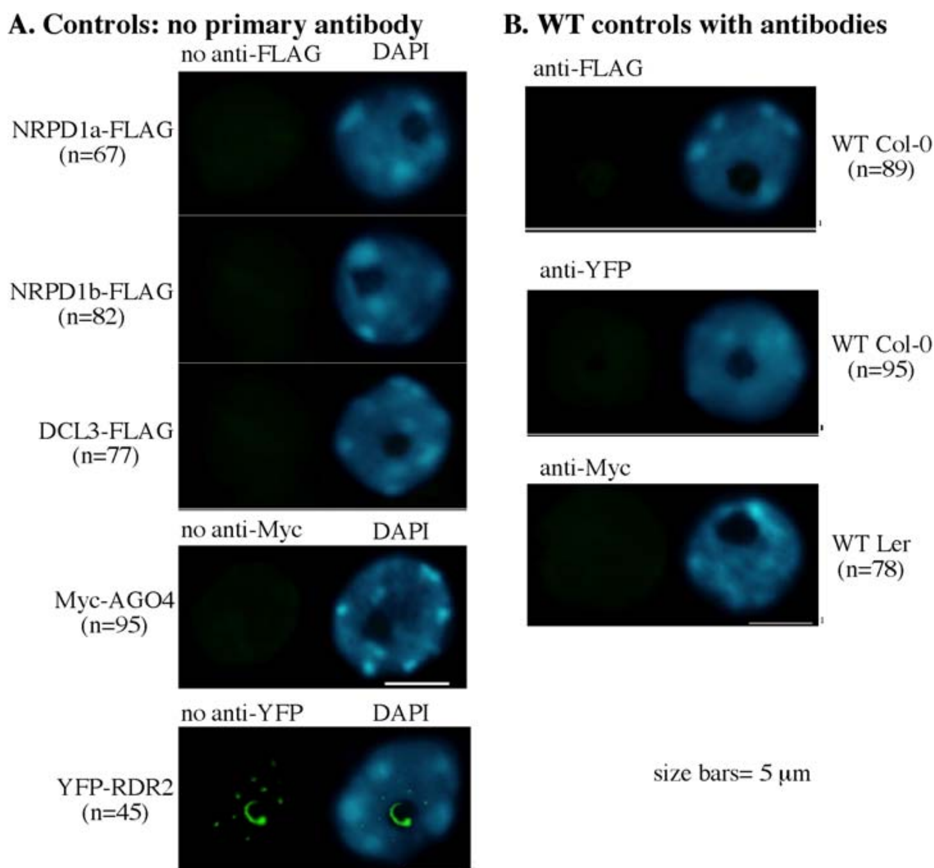


Figure S1. Antibody Specificity Controls

In part A of the figure, nuclei of transgenic lines expressing the indicated epitope tagged proteins were processed for protein immunolocalization as in Figure 3 of the paper except that the primary antibody was omitted prior to incubation with FITC-labeled secondary antibody (green). YFP fluorescence accounts for the YFP-RDR2 signal in the absence of anti-YFP antibody. In part B, non-transgenic, wild-type *A. thaliana* (ecotypes Col-0 or Ler) controls show that no signals are obtained upon immunolocalization using anti-FLAG, anti-YFP or anti-Myc primary antibodies. The images shown are representative of the nuclei observed, with the total number analyzed shown in parentheses. Nuclei were counterstained with DAPI (blue); the size bar corresponds to 5µm.

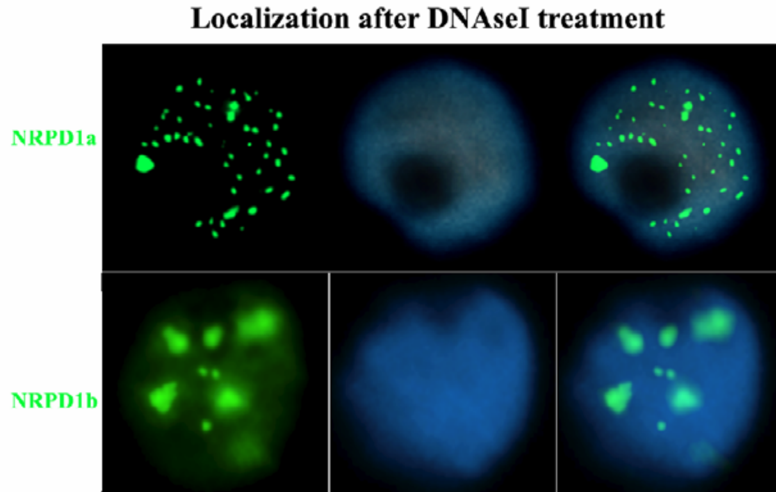


Figure S2. NRPD1a and NRPD1b Immunolocalization Signals Are Not Lost in DNase I-Treated Nuclei

Native NRPD1a and NRPD1b proteins were localized using anti-peptide antibodies in nuclei treated with DNase I as described in Figure 3B of the main paper.

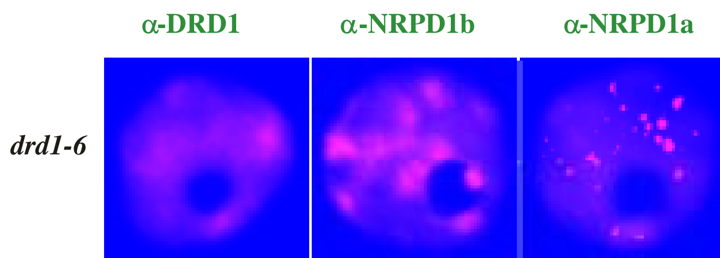


Figure S3. Immunolocalization of DRD1, NRPD1b, and NRPD1a in *drd1-6* Mutant Nuclei

Proteins were detected using anti-peptide antibodies. Note that DRD1 is not detected in the mutant, suggesting that the antibody specifically recognizes DRD1. The *drd1-6* mutation typically does not affect the NRPD1a pattern (85% yield the wild-type pattern for NRPD1a shown below; n = 90) but NRPD1b immunolocalization signals are typically more diffuse in drd1-6 (79%; n = 79) than in wild-type, suggesting that DRD1 may act upstream, or at the same step, as NRPD1b.

Table S1. Supporting Data for Figure 2A: siRNA Probe Hybridization Patterns and Frequencies

		Frequency (%) of phenotypes observed upon nuclease treatment or in different genetic backgrounds									
RNA probe	**Localization phenotypes**	**Col**	**Ler**	**+RNase A**	**+DNase I**	***nrpd1a***	***nrpd2***	***nrpd1b***	***rdr2-1***	***dcl3-1***	***ago4-1***
45S siR	**Nucleolar dot observed:**	100	100	0	100	0	0	0	0	0	0
	Dispersed nuclear signal:	0	0	0	0	29	13	56	8	9	29
	No signal:	0	0	100	0	71	87	44	92	91	71
	# nuclei observed	n = 75	n = 71	n = 71	n = 63	n = 65	n = 141	n = 132	n = 62	n = 72	n = 76
5S siR	**Nucleolar dot observed:**	100	100	0	100	0	0	0	0	0	0
	Dispersed nuclear signal:	0	0	0	0	0	6	75	11	3	17
	No signal:	0	0	100	0	100	94	25	89	97	83
	# nuclei observed	n = 56	n = 48	n = 62	n = 68	n = 81	n = 127	n = 162	n = 85	n = 62	n = 74
AtSN1	**Nucleolar dot + nucleoplasm:**	74	No data	0	89	No data	0	No data	0	No data	No data
	Nucleoplasm only:	26		0	11		0		0		
	No signal:	0		100	0		100		100		
	# nuclei observed	n = 67		n = 79	n = 85		n = 150		n = 123		
AtCopia4	**Nucleolar dot +nuclear spots:**	100	No data	0	100	No data	0	No data	0	No data	No data
	No signal:	0		100	0		100		100		
	# nuclei observed	n = 85		n = 53	n = 68		n = 103		n = 91		
45S precursor	**Diffuse nucleolar signals:**	100	100	100	100	100	100	100	100	100	100
	# nuclei observed	n = 63	n = 57	n = 64	n = 51	n = 86	n = 79	n = 127	n = 72	n = 74	n = 81

The table is organized as in Figure 2A except that the table includes two columns of data for wild-type nuclei (ecotypes Col-0 and Ler) whereas Figure 2A showed only the Col-0 wild-type control.

Table S2. Supporting Data for Figure 3A: Protein Localization and Effects of RNase

	NRPD1a	**NRPD2**	**NRPD1b**	**RDR2**	**DCL3**	**AGO4**
protein localization	100% of nuclei display pattern shown *n* = 82	100% of nuclei display pattern shown *n* = 245	100% of nuclei show the nucleolar dot. 57% display numerous puncta external to nucleolus, as shown; 43% show <10 puncta *n* = 77	100% of nuclei display pattern shown *n* = 87	100% of nuclei display pattern shown *n* = 125	100% of nuclei display pattern shown *n* = 96
Effect of RNase A	91% , protein not detectable 9% , WT pattern *n* = 85	81% , protein not detectable 19% , WT pattern *n* = 94	65% , protein not detectable 35% , WT pattern *n* = 93	85% , protein not detectable 15% , WT pattern *n* = 62	59% , protein not detectable 41% , WT pattern *n* = 89	72% , protein not detectable 28% , WT pattern *n* = 61

Table S3. Supporting Data for Figure 3C: Pairwise Detection of Nuclear siRNA Pathway Proteins

	Antibodies				
Epitope-tagged lines	**α-NRPD1a**	**α-NRPD2**	**α-NRPD1b**	**α-RDR2**	**α-DCL3**
NRPD1a-FLAG		Majority of the nucleoplasmic signals colocalized *n* = 93			
NRPD1b-FLAG	Few nucleoplasmic signals colocalized *n* = 71	Few nucleoplasmic signals colocalized *n* = 85			
YFP-RDR2	Few nucleoplasmic signals colocalized *n* = 54	Few nucleoplasmic signals colocalized *n* = 48	Nucleolar dot + Few nucleoplasmic signals colocalized *n* = 67		
DCL3-FLAG	Few nucleoplasmic signals colocalized *n* = 76	Few nucleoplasmic signals colocalized *n* = 81	Nucleolar dot + Few nucleoplasmic signals colocalized *n* = 73	Nucleolar dot + Few nucleoplasmic signals colocalized *n* = 86	
Myc -AGO4	Not colocalized *n* = 54	Few nucleoplasmic signals colocalized *n* = 61	Nucleolar dot + Few nucleoplasmic signals colocalized *n* = 58	Nucleolar dot colocalized *n* = 45	Nucleolar dot colocalized *n* = 59

Table S4. Supporting Data for Figure 4: Protein-siRNA Colocalization

RNA probes		Epitope-tagged lines: NRPD1b-Flag	YFP-RDR2	DCL3-Flag	cMyc-AGO4
45S siR	**Colocalized**	81%	82%	79%	91%
	Not colocalized	19% *n* = 46	18% *n* = 60	21% *n* = 75	9% *n* = 65
siR1003	**Colocalized**	76%	58%	85%	76%
	Not colocalized	24% *n* = 57	42% *n* = 72	15% *n* = 79	24% *n* = 57
AtSN1	**Colocalized**	85%	61%	76%	83%
	Not colocalized	15% *n* = 74	39% *n* = 56	34% *n* = 45	17% *n* = 56
AtCopia4	**Colocalized**	82%	54%	78%	72%
	Not colocalized	18% *n* = 57	46% *n* = 59	22% *n* = 49	28% *n* = 67
45S prec	**Colocalized**	25%	43%	21%	30%
	Not colocalized	75% *n* = 81	57% *n* = 64	79% *n* = 61	70% *n* = 75

Colocalization was considered to be when >50% of the RNA probe signal overlapped >50 % of the protein signal.

Table S5. Supporting Data for Figure 5: Localization of Proteins Relative to NORs and 5S Gene Loci

DNA loci		NRPD1a	NRPD2a	NRPD1b	RDR2	DCL3	DRD1
NORs	**Colocalized**	85%	93%	92%	22%	12%	87%
	Not colocalized	15% *n* = 71	7% *n* = 83	8% *n* = 89	78% *n* = 55	88% *n* = 66	13% *n* = 57
5S gene clusters	**Colocalized**	68%	72%	81%	13%	27%	72%
	Not colocalized	32% *n* = 58	28% *n* = 62	19% *n* = 76	87% *n* = 51	73% *n* = 65	28% *n* = 61

Colocalization was considered to be when at least two NORs and at least four 5S gene *loci* overlapped half of the protein signals outside the nucleolus.

Table S6. Supporting Data for Figure 6: Protein Localization in Various Nuclear siRNA Pathway Mutants

		NRPD2	NRPD1a	NRPD1b	RDR2	DCL3
WT	**Col**	100% of nuclei display pattern shown *n* = 245	100% of nuclei display pattern shown *n* = 160	71% of nuclei display pattern shown *n* = 185	77% of nuclei display pattern shown n = 96	100% of nuclei display pattern shown n = 125
Mutants	***nrpd1a***	Reduction in labeling intensity *n* = 181	No signal *n* = 123	WT pattern *n* = 87	Very faint to no signal *n* = 145	Very faint to no signal *n* = 61
	nrpd2	Not detected *n* = 155	Reduction in labeling intensity *n* = 178	Very faint to no signal *n* = 134	Very faint to no signal *n* = 141	Very faint to no signal *n* = 104
	nrpd1b	Very faint to no signal *n* = 138	WT pattern *n* = 67	No signal *n* = 149	Nucleolar dot is not detected *n* = 153	- Very strong reduction in labeling intensity (76%) - Mislocalization of the nucleolar dot to the nucleoplasm (24%) *n* = 84
	nrpd2, nrpd1a	Very faint to no signal *n* = 74	Very faint to no signal *n* = 81	Very faint to no signal *n* = 90	Very faint to no signal *n* = 67	Very faint to no signal *n* = 57
	rdr2-1	Small reduction in labeling intensity *n* = 121	WT pattern *n* = 112	- Nucleolar dot not detected (81%) - Reduction in labeling intensity (19%) *n* = 157	No signal *n* = 61	Very faint to no signal *n* = 87
	dcl3-1	Small reduction in labeling intensity *n* = 130	WT pattern *n* = 74	- Nucleolar dot not detected (78%) - Reduction in labeling intensity (22%) *n* = 72	WT pattern *n* = 89	No signal *n* = 91
	ago4-1	Small reduction in labeling intensity *n* = 109	WT pattern *n* = 65	- Nucleolar dot not detected (92%) - Reduction in labeling intensity (8%) *n* = 133	WT pattern *n* = 122	- WT pattern (67%) - Mislocalization of the nucleolar dot to the nucleoplasm (33%) *n* = 152

APPENDIX C

SEX-BIASED LETHALITY OR TRANSMISSION OF DEFECTIVE TRANSCRIPTION MACHINERY IN ARABIDOPSIS

Published in *Genetics*, vol. 180;1 207-18

My contributions to this work:

I performed the genotyping analysis of the *NRPB12* T-DNA line to show that this insertion mutation in *NRPB12* is lethal (data not shown). I also performed reciprocal crosses between heterozygotes and wild-type plants to show that the *NRPB12* T-DNA is lethal to the female gametophyte, but that the T-DNA is transmitted through the male gametophyte (Table S1). I also made critical comments on the manuscript.

Copyright © 2008 by the Genetics Society of America
DOI: 10.1534/genetics.108.090621

Sex-Biased Lethality or Transmission of Defective Transcription Machinery in Arabidopsis

Yasuyuki Onodera,*[,1] Kosuke Nakagawa,*[,1] Jeremy R. Haag,† Diane Pikaard,† Tetsuo Mikami,* Thomas Ream,† Yusuke Ito* and Craig S. Pikaard†[,2]

**Division of Applied Bioscience, Research Faculty of Agriculture, Hokkaido University, Kita 9, Nishi 9, Kita-ku, Sapporo 060-8589, Japan and* †*Biology Department, Washington University, St. Louis, Missouri 63130*

Manuscript received April 23, 2008
Accepted for publication June 24, 2008

ABSTRACT

Unlike animals, whose gametes are direct products of meiosis, plant meiotic products undergo additional rounds of mitosis, developing into multicellular haploid gametophytes that produce egg or sperm cells. The complex development of gametophytes requires extensive expression of the genome, with DNA-dependent RNA polymerases I, II, and III being the key enzymes for nuclear gene expression. We show that loss-of-function mutations in genes encoding key subunits of RNA polymerases I, II, or III are not transmitted maternally due to the failure of female megaspores to complete the three rounds of mitosis required for the development of mature gametophytes. However, male microspores bearing defective polymerase alleles develop into mature gametophytes (pollen) that germinate, grow pollen tubes, fertilize wild-type female gametophytes, and transmit the mutant genes to the next generation at moderate frequency. These results indicate that female gametophytes are autonomous with regard to gene expression, relying on transcription machinery encoded by their haploid nuclei. By contrast, male gametophytes make extensive use of transcription machinery that is synthesized by the diploid parent plant (sporophyte) and persists in mature pollen. As a result, the expected stringent selection against nonfunctional essential genes in the haploid state occurs in the female lineage but is relaxed in the male lineage.

IN flowering plants, three rounds of postmeiotic mitosis and development give rise to an eight-nucleate female gametophyte, one cell of which is the egg cell (Schneitz *et al.* 1995; Grossniklaus and Schneitz 1998; Drews and Yadegari 2002). Pollen, the male gametophyte, consists of three haploid cells, two of which are sperm cells. The three pollen cells are clonally related and are all descended from a single haploid meiotic product of a pollen mother cell (McCormick 1993, 2004). The male gametophyte can survive independent of the sporophyte (the parent plant) and upon landing on a receptive flower, the pollen germinates and develops a pollen tube that elongates through the transmitting tract of the pistil, the female floral organ, to reach the ovary. Within the ovary, the pollen tube grows toward chemical signals emanating from the two synergid cells of the female gametophyte (Higashiyama 2002; Higashiyama *et al.* 2001, 2003; Johnson and Preuss 2002). Upon reaching a synergid cell, adjacent to the egg, the pollen tube ruptures, releasing the sperm. One sperm cell fuses with the egg to give rise to the diploid embryo. The second sperm cell fuses with the female gametophyte's central cell, giving rise to the endosperm. Proper development of both embryo and endosperm as a result of double fertilization is required for seed maturation (Russell 1993; Grossniklaus and Schneitz 1998; Yadegari *et al.* 2000).

Large-scale analyses of cDNA libraries generated from mRNAs purified from maize and wheat female gametophytes have shown that thousands of genes are expressed in female gametophytes (Sprunck *et al.* 2005; Yang *et al.* 2006). Comparative microarray-based transcript profiling analyses using ovules of Arabidopsis wild-type plants and mutants lacking embryo sacs have similarly identified large numbers of female gametophyte-specific genes (Yu *et al.* 2005; Johnston *et al.* 2007; Jones-Rhoades *et al.* 2007; Steffen *et al.* 2007). Collectively, expression-profiling studies combined with analyses of female gametophytic mutants (Pagnussat *et al.* 2005) provide evidence for extensive transcriptional regulatory networks that are critical for the proper development of female gametophytes.

In Arabidopsis, ~62% of all genes in the genome are expressed during at least one stage of male gametophyte development, with ~10% of these transcripts being pollen specific (Honys and Twell 2003, 2004). Moreover, labeled UTP is incorporated into RNA in pollen and the transcription inhibitor, actinomycin D inhibits

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Department of Biology, Washington University, Campus Box 1137, 1 Brookings Dr., St. Louis, MO 63130.
E-mail: pikaard@biology2.wustl.edu

pollen tube growth (Mascarenhas 1989, 1993; Honys and Twell 2004). These observations indicate that male gametophytes are actively engaged in the transcription of their haploid genomes.

The enzymes central to nuclear gene expression are DNA-dependent RNA polymerases I, II, and III (Pol I, Pol II, and Pol III), each of which is composed of between 12 and 17 subunits. Pol I is responsible for transcribing the 45S preribosomal RNAs (rRNAs) that are then processed into the 18S, 5.8S, and 25–28S (the latter size depends on the species) rRNAs that form the catalytic core of ribosomes. Pol II transcribes messenger RNAs (mRNAs) as well as RNAs that do not encode proteins, such as micro RNAs and small nuclear RNAs that guide mRNA and rRNA processing events. Pol III is primarily responsible for transcribing transfer RNAs (tRNAs) and repetitive 5S rRNA genes (Kassavetis *et al.* 1994; Paule and White 2000).

For purposes of gene and subunit nomenclature, Arabidopsis Pol I is denoted as nuclear RNA polymerase A (NRPA), Pol II is denoted as NRPB, and Pol III is denoted as NRPC. Their second-largest subunits, denoted as NRPA2, NRPB2, and NRPC2, respectively, are homologs of the β-subunits of eubacterial RNA polymerase. Together with the largest subunits, the β-like second-largest subunits help form the active sites of the enzymes and are essential for RNA synthesis. In *Arabidopsis thaliana*, the Pol I, Pol II, and Pol III second-largest subunits are encoded by single-copy genes located on chromosomes 1, 4, and 5, respectively (Larkin and Guilfoyle 1993; Onodera *et al.* 2005); see also phylogenetic analyses by Craig S. Pikaard and Jonathan Eisen discussed in Arabidopsis Genome Initiative (2000).

Contrary to our expectation that loss-of-function mutations in *NRPA2*, *NRPB2*, or *NRPC2* genes would be unrecoverable due to lethality in both the haploid male and female gametophytes, transgenic lines hemizygous for T-DNA disruptions of each gene can be identified and maintained. Detailed analysis of these lines revealed that the mutant RNA polymerase alleles are not transmitted through the female lineage due to the failure of mutant female gametophytes to complete their development. By contrast, the mutant alleles are transmitted to subsequent generations through the male gametophyte at moderate efficiency compared to wild type. Our data indicate that pollen can develop to maturity, grow pollen tubes, and carry out fertilization in the absence of functional RNA polymerase genes, apparently by utilizing transcription machinery synthesized premeiotically in pollen mother cells. By contrast, female gametophyte development is autonomous and requires transcription machinery generated *de novo* in the haploid state.

MATERIALS AND METHODS

Plant strains and growth conditions: *Arabidopsis thaliana* wild-type and T-DNA insertion mutants (ecotype Columbia in both cases) were grown at 22° with a 16-hr photoperiod. Gene locus identifiers for *NRPA2*, *NRPB2*, and *NRPC2* are At1g29940, At4g21710, and At5g45140, respectively. The T-DNA insertion alleles we named *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, and *nrpb2-2* are carried within Torrey Mesa Research Institute (San Diego) transgenic lines: GARLIC_726_H01, GARLIC_918_C10, GARLIC_859_B04, and GARLIC_110_G08, respectively [GARLIC is the former name of the Syngenta Biotechnology's SAIL collection of T-DNA lines, available from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University]. The parental line for GARLIC_110_G08 was homozygous for the *qrt1-2* allele of the *QUARTET* gene (ecotype Columbia) (Preuss *et al.* 1994); other GARLIC lines are wild type at the *QRT* locus. The T-DNA allele *nrpc2-1* is present in Salk line 007865 (Alonso *et al.* 2003) obtained from the ABRC. Seeds of plants bearing the *nrpc2-2* (GABI_131_B09) allele were obtained from GABI-Kat (Rosso *et al.* 2003). The transgenic Arabidopsis line (SAIL _100_H07) carrying a *LAT52*::*GUS* reporter gene(s) inserted in an intergenic region was obtained from ABRC.

Genotyping: To identify T-DNA disrupted alleles in segregating families, PCR was carried out using primers complementary to the T-DNA left border (5′-GCATCTGAATTTCA TAACCAATCTC-3′, 5′-CGTCCGCAATGTGTTATTAAG-3′, or 5′-CCCATTTGGACGTGAATGTAGACAC-3′) and primers specific for *NRPA2* (5′-AGAGAGGTAGAGAAACTCACG-3′ or 5′-ATAAACAGTTAGGCAAGCGAA-3′), *NRPB2* (5′-CGATTTGAG CTTCTACCGTTT-3′ or 5′-CCTAGAACATACCATGCGAAA-3′) or *NRPC2* (5′-CTCGCACAATGAAGGATGTTT-3′ or 5′-TAATTC TTGCCGCAAATTGAC-3′). Wild-type alleles of *NRPA2*, *NRPB2*, and *NRPC2* were identified using the gene-specific primers above in combination with 5′-GATGAGTTGGATAACACGA AC-3′ or 5′-AGCACCCTTTAAGCTACAAAG-3′ for *NRPA2*; 5′-CCATCAGACTCTGTCATCATA-3′ or 5′-ACGAAGGGTAA GCATGCAGTT-3′ for *NRPB2*; and 5′-AGCTACTCCAGGGGA GATTAT-3′ or 5′-GGCAAGTACTATAGCCCCCTG-3′ for *NRPC2*.

The unique genomic DNA/T-DNA junction sequences at both ends of the single T-DNA loci in *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles were amplified by PCR and verified by sequencing.

Production of transgenic plants: Genomic sequences for *NRPA2* (positions −1433 to +7346 relative to the translation start site), *NRPB2* (positions −338 to +6514), or *NRPC2* (positions −1947 to +10295) were amplified by PCR. Amplified gene sequences included promoter regions and all introns and exons. Resulting PCR products were captured in pENTR/D-TOPO and recombined into the Gateway recombination (Invitrogen)-compatible expression vector pEarleyGate 302 (Earley *et al.* 2006). Resulting *NRPA2*, *NRPB2*, or *NRPC2* full-length transgenes were introduced into hemizygous plants bearing a corresponding mutant allele (+/*nrpa2-1*, +/*nrpb2-1*, or +/*nrpc2-1*). Progeny of transgenic plants that were homozygous for the *nrpa2-1*, *nrpb2-1*, or *nrpc2-1* mutations and were rescued by the full-length transgenes were identified by PCR genotyping.

Confocal laser scanning microscopy: Examination of specimens was carried out using a Zeiss LSM confocal microscope system equipped with a Helium/Neon laser. Images were processed using Adobe Photoshop 7.0 software. Floral stages were defined according to Bowman (1994). Developmental stages of female gametophytes were defined according to Christensen *et al.* (1997).

Cytological and histochemical analysis of pollen: *In vitro* pollen germination was carried out as described by Hashida *et al.* (2007). Pollen were stained with 1 μg/ml DAPI in 20 mM Tris-HCl pH 7.65, 0.5 mM EDTA, 1.2 mM spermidine, 7 mM 2-mercaptoethanol, 0.4 mM phenylmethylsulfonyl fluoride, 0.1 mg/ml FDA in 0.5 M sucrose, or Alexander solution (ftp://

ftp.arabidopsis.org/home/tair/Protocols/EMBOmanual/ch1.pdf). Pollen and self-pollinated pistils were incubated at 37° for 12 hr in GUS staining solution (50 mm sodium phosphate pH 7.2, 0.2% Triton X-100, 2 mm potassium ferrocyanide, 2 mm potassium ferricyanide, and 1 mg/ml X-Gluc).

RESULTS

Sex-biased defects in the transmission of mutant alleles encoding RNA polymerase I, II, and III second-largest subunits: We used a PCR-based strategy to verify the existence of T-DNA-disrupted alleles for the catalytic second-largest subunits of RNA polymerase I (alleles *nrpa2-1* and *nrpa2-2*), RNA polymerase II (alleles *nrpb2-1* and *nrpb2-2*), or RNA polymerase III (alleles *nrpc2-1* and *nrpc2-2*) (Figure 1A). We then genotyped the progeny resulting from self-fertilization of plants bearing these alleles. In all cases, individuals that carried a mutant RNA polymerase allele also carried a corresponding wild-type allele (Figure 1B and data not shown), indicating that these plants were hemizygous for the mutations. No plants homozygous for the Pol I (*nrpa2-1*, *nrpa2-2*), Pol II (*nrpb2-1, nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles were recovered, indicating that the alleles are all severe loss-of-function mutations in essential genes, consistent with the essential roles of Pol I, Pol II, and Pol III in nuclear gene expression.

Hemizygotes should outnumber homozygous wild-type siblings 67%:33% (2:1) among the progeny of a hemizygous parent bearing one copy of a defective essential gene, assuming that the homozygous mutant is inviable. However, as shown in Table 1, PCR-based genotyping revealed that only 8–38% of the progeny were hemizygous for Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrpb2-1* or *nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles (Table 1). Instead, the majority of the progeny possessed only wild-type alleles, indicating a defect in the transmission of the mutant RNA polymerase alleles.

To test for sex-biased defects in the transmission of the mutant alleles through the male or female gametophytes, Pol I hemizygotes (+/*nrpa2-1* or +/*nrpa2-2*), Pol II hemizygotes (+/*nrpb2-1* or +/*nrpb2-2, qrt1-2*; the latter is a Pol II mutant hemizygote in a homozygous quartet mutant background), or Pol III hemizygotes (+/*nrpc2-1* or +/*nrpc2-2*) were reciprocally crossed with wild-type (+/+) plants by hand-pollinating emasculated flowers. Resulting progeny were then genotyped by PCR. None of the mutant polymerase alleles were found to be transmitted to the progeny via the maternal parent (Figure 1, C–E; Table 2); instead all progeny of hemizygous (+/−) female plants crossed with wild-type (+/+) males were homozygous wild type (+/+). By contrast, the *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles were all pollen transmissible, such that 13–38% of the progeny inherited a mutant allele from the hemizygous paternal parent when crossed with a wild-type female (Table 2). Note, however, that equal numbers of hemizygous (+/−) and homozygous (+/+) progeny are expected from a (+/+) × (+/−) cross if the wild-type and mutant alleles are transmitted with equal efficiency; the male-transmitted Pol I, II, and III mutant alleles were not inherited at such high levels.

The reciprocal crossing data summarized in Tables 1 and 2 indicate a lack of transmission of the mutant polymerase second-largest subunit alleles through female gametophytes and a partial defect in their transmission through the male gametophyte. Similar allele transmission behavior was observed for the RNA polymerase subunit mutant *nrpb12a* (supplemental Table S1). The homolog of *NRPB12a* in yeast is a single-copy gene whose encoded protein is incorporated into all three nuclear polymerases (Pol I, II, and III). As was the case for the second-largest subunit mutants, homozygous *nrpb12a* mutants were not recoverable. Moreover, *nrpb12a* mutant alleles were transmitted via pollen but not through the female gametophytes. Collectively, our results indicate that male-specific transmissibility of defective RNA polymerase alleles is a general characteristic of RNA polymerase subunit genes and not a peculiarity of second-largest subunit genes.

Defective RNA polymerase alleles cause female gametophyte developmental arrest: Lack of maternal transmission of the Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrpb2-1* or *nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) alleles prompted an examination of siliques (seed pods) of self-pollinated hemizygous +/*nrpa2-1*, +/*nrpa2-2*, +/*nrpb2-1*, +/*nrpc2-1*, +/*nrpc2-2*, or +/*nrpb2-2*, *qrt1-2* plants. Siliques of these plants contain small unfertilized ovules interspersed with an equal number of normal seeds; as an example, a silique from a +/*nrpa2-1* plant is shown in Figure 2A. Whereas wild-type plants produce 51–58 seeds per silique, siliques of Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrpb2-1* or *nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant hemizygotes contain only 25–27 mature seeds (Figure 2B).

Defects in seed set caused by the polymerase mutations were rescued by transforming Pol I, Pol II, or Pol III hemizygotes with full-length *NRPA2*, *NRPB2*, or *NRPC2* genomic clone transgenes expressed from their endogenous promoters (Figure 2B). Southern blot and segregation analyses showed that the transgenes in each case were integrated in multiple copies at a single locus (data not shown) such that the plants tested in Figure 2B were hemizygous for the polymerase mutant alleles as well as being hemizygous for the rescuing transgene loci. As a result, seed set is rescued by the transgenes to a level intermediate between the mutant and wild-type phenotypes. This is due to the independent segregation of the transgenes and polymerase alleles such that only half of the gametophytes bearing a mutant polymerase allele inherit a rescuing transgene. Collectively, our data indicate that functional RNA polymerases are essential for one or more critical aspects of female gametophyte development, fertilization, or seed development.

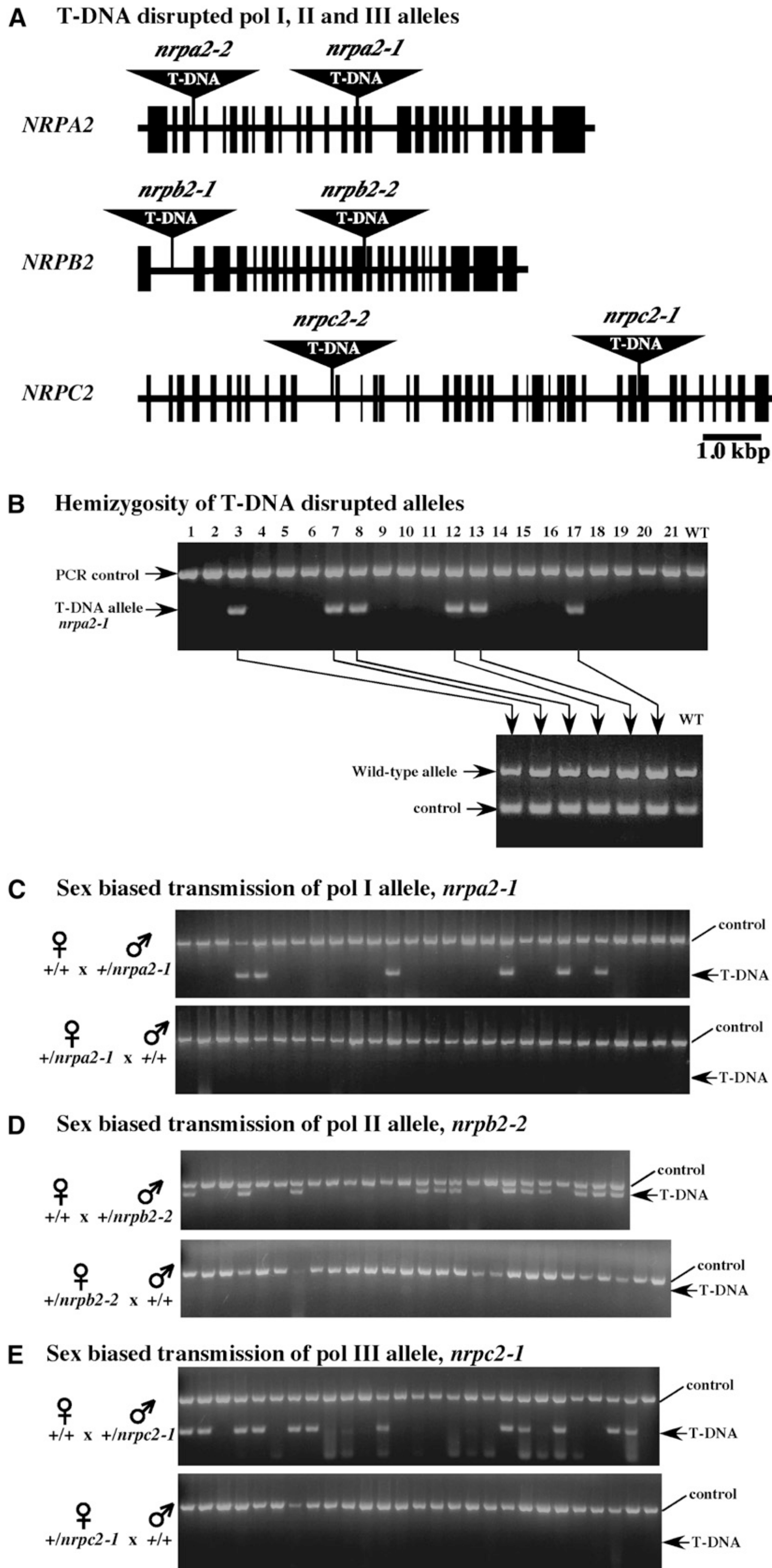


Figure 1.—Sex-biased transmission of disrupted alleles for second-largest subunits of RNA polymerases I, II, and III (*NRPA2*, *NRPB2*, and *NRPC2*, respectively). (A) Structures of the *NRPA2*, *NRPB2*, and *NRPC2* genes showing the positions of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* T-DNA insertions. Solid boxes represent exons. (B) PCR-based genotyping of progeny of a self-fertilized +/*nrpa2-1* hemizygote. Disrupted alleles were detected using a T-DNA-specific primer in conjunction with a gene-specific primer. Wild-type alleles were detected using primers that flank the T-DNA insertion site. (C–E) PCR-based detection of T-DNA disrupted alleles in progeny generated from reciprocal crosses between wild-type (+/+) and +/*nrpa2-1*, +/*nrpb2-2*, and +/*nrpc2-1* hemizygotes.

To further investigate the defects in ovule development and female transmission of mutant alleles (Figure 1, C–E; Table 2), ovaries of flowers at floral stage 13 (Bowman 1994), a stage just prior to flower opening, were examined by confocal laser scanning microscopy (CLSM). Female gametophytes develop relatively synchronously (Christensen *et al.* 1997) such that gametophytes that have undergone all three rounds of mitosis

TABLE 1

Genotypes of progeny of Pol I, II, and III hemizygotes

Parental genotype	% homozygous wt (+/+)	% hemizygous (+/−)	% homozygous mutant
+/*nrpa2-1*	76 (62/82)	24 (20/82)	0 (0/82)
+/*nrpa2-2*	63 (32/51)	37 (19/51)	0 (0/51)
+/*nrpb2-1*	86 (18/21)	14 (3/21)	0 (0/21)
qrt1-2, +/*nrpb2-2*	80 (67/84)	20 (17/84)	0 (0/84)
+/*nrpc2-1*	62 (39/63)	38 (24/63)	0 (0/63)
+/*nrpc2-2*	92 (45/49)	8 (4/49)	0 (0/49)

Mutant alleles *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* are underrepresented among the progeny of self-fertilized hemizygotes. Numbers in parentheses represent the number of individuals displaying a given genotype and the total number of individuals examined. wt, wild type.

(female gametophyte stages FG5–FG7; see Figure 3A) are observed at floral stage 13 in wild-type pistils (Figure 3B and Table 3). By contrast, in floral stage 13 pistils of hemizygous plants segregating mutant alleles for Pol I (+/*nrpa2-1* or +/*nrpa2-2*), Pol II (+/*nrpb2-1* or +/*nrpb2-2*, *qrt1-2*), or Pol III (+/*nrpc2-1* or +/*nrpc2-2*), ~50% of the female gametophytes arrest after only one or two rounds of mitosis (2–4 nuclei), at developmental stages FG2–FG4 (Table 3, Figure 3, C–G, and supplemental Figure S1). The other ~50% of the gametophytes in these ovaries display normal development, as in wild-type plants, consistent with the 1:1 segregation of wild-type and mutant alleles within the siliques of plants hemizygous for the mutations.

Detailed examination of ovules within +/*nrpa2-1* plants indicated that female gametophytes lacking functional Pol I arrest most frequently at the two-nucleus stage (FG2 and FG3; Figure 3, C and D and Table 3) and were not observed to progress beyond the four-nucleus stage. Similar results were observed for hemizygous plants bearing the *nrpa2-2* Pol I mutant allele (Table 3 and supplemental Figure S1, A and B).

As shown in Table 3, Figure 3, E–G, and supplemental Figure S1, most of the *nrpb2-1*, *nrpb2-2*, and *nrpc2-1* female gametophytes arrested after the second mitotic division (FG4), at the four-nucleus stage, whereas the majority of *nrpc2-2* female gametophytes displayed developmental arrest at the two-nucleus stage (FG2 and FG3). The difference in the severity of the *nrpc2-1* and *nrpc2-2* alleles is presumably due to the relative locations of the T-DNA insertions, with the T-DNA in the stronger *nrpc2-2* allele occurring in an earlier intron (see Figure 1).

Collectively, the microscopic analyses suggest that female gametophytes carrying defective alleles for RNA polymerases I, II, or III arrest early in development, at or prior to the four-nucleus stage, FG4.

Certation explains reduced male transmissibility of defective polymerase alleles: As shown in Table 2 and Figure 1, C–E, *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles are all transmitted via the male gametophyte. However, homozygous wild-type individuals outnumber hemizygous individuals among the progeny of self-fertilized hemizygotes or among the progeny of wild-type females outcrossed with a hemizygous male (Tables 1 and 2). These data indicate that male gametophytes bearing wild-type RNA polymerase alleles are either more viable or more successful at fertilization than are male gametophytes bearing mutant polymerase alleles.

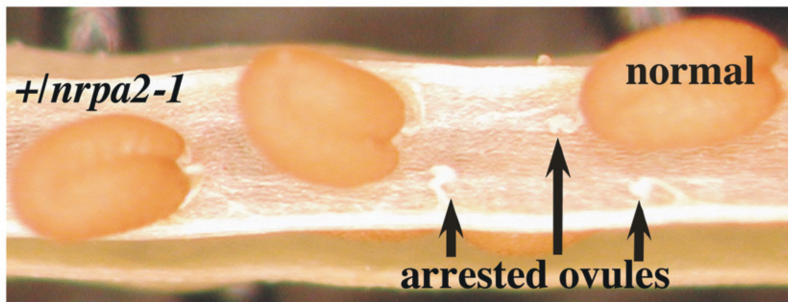
TABLE 2

Male-specific transmission of Pol I, II, and III mutant alleles

Parental genotype		Genotypes of progeny	
Female parent	Male parent	% homozygous wt (+/+)	% hemizygous (+/−)
+/*nrpa2-1*	+/+	100 (55/55)	0 (0/55)
+/*nrpa2-2*	+/+	100 (46/46)	0 (0/46)
+/*nrpb2-1*	+/+	100 (52/52)	0 (0/52)
qrt1-2, +/*nrpb2-2*	+/+	100 (42/42)	0 (0/42)
+/*nrpc2-1*	+/+	100 (56/56)	0 (0/56)
+/*nrpc2-2*	+/+	100 (47/47)	0 (0/47)
+/+	+/*nrpa2-1*	75 (42/56)	25 (14/56)
+/+	+/*nrpa2-2*	62 (24/39)	38 (15/39)
+/+	+/*nrpb2-1*	79 (38/48)	21 (10/48)
+/+	*qrt1-2*, +/*nrpb2-2*	70 (19/27)	30 (8/27)
+/+	+/*nrpc2-1*	67 (36/54)	33 (18/54)
+/+	+/*nrpc2-2*	87 (45/52)	13 (7/52)

Paternally biased transmission of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles. Wild-type (+/+) plants were reciprocally crossed with +/*nrpa2-1*, +/*nrpa2-2*, +/*nrpb2-1* (in *qrt1-2* mutant background); +/*nrpb2-2*, +/*nrpc2-1*, and +/*nrpc2-2* and resulting progeny were genotyped. Numbers in parentheses are the number of progeny displaying the specified genotype out of the total number of progeny examined.

A



B

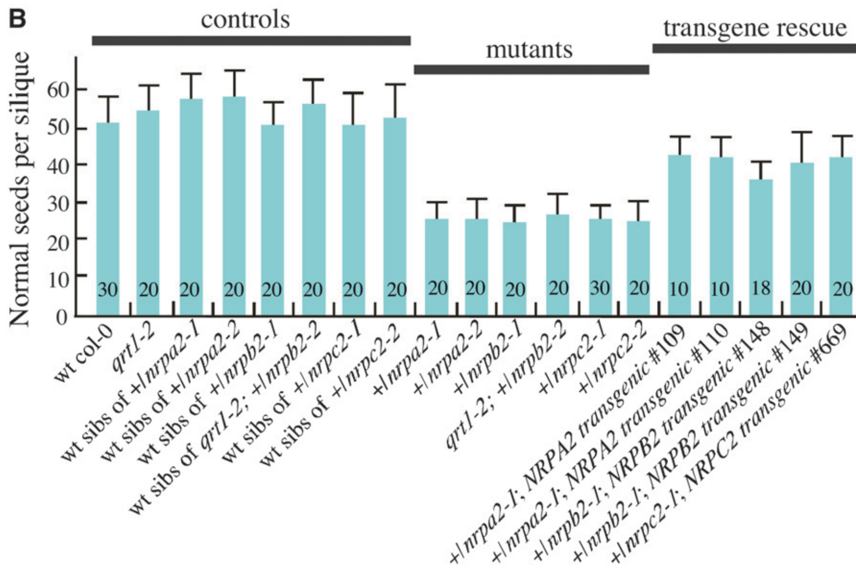


Figure 2.—Failed seed development in siliques of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* hemizygotes. (A) A silique of a hemizygous +/*nrpa2-1* plant. Normal seeds and undeveloped (arrested) ovules occur in a silique of a hemizygous plant. (B) Average amounts of normal seeds per silique from wild-type and hemizygous plants. Numbers of siliques examined are indicated.

To investigate the influence of defective RNA polymerase alleles on pollen development and viability using tetrad analysis, we generated lines that carry a Pol I (*nrpa2-1*), Pol II, (*nrpb2-1*), or Pol III (*nrpc2-1*) mutant allele in the *quartet* (*qrt*) mutant background. The quartet mutation causes the four pollen that develop from the four meiotic products (microspores) to remain associated with one another, rather than dissociating into individual pollen grains. Thus, pollen tetrads of plants hemizygous for the polymerase mutants include two pollen-bearing mutant polymerase alleles and two bearing wild-type polymerase alleles.

Pollen tetrads were examined by DAPI (4′,6-diamidino-2-phenylindole), FDA (fluorescein diacetate), or Alexander staining (Figure 4, A–H). DAPI staining of chromatin in pollen of *quartet* (*qrt1-2*) mutant plants; Pol I hemizygote *quartet* (+/*nrpa2-1; qrt1-2*), Pol II hemizygote *quartet* (+/*nrpb2-1; qrt1-2* as well as +/*nrpb2-2; qrt1-2*), or Pol III hemizygote *quartet* (+/*nrpc2-1; qrt1-2*) plants revealed the normal pattern of one diffuse vegetative cell nucleus and two compact sperm cell nuclei in each of the four attached pollen (Figure 4, B and F, and data not shown). FDA and Alexander staining detected no differences in viability among the individual pollen in tetrads of wild-type or mutant plants (Figure 4, C, D, G, and H, and data not shown).

Two of the pollen in each tetrad of a polymerase mutant hemizygote carry defective RNA polymerase alleles and lack wild-type alleles. In the case of the *nrpb2-2* hemizygotes, the mutant alleles are tagged by a *LAT52*::*GUS* reporter gene that is present within the T-DNA inserted into the Pol II *NRPB2* gene (Figure 4, I and J). The *LAT52* promoter is specifically expressed in mature pollen and pollen tubes, thereby allowing the pollen bearing the mutant *nrpb2-2* alleles to be visualized by GUS staining. Equal numbers of GUS-positive (blue) and GUS-negative pollen are present in *nrpb2-2*/+ pollen quartets, indicating that wild-type and mutant pollen develop in equal abundance and that the *nrpb2-2* mutant allele segregates normally (Figure 4, I and J).

It is noteworthy that mRNA-encoded proteins, such as the GUS enzyme, are synthesized by RNA polymerase II and require the distinctive 5′7-methylguanosine caps and poly A tails of Pol II transcripts to be translated. Pol I and Pol III transcripts lack these features and are not translated. Despite the disruption of the gene encoding the essential Pol II second-largest subunit (NRPB2), the GUS enzyme is clearly expressed from the *LAT52* promoter in *nrpb*2 mutant pollen (Figure 4, I and J). Expression of the GUS gene cannot be attributed to stored GUS mRNA transcribed premeiotically; if so, it would be

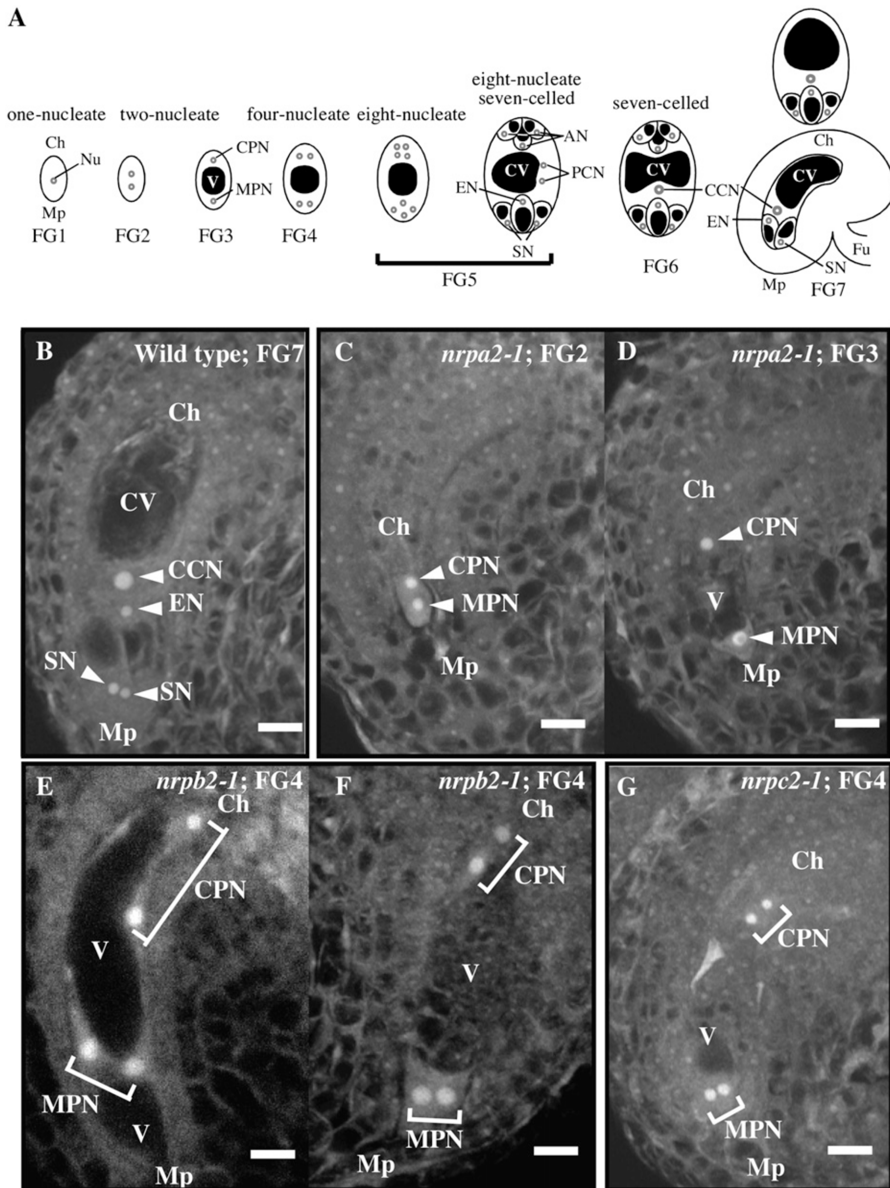


FIGURE 3.—Developmental arrest of mutant female gametophytes in flowers just prior to anthesis was visualized by confocal fluorescence microscopy. (A) Stages of female gametophyte development (FG1–FG7), according to CHRISTENSEN *et al.* (1997). Mp, micropylar pole; Ch, chalazal pole; Nu, nucleus; V, vacuole; CPN, chalazal pole nucleus; MPN, micropylar nucleus; AN, antipodal cell nucleus; CV, central cell vacuole; EN, egg cell nucleus; PCN, polar cell nucleus; CCN, central cell nucleus; SN, synergid cell nucleus; Fu, funiculus. (B) A wild-type female gametophyte, at floral stage 13, that is fully developed (FG7). The nuclei and vacuoles for the 2N central cell, the egg cell, and two synergid cells are apparent. (C and D) *nrpa2-1* female gametophytes arrested at the two-nucleate stage (FG2 and FG3). (E and F) *nrpb2-1* female gametophytes arrested at the four-nucleate stage. (G) A *nrpc2-1* female gametophyte arrested at the four-nucleate stage. Scale bars, 10 μm.

present in all four pollen of the tetrad. Moreover, the *LAT52* promoter has previously been shown to be expressed only postmeiotically, making it a useful male-gametophyte-specific marker (EADY *et al.* 1994; TWELL *et al.* 1990). We conclude that Pol II transcription takes place in *nrpb2-2* mutant pollen despite the lack of a functional *NRPB2* allele.

Examination of pollen germination and pollen tube growth *in vitro* revealed no differences among pollen tubes that grew from pollen quartets consisting of two pollen-bearing defective RNA polymerase alleles and two pollen-bearing wild-type alleles, at least up to a pollen tube length of 100–150 μm (Figure 4, K and L, and data not shown). Self-pollinated pistils of *qrt1-2; +/nrpb2-2* plants stained for GUS also reveal pollen tube growth from pollen bearing the disrupted allele *in vivo* (Figure 4, M and N). Most of the GUS-stained tubes from *nrpb2-2* pollen are observed at the stigma and upper portions of the ovary (Figure 4, M and N; Figure 5C). However, in rare cases, tubes from *nrpb2-2* pollen are observed in the distal portion of the ovary (Figure 4N, images at top right and bottom). Collectively, these observations suggest that in pollen that do not encode endogenous functional RNA polymerase II, Pol II-dependent GUS activity is sustained during pollen development and early pollen tube growth.

To test the hypothesis that pollen bearing Pol I (*nrpa2-1* or *nrpa2-2*), Pol II (*nrpb2-1* or *nrpb2-2*), or Pol III (*nrpc2-1* or *nrpc2-2*) mutant alleles are at a competitive disadvantage compared to wild-type pollen, we determined the distribution of seeds bearing mutant alleles within the siliques of self-pollinated hemizygous plants. Due to the previously demonstrated lethality of the 50% of female gametophytes that inherit a

TABLE 3

Female gametophyte development in polymerase mutants

Plant genotype	Pistil identification no.	No. of female gametophytes at specified developmental stages							Total
		FG1	FG2	FG3	FG4	FG5	FG6	FG7	
wt col-0	1					7	2	8	17
	2					1	2	11	14
	3					3	1	12	16
	4						1	14	15
qrt1-2	1						1	12	13
	2					5	3	9	17
	3							13	13
+/*nrpa2-1*	1		4	4	1		2	4	15
	2		2	4	2		1	3	12
	3		4	4	2	7	2	4	23
	4		1	6			1	9	17
+/*nrpa2-2*	1		3	6	1	2	1	7	20
	2		2	7	1	4	1	1	16
	3		1	4	1	7	3	1	17
	4		1	6	4	2	1	7	21
+/*nrpb2-1*	1			2	8		1	4	15
	2			1	4		2	2	9
	3			3	3	1	1	7	15
	4				9		2	8	19
	5			1	10		2	9	22
	6			4	10			8	22
qrt1-2, +/*nrpb2-2*	1			2	6	2	1	4	15
	2			5	8	3	2	3	21
	3			1	3			8	12
+/*nrpc2-1*	1			4	12	2	1	7	26
	2			2	7	2	1	5	17
	3			2	3	3	1	7	16
	4			4	4	1	2	6	17
+/*nrpc2-2*	1		3	10	1	3	1	6	24
	2		2	10			1	8	21
	3		3	6	1		3	6	19
	4		1	8			1	9	19
	5		2	7	1		1	8	19

Developmentally arrested *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* female gametophytes. Pistils from flowers just prior to anthesis (flower opening) were fixed, and female gametophytes within these pistils were classified according to their developmental stage (FG1–FG7). wt, wild type.

mutant polymerase allele (depicted as ovules with an "X" through them in Figure 5A), only the 50% of female gametophytes that bear wild-type alleles are available to be fertilized. Therefore, any mutant alleles detected in the seeds are inherited via the male gametophytes (refer to Figure 1, C–E, and Table 2). Seeds were collected from the top one-third of the silique, which is nearest to the stigma where the pollen germinates to initiate the growth of pollen tubes, or from the middle or bottom one-third of the silique. Following germination of the seeds, resulting plants were genotyped (Figure 5B). This test revealed that mutant alleles were found most frequently among seeds that developed within the top one-third of the siliques; 35–50% of these seeds develop as hemizygotes (note that a frequency of 50% is expected if there is no difference in the fitness of wild-type and mutant pollen). The frequency of hemizygous seeds within the middle portions of the siliques were significantly reduced (11–21%) in comparison with the top one-third, except for the *nrpa2-2* allele that was detected in 16 of the 23 sibs examined. In the bottom one-third of the siliques, where fertilization of the ovules would require the growth of the longest pollen tubes, hemizygotes represented only a small proportion of the seeds (0–11%).

The extent of mutant pollen tube growth fits with the distribution of hemizygous seeds following fertilization. A nonmutant transgenic line in which a T-DNA bearing the *LAT52*::GUS reporter gene inserted into an intergenic region was used as a control for comparison to

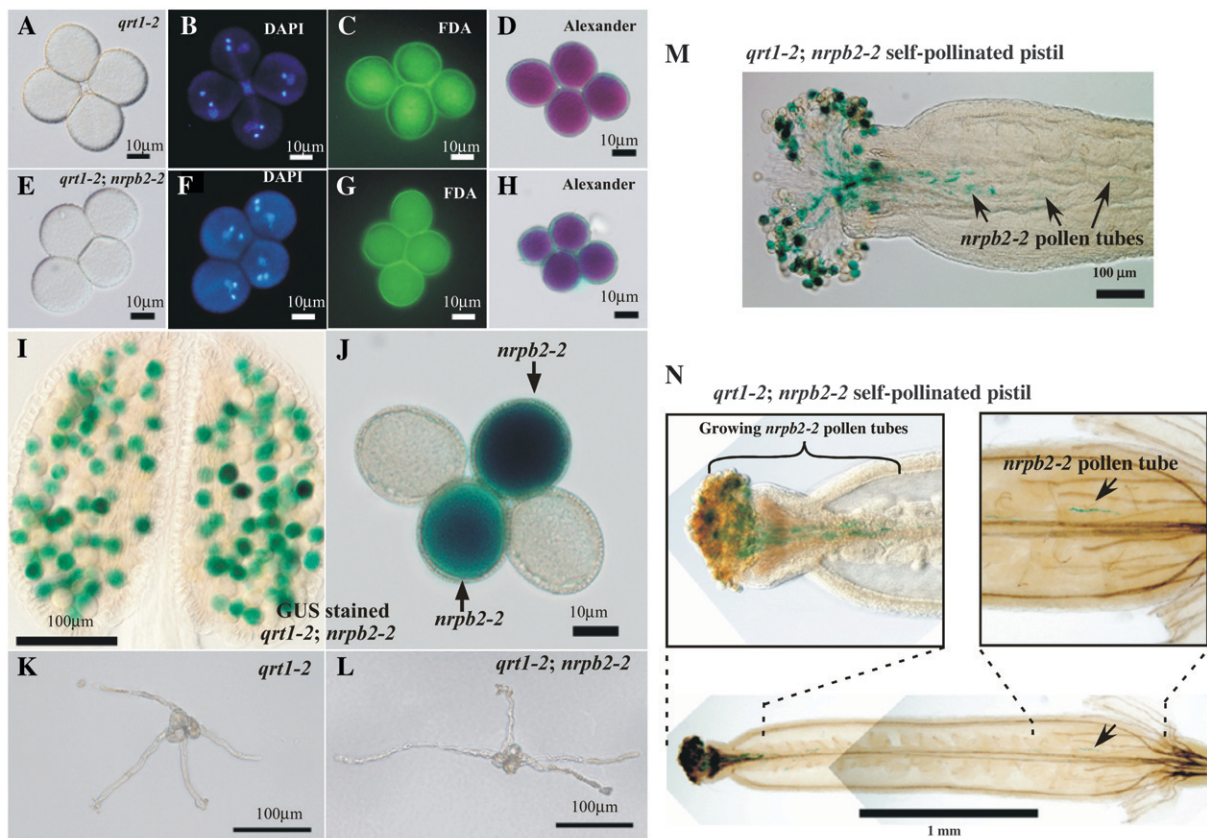


Figure 4.—Development and early tube elongation of pollen are unaffected by defects in RNA polymerases. (A–H) Cytological examination of mature pollen from *qrt1-2* (a–d) and *qrt1-2; +/nrpb2-2* (e–h). (a and e) Bright-field microscopy; (b and f) DAPI staining test; (c and g) FDA staining test; (d and h) Alexander staining test. (I and J) *LAT52*::*GUS* expression in pollen defective for the Pol II subunit (*nrpb2-2* pollen). (K and L) Germinating *qrt1-2* (k) and *qrt1-2; nrpb2-2* (l) pollen. Pollen was incubated for 18 hr at 22 ° in a germination medium and its images were captured. Note that four tubes of quartet pollen from wild-type (k, *qrt1-2*) and mutant (l, *qrt1-2; +/ nrpb2-2*) plants grew equally in this assay, to a length of ∼100–150 mm. (M and N) Self-pollinated pistils from *qrt1-2; +/ nrpb2-2* plants. *LAT52*::*GUS* was expressed during pollen tube growth in the absence of the functional allele of a catalytic subunit of Pol II. A considerable number of *nrpb2-2* pollen tubes (blue stained) was present in the top portions of the pistils. Note that a tube from *nrpb2-2* pollen grew into ∼2.0 mm in length (N).

nrpb2-2. Whereas GUS-stained *nrpb2-2* pollen tubes are rarely observed deeper than the top one-third of the pistil, GUS-stained control pollen tubes are easily detected throughout the top and middle one-thirds of the pistils and can be observed all the way to the base of the pistil (Figure 5C). Taken together, our results suggest that pollen germination, early pollen tube elongation, and fertilization are not severely affected by the lack of functional alleles for the RNA polymerase I, II, or III subunits. However, sustained pollen tube growth presumably requires *de novo* synthesis of essential RNA polymerase genes such that mutant pollen are at a competitive disadvantage compared to wild-type pollen, the phenomenon known as certation (Heribert-Nilsson 1920).

DISCUSSION

Genetic analyses have identified a large number of female gametophytic mutants in Arabidopsis, a significant fraction of which correspond to mutant alleles of transcription factors (Pagnussat *et al.* 2005). Our demonstration that mutations in RNA polymerases I, II, and III cause female gametophyte lethality are generally consistent with these findings and indicate that the female gametophyte is dependent on endogenous transcription machinery synthesized *de novo* during gametophyte development. In the absence of functional RNA polymerase subunits, female gametophytes can often progress to the two-nucleate stage, but typically arrest before, or shortly after, the second of the three mitotic divisions required for development of mature gametophytes. It is noteworthy that the SeedGenes Project database (http://www.seedgenes.org/index.html) (Tzafrir *et al.* 2003, 2004) includes information for two T-DNA insertion alleles of *nrpb2*, named *emb 1989-1* and *emb 1989-2*. Embryos fail to develop in 90–94% of ovules bearing these mutant alleles, consistent with the female gametophytic lethal phenotype we describe in this article. However, 6–10% of *emb 1989-1* and *emb 1989-2* ovules are reported to arrest as preglobular embryos, indicating that

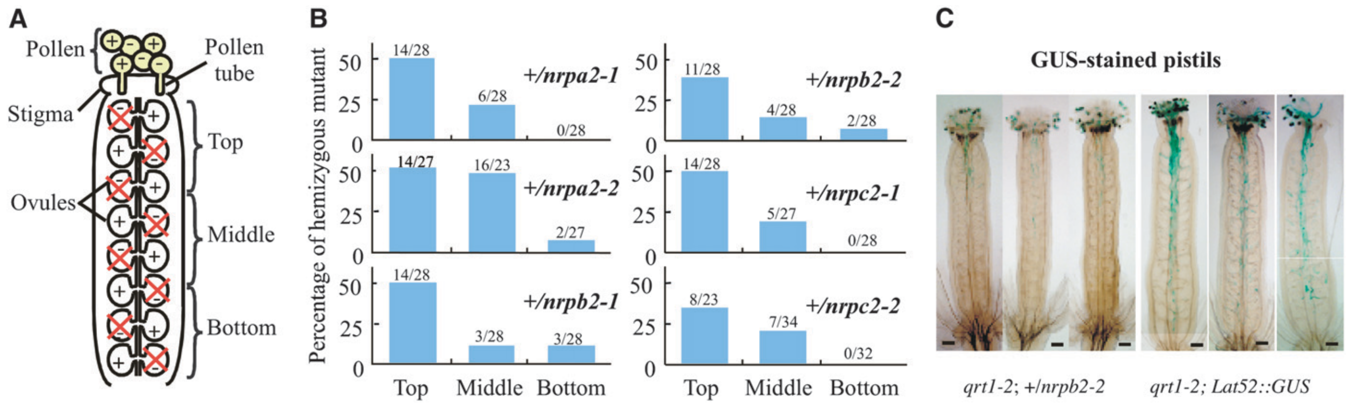


Figure 5.—Reduced paternal transmission of *nrpa2-1*, *nrpa2-2*, *nrpb2-1*, *nrpb2-2*, *nrpc2-1*, and *nrpc2-2* alleles relative to wild-type alleles in self-fertilized hemizygotes is due to decreased, competitive fertilization of ovules farthest from the stigma. (A) A diagram of the female floral organ (the pistil), whose surface (the stigma) is the site where a pollen grain germinates and initiates formation of a pollen tube. Half of the pollen of a hemizygote has wild-type (+) RNA polymerase alleles and half are mutant (−), but all develop and mature. Likewise, within the ovary of a hemizygote, half of the ovules are wild type and half are mutant with respect to the RNA polymerase alleles. However, the latter fail to develop (denoted with an "X") such that mutant alleles in fertilized ovules and seeds are derived from the male gametophyte. (B) Seeds collected from the top, middle, and bottom portions of siliques of the hemizygotes were germinated and resultant plants were genotyped. The numbers of plants of each genotype are indicated. Note that mutant alleles are more abundant in seeds developing nearest the stigma, at the top of the siliques, where the shortest pollen tubes would be needed to reach the ovules. (C) Self-pollinated pistils from *qrt1-2; +/ nrpb2-2* plants and transformants hemizygously carrying a *LAT52*::*GUS* reporter gene(s) inserted in an intergenic region (*qrt1-2; LAT52*::*GUS*). Pollen tubes from *qrt1-2; nrpb2-2* pollen (blue stained) were present in top portions of the pistils, while control pollen tubes (*qrt1-2; LAT52*::*GUS*) were observed all the way from the tops to the bottoms of the pistils.

the female gametophytes in these cases had completed development and had been fertilized, but produced embryos that were then unable to complete development. Cloning and sequencing of the region that defines the junction between the *NRPB2* gene and the T-DNA revealed that the T-DNA in *emb 1989-1* inserted 34 nucleotides upstream from the translation start site (Y. Onodera, data not shown). Because the protein coding region is not disrupted, it is possible that the *emb 1989-1* allele is partially functional, which may explain how development can sometimes proceed to stages beyond what we have observed for the *nrpb2-1* and *nrpb2-2* alleles. We currently lack analogous data concerning the precise location of the T-DNA in the *emb 1989-2* allele.

A recent study of developing and mature pollen showed that 61.9% of all Arabidopsis genes are expressed during at least one stage of male gametophyte development, with 9.7% of the transcripts being pollen specific (Honys and Twell 2004). A large number of transcription factors are expressed during pollen development, suggesting that orchestrated waves of transcription are essential for pollen maturation. Mature pollen is also known to contain proteins, ribosomes, mRNAs, rRNAs, and tRNAs that are synthesized postmeiotically during pollen maturation or pollen tube growth (Mascarenhas 1975, 1989). Therefore, we were surprised to find that functional alleles of RNA polymerases I, II, and III are not absolutely required in the haploid pollen genome to complete pollen development, germination, pollen tube growth, or fertilization. The simplest explanation is that transcription in pollen-bearing defective polymerase alleles is conducted using RNA polymerases, or stored mRNAs encoding RNA polymerase subunits, that are synthesized premeiotically in the hemizygous microspore mother cell and are then partitioned into the microspores following meiosis. The one functional allele is apparently sufficient for microspore mother cells to load microspores with enough polymerase to support subsequent pollen development and postgermination pollen functions, including pollen tube growth and fertilization.

Transcript profiling using DNA microarray technology has shown that mRNAs encoding the core subunits for nuclear RNA polymerases are present within unicellular microspores at similar or greater abundance than in sporophytic tissues (Honys and Twell 2004). However, in mature pollen, mRNAs encoding transcription factors, RNA processing proteins, and translation machineries are less abundant than in vegetative tissues of the plant (Honys and Twell 2003; Pina *et al.* 2005; Grennan 2007). This holds true for transcripts encoding the core subunits for nuclear RNA polymerases I, II, and III, which either are not detected in mature pollen or are present at very low levels (Honys and Twell 2003; Pina *et al.* 2005). The idea that maternally derived polymerase subunit mRNAs are stored for translation late in pollen development is not readily supported by these observations, but the possibility cannot be ruled out. An alternative hypothesis is that polymerase proteins derived from the microspore mother cell, or translated from mRNAs partitioned into the unicellular microspores, persist in mature pollen. Plants hemizygous for a

single-copy transgene expressing a polymerase subunit-GFP fusion protein would be useful for testing this hypothesis. If the transgene were capable of rescuing plants that were homozygous for null alleles of the corresponding endogenous genes, one would expect the GFP marker to segregate 2:2 among the pollen. If GFP were observed in all pollen, this would indicate maternal loading of the polymerase subunit. Regardless of whether stored mRNA or stored protein is responsible for allowing the transmission of mutant polymerase alleles through the pollen, there are enough of the stored molecules to complete pollen development, germination, and fertilization. These developmental events are thought to span a period of at least 90 hr (Bowman 1994). However, additional *de novo* synthesis of Pol I, II, and III is apparently needed for full pollen vigor and for growth of pollen tubes long enough to reach the ovules farthest from the stigma.

Given the reduced fitness of mutant pollen relative to wild-type pollen, deleterious mutant polymerase alleles are unlikely to become widespread among a population. However, some gene evolution phenomena would seem to be favored by allowing mutant alleles to persist in the population for some period of time. For instance, a characteristic of the RNA polymerase I transcription system is that it evolves rapidly, such that the transcription machinery of one species cannot transcribe the rRNA genes of an unrelated species (Grummt *et al.* 1982; Miesfeld and Arnheim 1984; Doelling and Pikaard 1996). Species specificity appears to be explained by the rapid evolution of rRNA gene sequences and the corresponding coevolution of the transcription machinery, such that changes in gene sequences can be tolerated as a result of compensatory changes in the proteins that bind these sequences (or vice versa). Because haploid selection against defective alleles is less stringent in the male gametophyte than in the female gametophyte, at least for subunits of RNA polymerases I, II, and III, it is tempting to speculate that the male lineage could be the conduit for transmitting mutations that might initially be deleterious but could be tolerated if a compensatory mutation in an interacting protein or DNA sequence were to occur. Transmitting mutations at moderate frequency via the pollen would presumably buy time for such compensatory mutations to occur. However, the null hypothesis is that the capacity to transmit mutations in essential housekeeping genes such as RNA polymerases via pollen has no evolutionary advantage and is merely an unintended consequence of pollen development.

Research at Hokkaido University was supported in part by a grant in aid for scientific research from the Ministry of Education, Science and Culture and grants from the Akiyama Foundation and Grant Program of The Sumitomo Foundation. Pikaard lab research was supported by National Institutes of Health grants GM-060380 and GM-077590. The content of this article is solely the responsibility of the authors and does not necessarily reflect the views of the funding agencies that supported the work.

LITERATURE CITED

Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen *et al.*, 2003 Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science **301:** 653–657.

Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408:** 796–815.

Bowman, J. (Editor), 1994 *Arabidopsis: An Atlas of Morphology and Development.* Springer-Verlag, New York.

Christensen, C., E. King, J. R. Jordan and G. N. Drews, 1997 Megagametogenesis in Arabdopsis wild type and the Gf mutant. Sex. Plant Reprod. **10:** 49–64.

Doelling, J. H., and C. S. Pikaard, 1996 Species-specificity of rRNA gene transcription in plants manifested as a switch in polymerase-specificity. Nucleic Acids Res. **24:** 4725–4732.

Drews, G., and R. Yadegari, 2002 Development and function of the angiosperm female gametophyte. Annu. Rev. Genet. **36:** 99–124.

Eady, C., K. Lindsey and D. Twell, 1994 Differential activation and conserved vegetative cell-specific activity of a late pollen promoter in species with bicellular and tricellular pollen. Plant J. **5:** 543–550.

Earley, K., J. Haag, O. Pontes, K. Opper, T. Juehne *et al.*, 2006 Gateway-compatible vectors for plant functional genomics and proteomics. Plant J. **45:** 616–629.

Grennan, A., 2007 An analysis of the Arabidopsis pollen transcriptome. Plant Physiol. **145:** 3–4.

Grossniklaus, U., and K. Schneitz, 1998 The molecular and genetic basis of ovule and megagametophyte development. Semin. Cell Dev. Biol. **9:** 227–238.

Grummt, I., E. Roth and M. R. Paule, 1982 rRNA transcription *in vitro* is species-specific. Nature **296:** 173–174.

Hashida, S., H. Takahashi, M. Kawai-Yamada and H. Uchimiya, 2007 Arabidopsis thaliana nicotinate/nicotinamide mononucleotide adenyltransferase (AtNMNAT) is required for pollen tube growth. Plant J. **49:** 694–703.

Heribert-Nilsson, N., 1920 Zuwachsgeschwindigkeit der Pollenschläuche und gestörte Mendelzahlen bei Oenothera Lamarckiana. Hereditas **1:** 41–67.

Higashiyama, T., 2002 The synergid cell: attractor and acceptor of the pollen tube for double fertilization. J. Plant Res. **115:** 149–160.

Higashiyama, T., S. Yabe, N. Sasaki, Y. Nishimura, S. Miyagishima *et al.*, 2001 Pollen tube attraction by the synergid cell. Science **293:** 1480–1483.

Higashiyama, T., H. Kuroiwa and T. Kuroiwa, 2003 Pollen-tube guidance: beacons from the female gametophyte. Curr. Opin. Plant Biol. **6:** 36–41.

Honys, D., and D. Twell, 2003 Comparative analysis of the Arabidopsis pollen transcriptome. Plant Physiol. **132:** 640–652.

Honys, D., and D. Twell, 2004 Transcriptome analysis of haploid male gametophyte development in Arabidopsis. Genome Biol. **5:** R85.

Johnson, M., and D. Preuss, 2002 Plotting a course: multiple signals guide pollen tubes to their targets. Dev. Cell **2:** 273–281.

Johnston, A., P. Meier, J. Gheyselinck, S. Wuest, M. Federer *et al.*, 2007 Genetic subtraction profiling identifies genes essential for Arabidopsis reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. Genome Biol. **8:** R204.

Jones-Rhoades, M., J. Borevitz and D. Preuss, 2007 Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. PLoS Genet. **3:** 1848–1861.

Kassavetis, G. A., C. Bardeleben, B. Bartholomew, B. R. Braun, C. A. P. Joazeiro *et al.*, 1994 Transcription by RNA polymerase III, pp. 107–126 in *Transcription Mechanisms and Regulation*, edited by R. C. Conaway and J. W. Conaway. Raven Press, New York.

Larkin, R., and T. Guilfoyle, 1993 The second largest subunit of RNA polymerase II from Arabidopsis thaliana. Nucleic Acids Res. **21:** 1038.

Mascarenhas, J. P., 1975 The Biochemistry of angiosperm pollen development. Bot. Rev. **41:** 259–314.

Mascarenhas, J. P., 1989 The Male gametophyte of flowering plants. Plant Cell **1:** 657–664.

Mascarenhas, J. P., 1993 Molecular mechanisms of pollen tube growth and differentiation. Plant Cell **5:** 1303–1314.

McCormick, S., 1993 Male gametophyte development. Plant Cell **5:** 1265–1275.

McCormick, S., 2004 Control of male gametophyte development. Plant Cell **16:** S142–S153.

Miesfeld, R., and N. Arnheim, 1984 Species-specific rDNA transcription is due to promoter-specific binding factors. Mol. Cell. Biol. **4:** 221–227.

Onodera, Y., J. R. Haag, T. Ream, P. C. Nunes, O. Pontes *et al.*, 2005 Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120:** 613–622.

Pagnussat, G., H. Yu, Q. Ngo, S. Rajani, S. Mayalagu *et al.*, 2005 Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development **132:** 603–614.

Paule, M. R., and R. J. White, 2000 Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res. **28:** 1283–1298.

Pina, C., F. Pinto, J. Feijó and J. Becker, 2005 Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol. **138:** 744–756.

Preuss, D., S. Rhee and R. Davis, 1994 Tetrad analysis possible in Arabidopsis with mutation of the QUARTET (QRT) genes. Science **264:** 1458–1460.

Rosso, M., Y. Li, N. Strizhov, B. Reiss, K. Dekker *et al.*, 2003 An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53:** 247–259.

Russell, S., 1993 The egg cell: development and role in fertilization and early embryogenesis. Plant Cell **5:** 1349–1359.

Schneitz, K., M. Hülskamp and R. E. Pruitt, 1995 Wild-type ovule development in Arabidopsis thaliana: a light microscope study of cleared whole-mount tissue. Plant J. **7:** 731–749.

Sprunck, S., U. Baumann, K. Edwards, P. Langridge and T. Dresselhaus, 2005 The transcript composition of egg cells changes significantly following fertilization in wheat (Triticum aestivum L.). Plant J. **41:** 660–672.

Steffen, J., I. Kang, J. MacFarlane and G. Drews, 2007 Identification of genes expressed in the Arabidopsis female gametophyte. Plant J. **51:** 281–292.

Twell, D., J. Yamaguchi and S. McCormick, 1990 Pollen-specific gene-expression in transgenic plants: coordinate regulation of 2 different tomato gene promoters during microsporogenesis. Development **109:** 705.

Tzafrir, I., A. Dickerman, O. Brazhnik, Q. Nguyen, J. McElver *et al.*, 2003 The Arabidopsis SeedGenes Project. Nucleic Acids Res. **31:** 90–93.

Tzafrir, I., R. Pena-Muralla, A. Dickerman, M. Berg, R. Rogers *et al.*, 2004 Identification of genes required for embryo development in Arabidopsis. Plant Physiol. **135:** 1206–1220.

Yadegari, R., T. Kinoshita, O. Lotan, G. Cohen, A. Katz *et al.*, 2000 Mutations in the FIE and MEA genes that encode interacting polycomb proteins cause parent-of-origin effects on seed development by distinct mechanisms. Plant Cell **12:** 2367–2382.

Yang, H., N. Kaur, S. Kiriakopolos and S. McCormick, 2006 EST generation and analyses towards identifying female gametophyte-specific genes in Zea mays L. Planta **224:** 1004–1014.

Yu, H., P. Hogan and V. Sundaresan, 2005 Analysis of the female gametophyte transcriptome of Arabidopsis by comparative expression profiling. Plant Physiol. **139:** 1853–1869.

Communicating editor: J. A. Birchler

Supplemental data

Table S1 methods:

NRPB12a (At5g41010) T-DNA insertion line SALK_049327 was obtained from ABRC. DNA was extracted from 1-3 leaves in microcentrifuge tubes using a modified version of a previously published protocol (2). Briefly, leaves were incubated 10 min in 300 ul of extraction buffer (200 mM Tris, pH 7.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) at 99˚C. Cell debris was cleared by centrifugation at 14,000 x g, 8 min. The supernatant was transferred to a new tube containing an equal volume of isopropanol, mixed and incubated at room temperature for fifteen minutes. DNA was pelleted by centrifugation at 14,000 x g, 15 min. Pellets were washed once in 70% ethanol before resuspending in 100 ul of 1x TE buffer, pH 8.0. Debris was pelleted by centrifuging one minute at top speed in a microcentrifuge. 2 ul of DNA was used in a 20 ul PCR reaction with GoTaqGreen (Promega) and appropriate primers. The wild-type *NRBP12a* gene was amplified using forward primer 5'-TTATAGCCAATCAAGGATTATAGCAATGTGAAC-3' and reverse primer 5'-GAAATCAAAGTTTTGTTAGTATCTGTAAAAGATTG-3'. The T-DNA inserted allele was detected using the reverse primer above in combination with the SALK line T-DNA Left border primer, LBa1: 5'-TGGTTCACGTAGTGGGCCATCG-3'.

Figure S1. Developmentally arrested mutant female gametophytes within pistils just prior to anthesis, visualized by confocal fluorescence microscopy. **(A and B)** *nrpa2-2* female gametophytes arrested at the two-nucleate stage **(C)** A *nrpb2-2* female gametophyte arrested at the two-nucleate stage **(D)** A *nrpb2-2* female gametophyte

arrested at the four-nucleate stage. **(E)** A *nrpc2-2* female gametophyte arrested at the two-nucleate stage and displaying a prominent vacuole. Abbreviations: Mp, micropylar pole; Ch, chalazal pole; CPN, chalazal pole nucleus; MPN, micropylar nucleus; V, vacuole; Nu, nucleus. Scale bars = 10 mm.

References for Supplemental Data

1. Alonso, J., Stepanova, A., Leisse, T., Kim, C., Chen, H., Shinn, P., Stevenson, D., Zimmerman, J., Barajas, P., Cheuk, R. et al. (2003). Genome-wide Insertional mutagenesis of Arabidopsis thaliana. *Science* 301, 653-657.

2. Herr, A., Molnar, A., Jones, A. and Baulcombe, D. (2006). Defective RNA processing enhances RNA silencing and influences flowering of Arabidopsis. *Pro. The Natl. Acad. Sci. USA* 103, 14994-15001.

Table S1. Male-specific transmission of *RPB12a* mutant alleles.

Parental genotype		Genotypes of progeny	
Female parent	Male parent	homozygous wt (+/+)	hemizygous (+/-)
+/nrpb12a	+/+	100% (20/20)	0% (0/20)
+/+	+/nrpb12a	40% (24/60)	60% (36/60)

Table S2. Transgene rescue allows maternal transmission of mutant alleles

Parental genotype		Genotypes of progeny	
Female parent	Male parent	Homozygous wt (+/+)	Hemizygous (+/-)
+/*nrpa2-1*, *NRPA2* transgenic #109	+/+	67% (12/18)	33% (6/18)
+/*nrpa2-1*, *NRPA2* transgenic #110	+/+	62% (16/26)	38% (10/26)
+/*nrpb2-1*, *NRPB2* transgenic #148	+/+	50% (9/18)	50% (9/18)
+/*nrpb2-1*, *NRPB2* transgenic #149	+/+	55% (11/20)	45% (9/20)
+/*nrpc2-1*, *NRPC2* transgenic #669	+/+	74% (31/42)	26% (11/42)

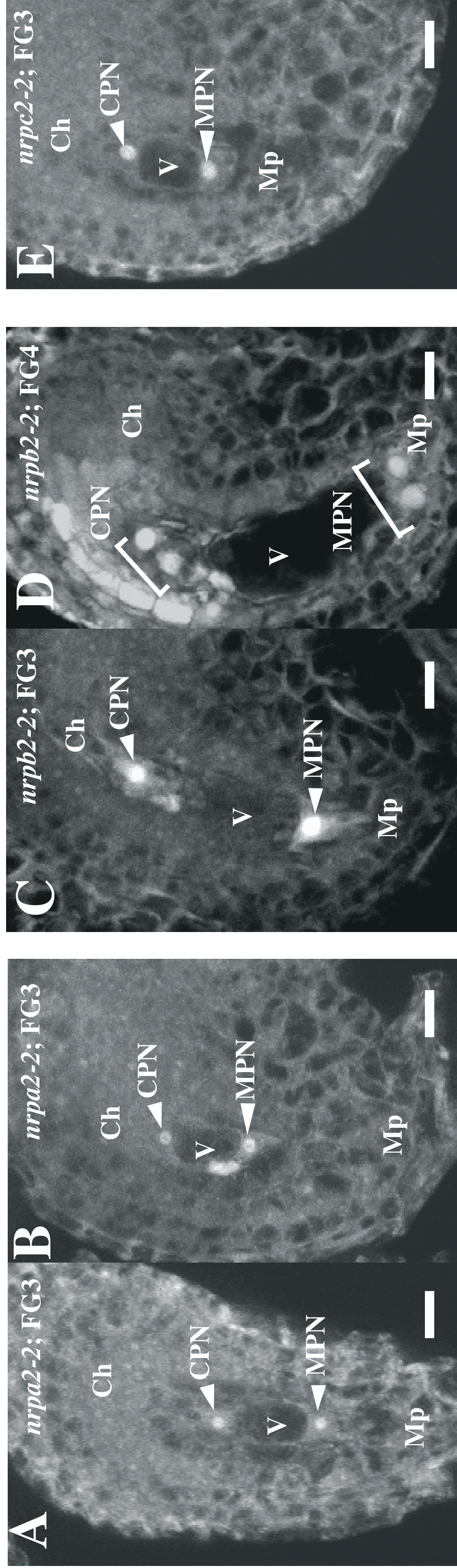


Figure S1

Copyright © 2009 by the Genetics Society of America

CORRIGENDUM

In the article by Y. Onodera, K. Nakagawa, J. R. Haag, D. Pikaard, T. Mikami, T. Ream, Y. Ito and C. S. Pikaard (Genetics **180:** 207–218) entitled "Sex-Biased Lethality or Transmission of Defective Transcription Machinery in Arabidopsis," on page 215, right column, the text beginning with "It is noteworthy that the SeedGenes" should be replaced with:

> It is noteworthy that the SeedGenes Project database (http://www.seedgenes.org/index.html) (Tzafrir *et al.* 2003, 2004) includes information for two T-DNA insertion alleles of *nrpb2*, named *emb1989-1* and *emb1989-2*. These mutant alleles behave as recessive embryonic lethals, not female gametophytic lethals. Homozygous mutant seeds typically contain embryos arrested at preglobular or early globular stages of development. Analysis of NRPB2 sequences flanking the T-DNA insert in *emb1989-1* confirmed that one T-DNA border is located 34 nucleotides upstream from the translation start site (Y. Onodera, data not shown). The location of a second border noted at SeedGenes remains unresolved. Because the protein coding region is not disrupted, it is possible that the *emb 1989-1* allele is partially functional, which may explain how development can sometimes proceed to stages beyond what we have observed for the *nrpb2-1* and *nrpb2-2* alleles. We currently lack analogous data concerning the precise location of the T-DNA in the *emb 1989-2* allele.

Genetics **181:** 353 (January 2009)

APPENDIX D

RNA POLYMERASE V GUIDES AGO4 TO CHROMATIN

Published in *Nature Genetics*, vol. 41;5 630-634

My contributions to this work:

I designed and constructed the polyclonal AGO4 antibody that was instrumental in performing the RNA and DNA ChIP analyses to show that AGO4 associates with Pol V chromatin templates and Pol V RNA transcripts. I also made comments on the manuscript and gave technical advice.

nature genetics

RNA polymerase V transcription guides ARGONAUTE4 to chromatin

Andrzej T Wierzbicki, Thomas S Ream, Jeremy R Haag & Craig S Pikaard

© 2009 Nature America, Inc. All rights reserved.

Retrotransposons and repetitive DNA elements in eukaryotes are silenced by small RNA–directed heterochromatin formation. In *Arabidopsis,* this process involves 24-nt siRNAs that bind to ARGONAUTE4 (AGO4) and facilitate the targeting of complementary loci[1,2] via unknown mechanisms. Nuclear RNA polymerase V (Pol V) is an RNA silencing enzyme recently shown to generate noncoding transcripts at loci silenced by 24-nt siRNAs[3]. We show that AGO4 physically interacts with these Pol V transcripts and is thereby recruited to the corresponding chromatin. We further show that DEFECTIVE IN MERISTEM SILENCING3 (DMS3), a structural maintenance of chromosomes (SMC) hinge-domain protein[4], functions in the assembly of Pol V transcription initiation or elongation complexes. Collectively, our data suggest that AGO4 is guided to target loci through base-pairing of associated siRNAs with nascent Pol V transcripts.

Arabidopsis Pol V, AGO4 (ref. 5), DMS3 (ref. 4) and the putative chromatin remodeller DRD1 (ref. 6) function in the silencing of siRNA-homologous loci at one or more steps downstream of siRNA biogenesis[3,7–10]. Recently, we showed that DRD1 facilitates Pol V transcription of noncoding RNAs at target loci, revealing a functional relationship between these two activities[3]. However, the functional relationships, if any, between AGO4, DMS3 and Pol V transcription are unclear.

Mutations disrupting *NRPE1* (encoding the largest Pol V subunit), *AGO4* or *DMS3* cause similar losses of RNA-directed DNA methylation at *AtSN1* retrotransposons, *IGN5* (*INTERGENIC REGION 5*) and a retroelement *solo LTR* locus (**Fig. 1a**,**b**). Likewise, histone H3 lysine 27 monomethylation (H3K27me1), a characteristic of silenced heterochromatin, is reduced at these loci in *nrpe1*, *ago4* and *dms3* mutants compared to wild-type plants (ecotype Col-0) (**Fig. 1c**). These results indicate that Pol V, AGO4 and DMS3 collaborate in the establishment of repressive chromatin modifications. At the *solo LTR* locus transcribed by RNA polymerase II (Pol II), chromatin immunoprecipitation (ChIP) shows that levels of diacetylated histone H3 (H3Ac2; acetylated on lysines 9 and 14), a mark of active chromatin, increase in the mutants (**Fig. 1d**), coincident with increased Pol II occupancy of the locus (**Fig. 1e**; compare to no-antibody controls in **Fig. 1f**). At *IGN5* and *AtSN1*, which lack associated Pol II (**Fig. 1e**), no increase in histone H3 acetylation is observed in the mutants (**Fig. 1d**). *AtSN1* elements are thought to be transcribed by Pol III; therefore, differences in H3 acetylation at the *solo LTR* and *AtSN1* loci may reflect the different polymerases involved.

AGO4 and Pol V colocalize in a nucleolus-associated Cajal body[7,8] that is distant from the target loci subjected to siRNA-mediated silencing. These observations have suggested that AGO4–siRNA complexes might guide Pol V to the target loci[7,8]. To test this hypothesis, we asked whether production of Pol V transcripts is AGO4 dependent. At intergenic regions *IGN5* and *IGN6* (ref. 3), Pol V transcripts are lost or substantially reduced in the Pol V mutant (*nrpe1*) but not in the *ago4* mutant (**Fig. 2a**); in fact, *IGN5* transcript levels increase by ~50% in *ago4* (**Fig. 2b**). This increase in transcript levels is dependent on Pol V, as shown by analysis of the *nrpe1 ago4* double mutant (**Fig. 2a**). In the *rdr2 (rna-dependent rna polymerase 2)* mutant, which abolishes 24-nt siRNA biogenesis[11,12], or in an *rdr2 ago4* double mutant, Pol V transcript levels are unaffected compared to wild-type (Col-0) plants. We conclude that AGO4–siRNA complexes are dispensable for Pol V transcription at target loci, arguing against the hypothesis that AGO4–siRNA complexes guide Pol V to target loci. The functional significance of AGO4 and Pol V colocalization in Cajal bodies is unclear but could reflect independent protein processing/assembly or storage functions that are unrelated to RNA-induced silencing complex (RISC) assembly.

To test an alternative hypothesis, that AGO4–siRNA complexes are recruited to chromatin in a Pol V-dependent manner, we assayed AGO4 associations with target loci using ChIP (**Fig. 3**). In wild-type (Col-0) plants, *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci are all enriched upon AGO4-ChIP, whereas only background levels are observed in *ago4* or *nrpe1* mutants or in control ChIP reactions lacking antibody to AGO4 (anti-AGO4, **Fig. 3a**). These findings indicate that AGO4 interacts with target locus chromatin and does so in a Pol V–dependent manner. AGO4–chromatin interactions are not diminished by mutation of *DRM2* (**Fig. 3a**), which encodes the *de novo* DNA methyltransferase that carries out siRNA and AGO4-dependent cytosine methylation[13,14]. Collectively, these data indicate that Pol V, but not preexisting DNA methylation, is required to recruit AGO4 to chromatin.

To test whether Pol V enzymatic activity is required for AGO4 binding to chromatin, we examined AGO4–chromatin associations in *nrpe1* mutants that had been transformed with either a full-length,

Biology Department, Washington University, St. Louis, Missouri, USA. Correspondence should be addressed to C.S.P. (pikaard@biology.wustl.edu).

Received 1 December 2008; accepted 25 February 2009; published online 19 April 2009; doi:10.1038/ng.365

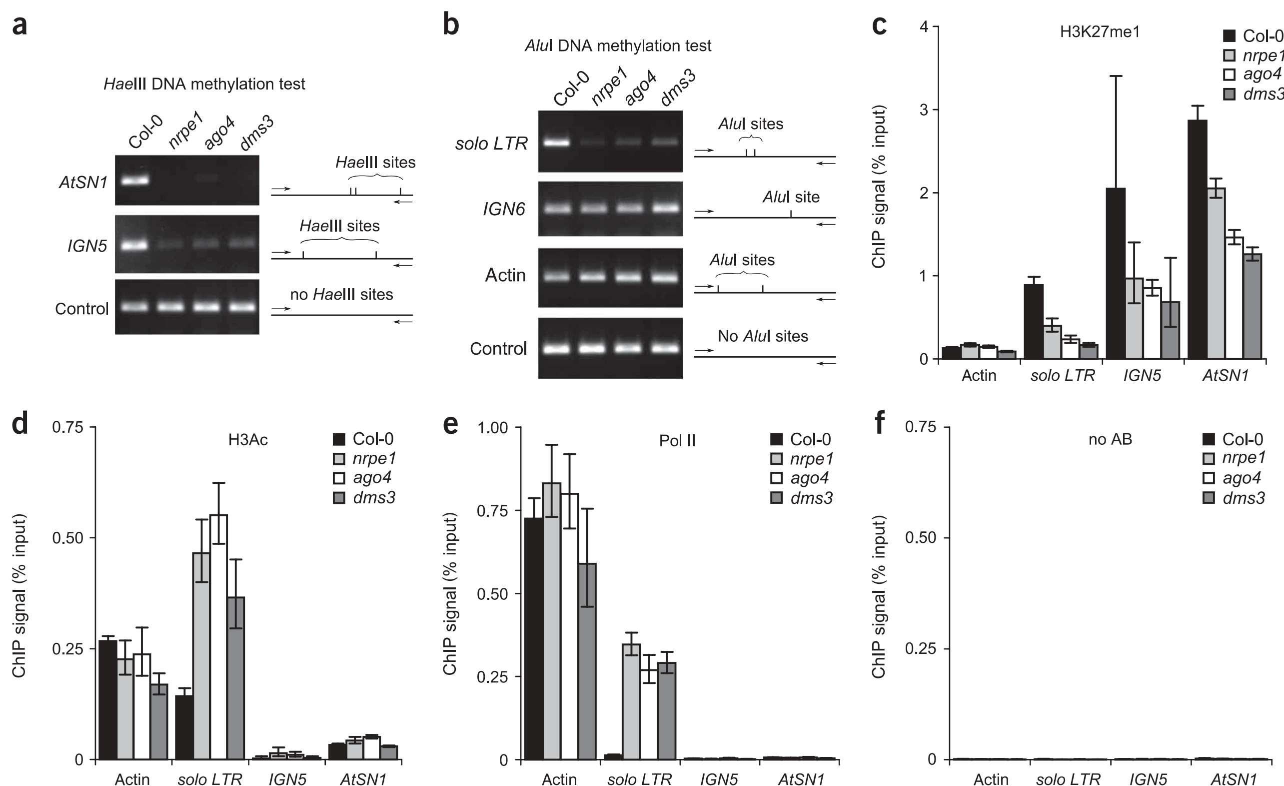


Figure 1 Pol V, AGO4 and DMS3 work nonredundantly in heterochromatin formation. (**a**,**b**) DNA methylation analysis at the *AtSN1*, *IGN5* and *solo LTR* loci in *nrpe1*, *ago4* and *dms3* mutants. Genomic DNA was digested with *Hae*III (**a**) or *Alu*I (**b**) methylation-sensitive restriction endonucleases followed by PCR. Sequences lacking *Hae*III sites (actin 2; **a**) or *Alu*I sites (*IGN5*, **b**) served as controls to show that equivalent amounts of DNA were tested in all reactions. (**c**,**d**) ChIP analysis of H3K27me1 (**c**) and H3Ac2 (**d**) levels in *nrpe1*, *ago4* and *dms3* mutants. Histograms show means ± s.d. obtained from three independent amplifications. (**e**) ChIP analysis of Pol II binding to chromatin in *nrpe1*, *ago4* and *dms3* mutants. Histograms show means ± s.d. obtained from three independent amplifications. (**f**) Control ChIP reactions carried out in the absence of antibody reveal background signal levels.

© 2009 Nature America, Inc. All rights reserved.

wild-type *NRPE1* transgene or an equivalent transgene bearing point mutations within the metal A motif of the active site (*NPRE1 ASM* transgene). The active site point mutations do not affect NRPE1 stability or its association with the second-largest subunit but eliminate Pol V transcripts and Pol V biological activity[3,15]. Whereas the wild-type *NRPE1* genomic transgene (*NRPE1 wt*) restored AGO4 interaction with the *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in the *nrpe1* mutant background (**Fig. 3b**), the active-site mutant (*NRPE1 ASM*) failed to do so. Immunoblotting ruled out the trivial explanation that AGO4 protein levels might be differentially affected by the *nrpe1* mutation or the *NRPE1* transgenes (**Fig. 3c**) and also demonstrated that the antibody specifically recognizes AGO4, which is absent in the *ago4* mutant. Collectively, the data indicate that Pol V transcriptional activity is required to recruit AGO4 to chromatin.

Base-pairing between AGO4-associated siRNAs and nascent Pol V transcripts could be a mechanism by which Pol V transcription recruits AGO4 to target loci. To test this hypothesis, we used RNA immunoprecipitation to ask whether AGO4 associates with Pol V transcripts *in vivo*. In wild-type (Col-0) plants, anti-AGO4 immunoprecipitates *IGN5* and *IGN6* Pol V transcripts[3] (**Fig. 4a**). Important controls show that Pol V transcripts are not immunoprecipitated in the *ago4* or *nrpe1* mutant backgrounds. Anti-AGO4

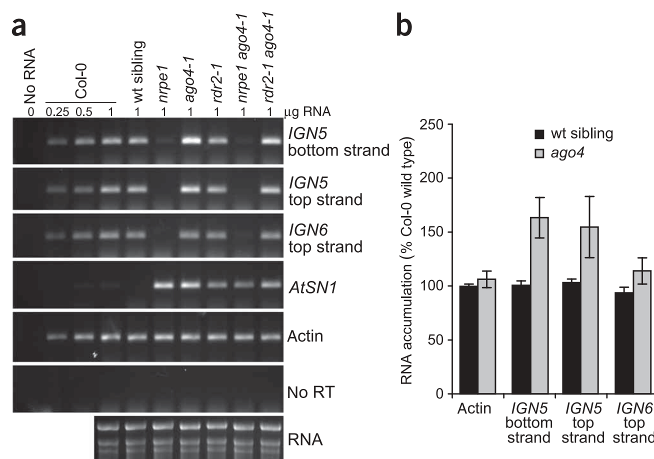


Figure 2 AGO4 is not required for Pol V transcription. (**a**) Strand-specific RT-PCR of Pol V transcription at *IGN5*, *IGN6* and *AtSN1* in *ago4* and *rdr2* mutants as well as *nrpe1 ago4* and *rdr2 ago4* double mutants. Wild-type sibling is a wild-type sibling of the *ago4* mutant identified in a segregating family. Actin RT-PCR products and ethidium bromide–stained rRNAs resolved by agarose gel electrophoresis serve as loading controls. To control for background DNA contamination, we carried out a reaction using *IGN5* top strand primers but no reverse transcriptase (no RT). No-RNA (0 µg) controls are provided for all primer pairs. (**b**) Densitometric analysis of RT-PCR data for the *ago4* mutant presented in **a**. The histogram provides mean band intensities relative to wild type Col-0, ± s.d. obtained from three independent experiments.

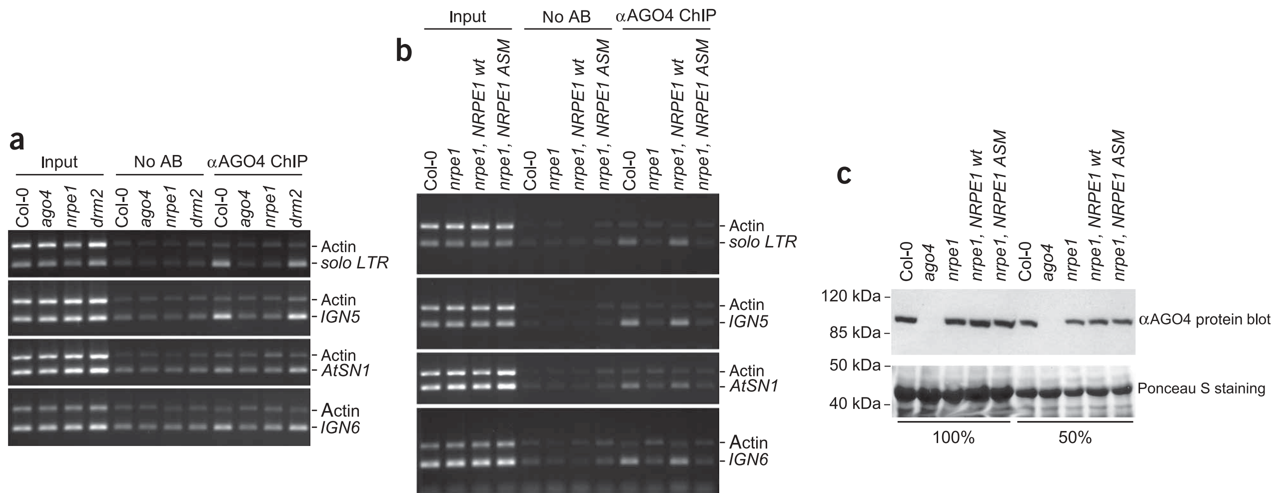


Figure 3 Pol V transcription is necessary for AGO4–chromatin interactions. (**a**) ChIP data showing AGO4 binding to chromatin at *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in *ago4*, *nrpe1* and *drm2* mutants. DNA purified from input chromatin samples, chromatin subjected to the immunoprecipitation procedure in the absence of antibody (no Ab) and chromatin immunoprecipitated using anti-AGO4 (αAGO4) was amplified by PCR using locus-specific primers. Primers amplifying the *Actin2* locus served as an internal control. (**b**) ChIP data showing AGO4 binding to chromatin at *solo LTR*, *IGN5*, *AtSN1* and *IGN6* loci in *nrpe1* mutant, *nrpe1* mutant transformed with a wild-type *NRPE1* transgene (*NRPE1 wt*), and *nrpe1* mutant transformed with an *NRPE1* active site mutant transgene (*NRPE1 ASM*). (**c**) Immunoblot detection of AGO4 in protein extracts of wild type (Col-0), *ago4*, *nrpe1*, or *nrpe1* transformed with either a wild-type *NRPE1* transgene (*NRPE1 wt*) or an *NRPE1* active site mutant transgene (*NRPE1 ASM*). Ponceau S staining revealed equal loading of lanes; 100% and 50% sample loadings indicate that the assay is semiquantitative.

© 2009 Nature America, Inc. All rights reserved.

immunoprecipitation of *IGN5* or *IGN6* RNAs was also reduced or eliminated in *rdr2* mutant plants, indicating that AGO4–Pol V transcript interactions are dependent on siRNAs. However, in the absence of siRNA biogenesis, as in the *rdr2*, *nrpd1*, *nrpd2/nrpe2* or *dcl2,3,4* mutants, AGO4 protein levels drop below the limits of immunoblot detection[7,8] (**Fig. 4b–d**). By contrast, AGO4 protein levels are unaffected in *nrpe1* (**Fig. 4b–d**) or *drm2* mutants (ref. 7), which act downstream of siRNA biogenesis. The instability of AGO4 in the absence of siRNAs complicates the interpretation of these results. Although we favor the hypothesis that siRNA–Pol V transcript base-pairing is responsible for AGO4 association with Pol V transcripts, we cannot rule out the possibility that AGO4 binds Pol V transcripts directly, with siRNAs merely being required for AGO4 stability.

DMS3 was recently identified as a gene required for RNA-directed DNA methylation that acts at an unspecified step downstream of siRNA biogenesis[4]. The encoded protein shares sequence similarity with the hinge-domain regions of SMC proteins, such as the core proteins of cohesin and condensin complexes[16], suggesting a chromatin-related function. We found that at *IGN5*, *IGN6* and *AtSN1* loci, Pol V transcripts are substantially reduced or absent in *dms3* mutant plants, as in *nrpe1* (**Fig. 5a**) or *drd1* mutants[3]. Likewise, transcriptional suppression of *AtSN1* and *solo LTR* elements is similarly disrupted in *dms3* and *nrpe1* mutants (**Fig. 5b**). ChIP using an antibody to NRPE1 revealed that, in the *dms3* mutant, Pol V–chromatin associations are reduced to background levels, resembling the actin and *nrpe1* mutant controls (**Fig. 5c**). Collectively, these data (**Fig. 5**) indicate that DMS3 is required for Pol V transcription, as shown previously for the chromatin remodeller DRD1 (ref. 3). The loss of detectable Pol V–chromatin association in *dms3* or *drd1* mutants suggests that these chromatin proteins participate in the assembly of Pol V transcription complexes.

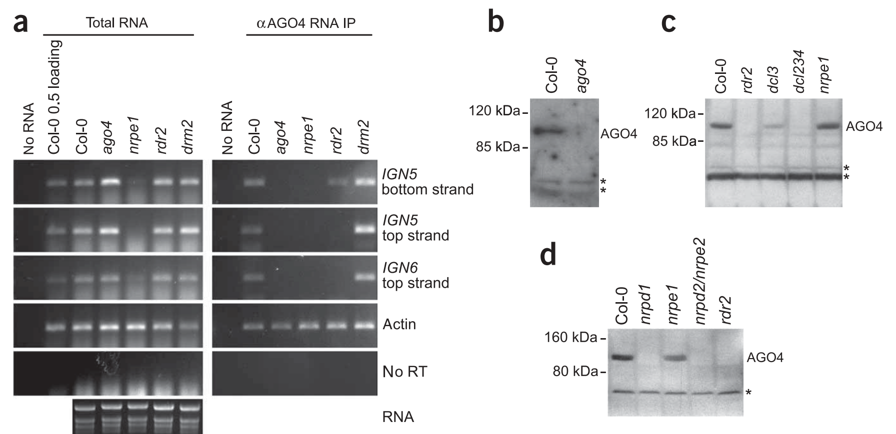


Figure 4 AGO4 physically interacts with Pol V transcripts. (**a**) RNA immunoprecipitation using anti-AGO4 (αAGO4). Immunoprecipitated RNA isolated from the indicated mutants was digested with DNaseI and amplified by RT-PCR. Total RNA controls show that the Pol V transcripts are present in equivalent amounts in all mutants tested except *nrpe1*. Ethidium bromide–stained rRNAs (bottom left) show that equal amounts of RNA were tested. The no reverse transcriptase (no RT) control was done with *IGN5* bottom-strand primers. No-RNA controls were carried out for all primer pairs tested. RT-PCR amplification of actin RNA serves as a loading control. (**b**) Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0) plants or *ago4* mutant. Asterisks denote nonspecific bands. (**c**) Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0), *rdr2*, *dcl3*, *dcl234* or *nrpe1* mutants. Asterisks denote nonspecific bands. (**d**) Immunoblot detection of AGO4 in protein extracts of wild-type (Col-0), *nrpd1* (Pol IV), *nrpe1* (Pol V), *nrpd2/nrpe2* (shared subunit of Pol IV and Pol V) or *rdr2* mutants. Asterisks denote nonspecific bands.

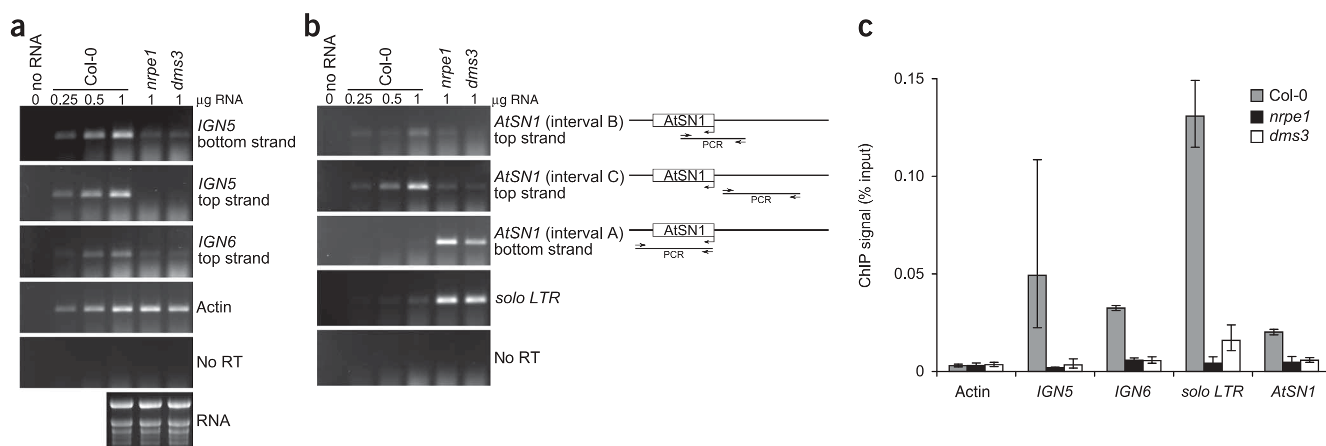


Figure 5 The SMC hinge-domain protein DMS3 is required for Pol V transcription and detectable Pol V-chromatin interactions. (**a**,**b**) Strand-specific RT-PCR detection of Pol V transcripts at *IGN5* and *IGN6* (**a**) and *AtSN1* (**b**) in wild-type (Col-0) and *nrpe1* and *dms3* mutants. Derepression of Pol II transcripts at the solo LTR and putative Pol III transcripts at *AtSN1* in the *nrpe1* and *dms3* mutants is shown in the right panel. Actin RT-PCR products and ethidium bromide–stained rRNAs resolved by agarose gel electrophoresis serve as loading controls. To control for background DNA contamination, we carried out a reaction using *IGN5* bottom strand (**a**) or *AtSN1* (interval B) primers (**b**) but no reverse transcriptase (no RT). No-RNA (0 µg) controls are provided for all primer pairs. (**c**) ChIP with anti-NRPE1 in Col-0 wild-type, *nrpe1* and *dms3* mutants followed by real-time PCR. Histograms show means ± s.d. obtained from three independent amplifications.

© 2009 Nature America, Inc. All rights reserved.

npg

Our results suggest that siRNAs and Pol V transcripts are produced by independent pathways that intersect to bring about heterochromatin formation and gene silencing (**Fig. 6**). In one pathway, Pol IV, RDR2 and DCL3 collaborate to produce 24-nt siRNAs that associate with AGO4 (ref. 1). Independent of this pathway, DRD1 and DMS3 facilitate noncoding Pol V transcription at target loci. AGO4's interaction with Pol V transcripts, and the fact that AGO4 association with chromatin requires the Pol V active site, suggests that siRNA–AGO4 complexes are guided to target loci by interacting with Pol V transcripts. It has also been reported that AGO4 can interact with the C-terminal domain (CTD) of NRPE1 *in vitro*[7,17] and *in vivo*[7], suggesting that Pol V might recruit AGO4 directly, in an RNA-independent manner. However, we have been unable to detect

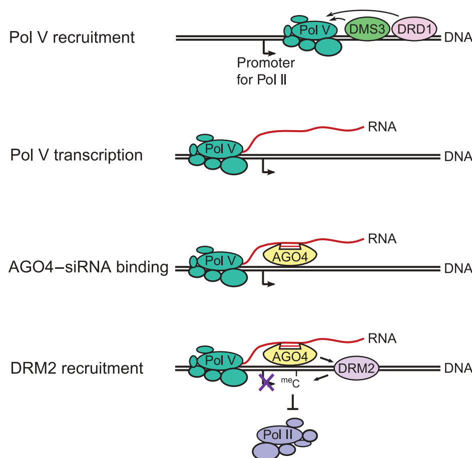


Figure 6 A model for Pol V and siRNA-dependent heterochromatin formation. DMS3 and DRD1 mediate the assembly of Pol V initiation and/or elongation complexes and the production of Pol V transcripts. AGO4–siRNA complexes recognize target loci via base-pairing of siRNAs with nascent Pol V transcripts. AGO4 subsequently recruits chromatin modifying activities including the de novo DNA methyltransferase DRM2 and histone modifying enzymes via unknown mechanisms.

AGO4–Pol V associations *in vivo* using immunoprecipitation and subsequent immunoblotting nor by mass spectrometric analysis of affinity-purified Pol V (data not shown), suggesting that any interactions between AGO4 and Pol V may be weak or transient. We suggest that AGO4 recruitment to chromatin is primarily an RNA-mediated process but may also involve protein–protein interactions.

In fission yeast, artificial tethering of the RNA-induced transcriptional silencing (RITS) complex to *ura4* pre-mRNAs is sufficient to induce heterochromatin formation at the normally euchromatic $ura4^+$ locus[18]. These and other results are consistent with the hypothesis that fission yeast silencing complexes are guided to chromatin via associations with nascent Pol II transcripts[19]. Our findings suggest that plants and yeast are fundamentally similar in their use of RNA guidance mechanisms for recruiting Argonaute-containing transcriptional silencing complexes to target loci. It is intriguing that plants should have evolved a unique RNA polymerase, Pol V, whose specialized role seems to be the generation of noncoding RNAs that can serve as scaffolds for Argonaute recruitment.

METHODS

Plant strains. *Arabidopsis thaliana nrpe1* (*nrpd1b-11*) was described previously[8]. The *dms3-4* mutant (SALK_125019C) of locus At3g49250 was obtained from the *Arabidopsis* Biological Resource Center. The *dcl2*, *dcl3*, *dcl4* triple mutant (*dcl2,3,4*) was provided by T. Blevins (Washington University, St. Louis). The *ago4-1* mutant (Ler ecotype background) was provided by S. Jacobsen (University of California, Los Angeles) and was introgressed into the Col-0 background by three rounds of backcrossing.

Antibodies. Anti-Pol II (anti-NRPB2) was described previously[20]. Anti-H3K27me1 #8835 (ref. 21) was provided by T. Jenuwein (Max Planck Institute of Immunobiology). Antibody against diacetyl-H3 (K9 and K14) was obtained from Millipore (cat. #06599, lot #JBC1349702). Rabbit anti-NRPE1 has been described[9]. Rabbit anti-AGO4 was raised against a C-terminal portion of the protein (amino acids 573–924) expressed in bacteria.

RNA and DNA analysis. RNA isolation, RT-PCR and real-time quantitative PCR were carried out as described[3] except that real-time quantitative PCR analysis of the *IGN5* locus was done using the following oligonucleotide primers: A195, 5′-ACATGAAGAAAGCCCAAACCA-3′; A196,

5′-GGCCGAATAACAGCAAGTCCT-3′. Densitometric analysis of DNA resolved by agarose gel electrophoresis was performed using ImageJ.

ChIP and RNA IP. ChIP and RNA IP were carried out as described[3] except that for ChIP with anti-AGO4, RNase A was added during immunoprecipiation, washes with TE buffer were omitted, immune complexes were eluted with 100 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS for 10 min at room temperature and a second elution at 65 °C was performed. Crosslinking was reversed at 65 °C for 1 h in the presence of 40 μg Proteinase K (Invitrogen). DNA was purified by extraction with phenol:chloroform and ethanol precipitation. DNA recovery was assayed by PCR using 1.5 u Platinum Taq (Invitrogen).

ACKNOWLEDGMENTS
Our work is supported by US National Institutes of Health grant GM077590. The content of the paper is the sole responsibility of the authors and does not necessarily reflect the views of the NIH. We thank T. Blevins (Washington University, St. Louis), S. Jacobsen (University of California, Los Angeles) and T. Jenuwein (Max Planck Institute of Immunobiology) for providing reagents.

AUTHOR CONTRIBUTIONS
T.S.R. generated anti-AGO4; J.R.H. and T.S.R. assayed NRPE1–AGO4 interactions; J.R.H. produced **Figure 4d**; A.T.W. performed all remaining experiments. A.T.W. and C.S.P. wrote the manuscript.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

© 2009 Nature America, Inc. All rights reserved.

1. Chapman, E.J. & Carrington, J.C. Specialization and evolution of endogenous small RNA pathways. *Nat. Rev. Genet.* **8**, 884–896 (2007).
2. Girard, A. & Hannon, G.J. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol.* **18**, 136–148 (2008).
3. Wierzbicki, A.T., Haag, J.R. & Pikaard, C.S. Noncoding transcription by RNA Polymerase Pol IVb/ Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* **135**, 635–648 (2008).
4. Kanno, T. *et al.* A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nat. Genet.* **40**, 670–675 (2008).
5. Vaucheret, H. Plant ARGONAUTES. *Trends Plant Sci.* **13**, 350–358 (2008).
6. Kanno, T. *et al.* A SNF2-like protein facilitates dynamic control of DNA methylation. *EMBO Rep.* **6**, 649–655 (2005).
7. Li, C.F. *et al.* An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in *Arabidopsis thaliana*. *Cell* **126**, 93–106 (2006).
8. Pontes, O. *et al.* The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79–92 (2006).
9. Pontier, D. *et al.* Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* **19**, 2030–2040 (2005).
10. Kanno, T. *et al.* Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **37**, 761–765 (2005).
11. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2**, e104 (2004).
12. Kasschau, K.D. *et al.* Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* **5**, e57 (2007).
13. Cao, X. *et al.* Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Curr. Biol.* **13**, 2212–2217 (2003).
14. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
15. Haag, J.R., Pontes, O. & Pikaard, C.S. Metal A and metal B sites of nuclear RNA polymerases Pol IV and Pol V are required for siRNA-dependent DNA methylation and gene silencing. *PLoS ONE* **4**, e4110 (2009).
16. Peric-Hupkes, D. & van Steensel, B. Linking cohesin to gene regulation. *Cell* **132**, 925–928 (2008).
17. El-Shami, M. *et al.* Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* **21**, 2539–2544 (2007).
18. Bühler, M., Verdel, A. & Moazed, D. Tethering RITS to a nascent transcript initiates RNAi- and heterochromatin-dependent gene silencing. *Cell* **125**, 873–886 (2006).
19. Bühler, M. & Moazed, D. Transcription and RNAi in heterochromatic gene silencing. *Nat. Struct. Mol. Biol.* **14**, 1041–1048 (2007).
20. Onodera, Y. *et al.* Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* **120**, 613–622 (2005).
21. Peters, A.H.F.M. *et al.* Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* **12**, 1577–1589 (2003).

APPENDIX E

CURRICULUM VITAE

Thomas S. Ream

444 N. Clay #3N
Kirkwood, MO 63122
tsream@artsci.wustl.edu
(314) 909-4145 (home)
(314) 489-6795 (cell)
(314) 935-8529 (office)
Date of birth: August 9, 1980

Washington University in St. Louis
Plant Biology Graduate Program
Division of Biology and Biomedical Sciences
Campus Box 1137, 1 Brookings Drive
St. Louis, MO 63122

EDUCATION:

Ph.D. **Washington University in St. Louis**
Plant Biology, April 23, 2009

B.S. **University of Missouri, Columbia**
Biological Sciences, December 2002

PhD ABSTRACT:

In addition to RNA polymerases I, II and III, the essential RNA polymerases present in all eukaryotes, plants have two additional nuclear RNA polymerases, abbreviated as Pol IV and Pol V. Pol IV and Pol V play non-redundant roles in siRNA-directed DNA methylation and gene silencing in *Arabidopsis*. Using a combination of affinity purification and protein identification by mass spectrometry, my studies show that Pol IV and Pol V are composed of RNA polymerase subunits that are parologous or identical to the twelve subunits of Pol II. Four subunits of Pol IV are distinct from their Pol II paralogs, six subunits of Pol V are distinct from their Pol II paralogs, and four subunits differ between Pol IV and Pol V. Importantly, the subunit differences occur in key positions relative to the template entry and RNA exit paths. My findings support the hypothesis that Pol IV and Pol V are Pol II-like enzymes that evolved specialized roles in the production of noncoding transcripts for RNA silencing and genome defense. In addition, I have determined the subunit compositions of *Arabidopsis* RNA polymerases I, II and III, providing a novel framework for comparative and functional analyses of subunits from all five DNA-dependent RNA polymerases in plants.

RESEARCH EXPERIENCE:

2004-present **Ph.D. Thesis.** Characterization of Pol IV and Pol V subunits in RNA-mediated gene silencing and heterochromatin formation in *Arabidopsis*. Determined the subunit compositions of all five DNA-dependent RNA polymerases in *Arabidopsis*.

2004 (summer) **Ph.D. rotation.** Sigma-Aldrich (St. Louis), Plant Biotechnology Research & Development Team. Tested FLAG and c-myc affinity resins for their utility in purifying plant protein complexes.

2003 **Research technician.** Collaborated with Monsanto researchers to test for ectopic Cre/Lox-mediated recombination in maize (James Birchler laboratory).

2001-2003 **Undergraduate research associate.** Analyzed gene expression patterns in different ploidy series of maize inbreds vs. hybrids (James Birchler laboratory).

RESEARCH SKILLS:

Epitope tagging and affinity purification of protein complexes from plants.

Experience with DNA methylation and small RNA analyses, RT-PCR, genotyping, DNA sequencing, and collaborative mass spectrometry analyses.

Developed protocols for protein expression, immunoprecipitation, purification and characterization in plants and bacteria. Experience in using various chromatographic methods, FPLC and standard plant molecular biology protocols.

Experience with antibody design, antigen production and affinity purification.

TEACHING:

Washington University

Teaching Assistant **Plant Cells and Proteins**
A protein engineering lab course which emphasized cloning, purification and characterization of recombinant proteins. Designed and prepared laboratory exercises and gave lectures.

Teaching Assistant **Plant biology and genetic engineering**
A lecture based course focusing on plant development, plant molecular biology techniques and special topics in plant biology. Assisted with grading, test design, lecturing and conducted student help sessions.

University of Missouri

Peer Learning Assistant **Plant Taxonomy**
A lab course highlighting the diversity of the plant kingdom by observation and examination of different plant families from the micro- to the

macro-level. Introduction to using taxonomic keys to identify plant species. Assisted students with plant species and organ identification. Led interpretive hikes to identify native plant species.

PUBLICATIONS:

First author publications:

Ream, T.S., Pontvianne, F., Nicora, C.D., Norbeck, A.D., Haag, J.R., Pasa-Tolic, L., and Pikaard, C.S. (2009). Subunit compositions of *Arabidopsis* RNA polymerases I and III reveal insights into the evolution, functional diversification and redundancy of subunits among all five DNA-dependent RNA polymerases. **In preparation**.

Ream, T.S., Haag, J.R., Wierzbicki, A.T., Nicora, C.D., Norbeck, A., Zhu, J.K., Hagen, G., Guilfoyle, T.J., Pasa-Tolic, L., and Pikaard, C.S. (2008). Subunit Compositions of the RNA-Silencing Enzymes Pol IV and Pol V Reveal Their Origins as Specialized Forms of RNA Polymerase II. **Mol Cell**.

Onodera, Y.*, Haag, J.R.*, **Ream, T.***, Nunes, P.C., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. **Cell** *120*, 613-622.
*co-first authors

Ream, T.S., Strobel, J., Roller, B., Auger, D.L., Kato, A., Halbrook, C., Peters, E.M., Theuri, J., Bauer, M.J., Addae, P., *et al.* (2005). A test for ectopic exchange catalyzed by Cre recombinase in maize. **Theor Appl Genet** *111*, 378-385.

Other publications:

Haag, J.R., **Ream, T.S.**, Pontes, O., Nicora, C.D., Norbeck, A.D., Pasa-Tolic, L., and Pikaard, C.S. (2009) RNA silencing enzymes Pol IV and RDR2 are physically coupled for dsRNA production. **In preparation**.

Wierzbicki, A.T., **Ream, T.S.**, Haag, J.R., and Pikaard, C.S. (2009). RNA Polymerase V transcription guides ARGONAUTE4 to chromatin. **Nature Genetics** *41*, 630-634.

Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., **Ream, T.**, Ito, Y., and Pikaard, C.S. (2008). Sex-biased lethality or transmission of defective transcription machinery in *Arabidopsis*. **Genetics** *180*, 207-218.

Pikaard, C.S., Haag, J.R., **Ream, T.**, and Wierzbicki, A.T. (2008). Roles of RNA polymerase IV in gene silencing. **Trends Plant Sci** *13*, 390-397. (Review)

Lough, A.N., Roark, L.M., Kato, A., **Ream, T.S.**, Lamb, J.C., Birchler, J.A., and Newton, K.J. (2008). Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. **Genetics** *178*, 47-55.

Pontes, O., Li, C.F., Nunes, P.C., Haag, J., **Ream, T.**, Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The *Arabidopsis* chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. **Cell** *126*, 79-92.

Auger, D.L., Gray, A.D., **Ream, T.S.**, Kato, A., Coe, E.H., Jr., and Birchler, J.A. (2005). Nonadditive gene expression in diploid and triploid hybrids of maize. **Genetics** *169*, 389-397.

Auger, D.L., **Ream, T.S.**, and Birchler, J.A. (2004). A test for a metastable epigenetic component of heterosis using haploid induction in maize. **Theor Appl Genet** *108*, 1017-1023.

CONFERENCES:

Speaker:

2009 17th International Chromosome Conference (June 2009-Boone, NC)
Tentative title: "*Subunit structures of the RNA silencing enzymes, RNA Polymerase IV and V, reveal their origins as specialized forms of RNA Polymerase II*"

First author posters have been presented at:

2007 24th Symposium in Plant Biology (Riverside, CA)
"Subunit structures of RNA polymerase IVa and IVb in *Arabidopsis*."

2005 International *Arabidopsis* Conference (Madison, WI)
"Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation."

2003 Maize Genetics Conference (Lake Geneva, WI)
"A test for a metastable epigenetic component of heterosis using haploid induction in maize."

2002 Maize Genetics Conference (Orlando, FL)
"Nonadditive gene expression in diploid and triploid hybrids of maize."

Unlisted: Numerous graduate presentations, including journal clubs, departmental research seminars and program retreats.

AWARDS AND HONORS:

Environmental Molecular Sciences Laboratory Grant (USDOE partnership)
A grant awarded by Pacific Northwest National Lab in conjunction with the Department of Agriculture to analyze complex protein mixtures using mass spectrometry to identify protein complexes (used for Pol IV and Pol V purification).

Professor Stanley Zimmering Prize in Biology (University of Missouri)

The Professor Stanley Zimmering Prize in Biology is awarded annually to one or more graduating seniors with outstanding records who plan academic careers in research and teaching, and who have been accepted into PhD programs in genetics or molecular biology at universities outside the University of Missouri system.

Undergraduate Major Field Test (>90th percentile in Biology)

Magna cum laude (University of Missouri)