1-1-2011

Molecular Genetic Analysis of Non-Catalytic Pol IV and V Subunits

Ek Han Tan
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Program in Molecular Genetics and Genomics

Dissertation Examination Committee:
Craig S. Pikaard, Ph.D., co-Chair
Douglas L. Chalker, Ph.D., co-Chair
Sarah C.R. Elgin, Ph.D.
Susana Gonzalo, Ph.D.
Joseph M. Jez, Ph.D.
Robert G. Kranz, Ph.D.

MOLECULAR GENETIC ANALYSIS OF NON-CATALYTIC POL IV AND V SUBUNITS IN ARABIDOPSIS

by

Ek Han Tan

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 2011

Saint Louis, Missouri

ABSTRACT OF DISSERTATION

Molecular genetic analysis of non-catalytic RNA polymerase IV and V subunits in *Arabidopsis*

by

Ek Han Tan

Doctor of Philosophy in Molecular Genetics and Genomics

Washington University in St. Louis, 2011

Professor Craig S. Pikaard, Ph.D., co-Chair

Professor Doug L. Chalker, Ph.D., co-Chair

Among eukaryotes, plants have the distinction of encoding multisubunit RNA polymerases used exclusively for RNA directed DNA Methylation (RdDM) in addition to Pol I, II, and III. In *Arabidopsis thaliana*, Pol IV is required for the biogenesis of 24nt siRNAs whereas Pol V transcription is needed for cytosine methylation of the DNA sequences corresponding to these siRNAs. The ancestry of Pol IV and V can be traced back to Pol II, and Pol II, IV and V still utilize multiple non-catalytic subunits encoded by the same genes. Genetic analysis of non-catalytic subunits that are highly similar reveals that these subunits are not necessarily redundant. For instance, NRPB9b but not its 97% similar paralog, NRPB9a is required for RdDM. Likewise, Pol IV and Pol V-specific 7th largest subunits are very similar yet have different involvements in RdDM. In some of the non-catalytic subunit mutants of Pol IV, 24nt siRNA accumulation is not dramatically reduced, yet RNA silencing is disrupted. This contrasts with Pol IV catalytic subunit mutants in which siRNA biogenesis and RdDM are coordinately disrupted. Taken

together, these results suggest that Pol IV might possess functions in RdDM that are in addition to, and separable from siRNA biogenesis. Differences in Pol V subunit composition based on the use of alternative non-catalytic subunit variants might also have functional consequences for RdDM. The evidence suggests that alternative non-catalytic subunits in Pol IV and V are likely to influence interactions with other proteins needed for RdDM.

Many thanks I owe,

my fam, friends, mentors.. new paths,

to run along now.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER ONE

INTRODUCTION

i.

PROLOGUE

The genomes of prokarya, archaea and eukarya are transcribed by structurally conserved, DNA dependent RNA polymerase complexes. In concordance with the central dogma, transcription of DNA into RNA is catalyzed by these RNA polymerases (Pol II in eukaryotes), allowing for the translation of the resulting messenger RNA into protein. While prokaryotes and archaea use a single RNA polymerase for all transcriptional activities, eukaryotes use three functionally divergent RNA polymerases: Pol I, II and III, transcripts from two of which don't make proteins (Pol I and III).

A pivotal shift from the dogmatic views on RNAs was sparked by the discovery of RNA interference, a method of gene silencing mediated by short, 20-40 nucleotide small RNA species. Since then, many analogous small RNA pathways, which are very diverse in eukaryotes, have been characterized. Among the most astounding discoveries is that plants encode dedicated RNA silencing RNA polymerases in addition to RNA polymerase I, II and III. In *Arabidopsis*, two additional RNA polymerases, Pol IV and Pol V are involved in a silencing pathway known as RNA directed DNA methylation, where 24nt siRNAs produced by Pol IV are targeted for de novo DNA methylation through Pol V.

At the molecular level, epigenetic phenomena associated with small RNAs include de novo DNA methylation and targeting of repressive histone modifications. I will begin this introduction with a broad survey of RNA silencing pathways, followed by closer look at DNA methylation before reviewing current views on the plant-specific RNA polymerases, Pol IV and Pol V.

GLOSSARY OF TERMS

Adapted from (Pikaard, Haag et al. 2008) and updated to reflect recent advances.

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.

AGO1: ARGONAUTE1, binds primarily 21nt microRNAs and is involved in post-transcriptional gene silencing

AGO4: ARGONAUTE4, binds primarily 24nt heterochromatic siRNAs and is involved in RNA directed DNA methylation

CLSY1: CLASSY1, a putative chromatin remodeling protein involved in RNA-directed DNA methylation.

CMT3: CHROMOMETHYLASE3, a plant specific de novo DNA methyltransferase cooperates with histone methyltransferase KYP to maintain CHG methylation.

CTD: C-terminal domain, of Pol II, Pol IV and Pol V largest subunits.

DCL: DICER-LIKE enzyme, family of RNAse III domain-containing endoribonucleases that process hairpin RNA or double-stranded RNA into duplexes of 20-26nt.

DCL1: Arabidopsis DICER-LIKE 1, required primarily for miRNA biogenesis.

DCL2: Arabidopsis DICER-LIKE 2, generates 22-nt siRNAs.

DCL3: Arabidopsis DICER-LIKE 3, generates 24-nt siRNAs.

DCL4: Arabidopsis DICER-LIKE 4, generates 21-nt siRNAs.

DDM1: DEFICIENT IN DNA METHYLATION1, a SWI/SNF chromatin remodeler required for global maintenance of CG and CHG methylation.

DME: DEMETER, maternally expressed DNA glycosylase/lyase required for active DNA demethylation in the central cell, resulting in maternal-specific gene expression.

DML: DEMETER-Like, DNA glycosylases involved in DNA methylation. Family includes ROS1, DML2 and DML3.

DMS3: A SMC-hinge domain protein required for RdDM and Pol V dependent transcription.

DNA: Deoxyribonucleic acid.

DNA methylation (m5C): Cytosine with a covalently attached methyl group at the 5 position.

DNMT1: DNA Methyltransferase 1, mammalian CG methyltransferase.

DNMT3a/3b: DNA Methyltransferase 3a and 3b, mammalian de novo methyltransferases.

DNMT3L: Catalytically inactive de novo methyltransferase, but heterodimerizes with DNMT3 for enhanced activity.

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 de novo DNA methyltransferase in *Arabidopsis*, related to mammalian DNMT3a/b.

DRM3: DOMAINS REARRANGED METHYLTRANSFERASE 3, catalytically inactive paralog of DRM2, related to mammalian DNMT3L.

dsRNA: Double-stranded RNA.

HDA6: Histone Deacetylase 6, a histone deacetylase required for transcriptional gene silencing of many loci.

HEN1: HUA ENHANCER 1; methylates the 2’ hydroxyl groups of siRNA and miRNA 3’-terminal nucleotides.

hmC: hydroxylmethylcytosine, oxidized form of methylcytosine.

HST1: HASTY1, an exportin 5 homolog implicated in nuclear export of miRNAs.

HYL1: HYPONASTIC LEAVES 1, a dsRNA-binding protein that interacts with DCL1.

IDN2: also known as RDM12. An SGS3-like coiled-coil protein involved in RdDM and displays dsRNA binding in vitro.

KTF1: SPT5-like protein with that interacts with AGO4 and Pol V for RdDM.

KYP: KRYPTONITE, histone H3K9 methyltransferase, works with CMT3 in a feed forward loop coupling CHG methylation with H3K9 methylation.

l-siRNA: long siRNA of around 40nt, as opposed to the predominant 21–24nt size range.

MBD: Methyl-CpG-Binding Domain protein, binds 5-methylcytosine residues.

MET1: Methytransferase1, the major CG maintenance DNA methyltransferase in plants.

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.

NOR: Nucleolus organizer region.

NRPB9b: Ninth largest subunit shared by Pol II, IV and V, required for RdDM.

NRPD4: Fourth largest subunit shared by Pol IV and V, required for RdDM.

NRPD7: Seventh largest subunit of Pol IV, not required for RdDM, may be substituted by NRPE7.

NRPE7: Seventh largest subunit shared by Pol IV and V, required for RdDM.

Pol I: DNA-DEPENDENT RNA POLYMERASE I, synthesizes the precursor for the three largest rRNAs.

Pol II: DNA-DEPENDENT RNA POLYMERASE II, transcribes most protein-coding genes, encoded by mRNAs as well as miRNAs.

Pol III: DNA-DEPENDENT RNA POLYMERASE III, mostly transcribes 5S rRNA genes and tRNA genes.

Pol IV: nuclear RNA polymerase IV, includes the NRPD1 and NRPD2/E2 catalytic subunits.

Pol V: nuclear RNA polymerase V, includes the NRPE1 and NRPD2/E2 catalytic subunits.

RdDM: RNA-directed DNA methylation, one of several gene silencing pathways in the nucleus.

RDM1: Methylcytosine binding protein that forms a complex with DRD1 and DMS3 and is required to generate Pol-dependent transcripts.

RDM4: also known as DMS4, an IWR1-like protein that interacts with Pol II and Pol V, and required for RdDM.

RDR2: RNA-DEPENDENT RNA POLYMERASE 2, required for the biogenesis of 24-nt siRNAs that mediate RNA-directed DNA methylation in *Arabidopsis*

RDR6: RNA-DEPENDENT RNA POLYMERASE 6, involved in the ta-siRNA, nat-siRNA and l-siRNA, transgene and anti-viral silencing, and long-distance spread of RNA silencing pathways.

RISC: RNA-induced silencing complex, includes an ARGONAUTE protein association with an siRNA (siRISC) or miRNA (miRISC).

RNA: Ribonucleic acid.

RNAP: DNA-dependent RNA polymerase.

RNP: ribonucleoprotein, a complex of RNA and proteins.

ROS1: Repressor of silencing 1, DNA glycosylase for active DNA demethylation.

ROS3: Repressor of silencing 3, RRM domain protein required for active DNA demethylation.

rRNA: ribososomal RNA, four different 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.

Spt5: A subunit of a yeast elongation complex, conserved in archaea and similar to NusG in bacteria.

SUVH: SET-domain containing proteins, H3K9 methyltransferases.

TFIIS: Transcription factor IIS, a Pol II elongation factor.

iii.

THE WORLD OF RNA SILENCING

When dsRNA was first demonstrated to induce sequence specific gene silencing in *Caenorhabditis elegans* and plants (Fire, Xu et al. 1998; Timmons and Fire 1998; Waterhouse, Graham et al. 1998), this breakthrough postulated that dsRNA could potentially be the source of a gene silencing mechanism based on the Watson-Crick base pairing of homologous nucleic acid sequences, a mechanism know as RNA interference (RNAi) (Hannon 2002). This discovery led a link between RNAi and the generation of 20-40nt small RNAs, which were discovered in plants and found to be present in diverse organisms including unicellular ciliates, fungi and animals (Hamilton and Baulcombe 1999; Carthew 2001; Hamilton, Voinnet et al. 2002; Mochizuki, Fine et al. 2002; Martienssen 2003; Pal-Bhadra, Leibovitch et al. 2004). Since RNA silencing is widespread, it is proposed to be an ancient mechanism used by eukaryotes for genome defense against viruses and virus derived repeats but has since become important for other aspects of gene regulation and nuclear organization (Tijsterman, Ketting et al. 2002; Beisel and Paro 2011). RNA silencing is able to operate at both transcriptional and post-transcriptional levels, preventing the detrimental expression of transposable elements that are abundant in eukaryotic genomes (Slotkin and Martienssen 2007). In plants, Pol IV and V activities cooperate to generate 24nt siRNAs that guide de novo DNA methylation and heterochromatic silencing of transposons and other repeats (Pikaard, Haag et al. 2008; Matzke, Kanno et al. 2009; Lahmy, Bies-Etheve et al. 2010).

Recently, small RNAs from clustered, regularly interspaced short palindromic repeat (CRISPR) loci in diverse species of bacteria and archaea have been described with some provocative parallels to RNAi in eukaryotes (Jore, Brouns et al. 2011). CRISPR small RNAs (crRNAs) are used to target the destruction of invading viruses and DNA elements via a multimeric protein complex known as Cascade (Brouns, Jore et al. 2008; van der Oost and Brouns 2009; Deltcheva, Chylinski et al. 2011; Jore, Lundgren et al. 2011). Bioinformatic studies have indicated that CRISPR loci are widespread, and are predicted from about 90% of sequenced archaeal species and 40% of bacterial species (Grissa, Vergnaud et al. 2008; Marraffini and Sontheimer 2010). In light of this finding, the world of RNA silencing may not be restricted eukaryotes, but maybe common to all three kingdoms (Karginov and Hannon 2010; Marraffini and Sontheimer 2010).

In this section, I review the fundamentals of RNAi in the context of eukaryotic RNA silencing pathways, focusing on the unique adaptations that are employed by different eukaryotes and important findings in different organisms. Nucleolar dominance will be given some focus, due to my involvement in experiments exploring this phenomenon, as well as small RNA pathways that involve Pol IV and V in *Arabidopsis*, which is the main focus of this thesis.

Epigenetic Phenomena

In most cases, the unconventional manifestations of epigenetic traits become associated with ‘phenomena’ because they cannot be explained by Mendel’s rules and can sometimes be sporadic in nature. Epigenetic phenomena can be described in the broadest sense as alternative, heritable state of gene expression or molecular function that

can be inherited from the same DNA sequences. Random X-chromosome inactivation in female eutherian mammals (illustrated by the coat color of calico cats), DNA elimination in ciliates, paramutation, nucleolar dominance and position effect variegation (PEV) in flies are a few examples of epigenetic phenomena. The examples given here have RNA silencing components which are attributed to their phenomenology, but other forms of inheritance that are not RNA-mediated exist as well. For instance, prions are proteinaceous agents that can propogate as alternative protein conformers (Tuite and Serio 2010) and the cortical inheritance of ciliary patterns in ciliates is a form of epigenetic inheritance not occurring at the level of nucleic acids (Beisson and Sonneborn 1965; Beisson 2008).

Our understanding of modern epigenetic regulation stems from studying epigenetic phenomena which has led to major discoveries of genes that regulate the nuclear genome structurally and transcriptionally (Taverna, Li et al. 2007;Henikoff 1990).

Chromatin dynamics

The basic unit of eukaryotic chromatin is a nucleosome, which is approximately 147bp of DNA wrapped around a histone octamer (two molecules each of histones H2A, H2B, H3 and H4). Nucleosomes can be packed into higher order structures that are essential for both gene regulation and chromosome function (Luger, Mader et al. 1997; Richmond and Davey 2003). The term epigenetic is often used to describe chromatin modifications, which are post-transcriptional modifications of histone tails that include methylation, acetylation, ubiquitination, sumoylation and phosphorylation (Bestor,

Chandler et al. 1994; Taverna, Li et al. 2007; Marmorstein and Trievel 2009). In addition to promoter regulatory sequences and cognate transcription factors, RNA polymerase transcription is controlled by alternate states of chromatin modification. A large number of specialized enzymes catalyze histone modifications, and small RNAs can guide these epigenetic marks to specific target sequences (Stevenson and Jarvis 2003). Long non-coding RNAs are also emerging as being important for establishing silent chromatin, providing a platform for recruitment of chromatin modifying enzymes (Hannon, Rivas et al. 2006; Tsai, Manor et al. 2010). Finally, specific H2A and H3 protein variants can be incorporated into chromatin, adding to the complexity of epigenetic regulation (Banaszynski, Allis et al. 2010).

Actively transcribed chromatin or euchromatin is typically marked by H3K4 methylation at promoter regions and histone hyperacetylation (Fuchs, Demidov et al. 2006). On the other hand, condensed, silent chromatin (heterochromatin) is not permissive for RNA polymerase II transcription (Richards and Elgin 2002; Elgin and Grewal 2003) and H3K9 and H3K27 methylation (mono-, di- or tri-methylation) are conserved hallmarks of heterochromatin.

Two modes silencing are involved in heterochromatin formation; the first involves H3K9 methylation and often coincides with the presence of heterochromatin protein 1, HP1 (James and Elgin 1986; Eissenberg and Elgin 2000; Pal-Bhadra, Leibovitch et al. 2004; Wang, Fischle et al. 2004). Heterochromatin associated with H3K9 methylation and HP1 often involves the RNAi machinery and is associated with regions that are depleted in protein-coding genes, but enriched in repetitive sequences; examples include transposable elements, pericentromeric regions and telomeres

(Richards and Elgin 2002). DNA methylation in plants and mammals is typically associated with methylated H3K9-associated heterochromatin as well (Ooi, O'Donnell et al. 2009; He, Chen et al. 2011).

The second mode of silencing, not involving HP1, involves Polycomb group (PcG) proteins (Schwartz and Pirrotta 2007), H3K27 methylation and long non-coding RNA scaffolds (Beisel and Paro 2011). Plant ATXR5/ATXR6 proteins have recently been implicated in H3K27 methylation, and they are involved in silencing heterochromatic regions without any dependence on either DNA methylation or H3K9 methylation (Jacob, Feng et al. 2009; Jacob, Stroud et al. 2010).

Chromatin-modifying enzymes can be divided conceptually into readers, writers and erasers, which work together to regulate gene expression via the hypothesized "histone code" (Jenuwein and Allis 2001). For example, histone acetyltransferases (HATs) transfer acetyl groups to lysine tails of histones, thereby "writing" a code corresponding to a euchromatic state. This state is "read" by bromodomain proteins that recognize acetylated lysine tails in histones. However, the euchromatic state can be "erased" by histone deacetylases (HDACs), which remove acetyl groups from histones.

The reversible nature of histone modifications is an important feature of chromatin dynamics, involving numerous enzymes that perform specific histone modifications to modulate chromain states (Agger, Christensen et al. 2008; Cloos, Christensen et al. 2008; Atanassov, Koutelou et al. 2010). Addition and removal of these epigenetic marks is especially important in the germline where gametes must acquire appropriate epigenetic marks to ensure proper development of the zygote (Feng, Cokus et al. 2010; Feng, Jacobsen et al. 2010; He, Chen et al. 2011). After fertilization, a condition

which is quite prevalent in undifferentiated cells is chromatin bivalency, where both active and repressive modifications occur simultaneously (Bernstein, Mikkelsen et al. 2006). Chromatin is truly a crucial component of eukaryotic cells, for packaging large eukaryotic genomes and relaying important epigenetic information while doing so, as an extension of the DNA sequence.

Nucleolar dominance

Heterochromatin formation is central to the epigenetic phenomenon known as nucleolar dominance, in which one of the two sets of ribosomal RNA genes in a hybrid is selectively expressed (Pikaard 2000). Nucleolar dominance is widespread, and has been shown to occur in hybrids of plants, flies, mammals, amphibians and invertebrates (Pikaard 1999; Preuss and Pikaard 2007; Tucker, Vitins et al. 2010). The study of nucleolar dominance may inform our understanding of interspecies reproductive barriers in plants where connections to small RNAs are just beginning to emerge (Ha, Lu et al. 2009; Martienssen 2010). The selective nature of nucleolar dominance, whereby the ribosomal RNA (rRNA) genes from one species are always dominantly expressed over the other is in contrast to the random nature of X-chromosome inactivation in somatic cells of female mammals (Pikaard 2000).

The 40-45S large ribosomal RNA genes are essential genes that arranged in tandem arrays of hundreds to thousands of copies (Moss and Stefanovsky 2002). The transcription of 45S rRNA genes by Pol I forms the nucleolus, a distinct nuclear compartment which is the site of ribosomal assembly, synthesis and maturation (Grummt 2003). Hence, the term nucleolus organizing region (NOR), coined by Barbara

McClintock, describes the chromosomal locations where rRNA genes are tandemly arrayed (McClintock 1934; Dimario 2004).

Arabidopsis suecica is an allotetraploid hybrid of *Arabidopsis arenosa* and *Arabidopsis thaliana* that exhibits nucleolar dominance. In *A. suecica*, the *A. arenosa*-derived rRNA genes are expressed while the *A. thaliana*-derived rRNA genes are silenced (Preuss and Pikaard 2007). Polymorphisms in rRNA genes of the two species may provide a way to distinguish the NORs in *A. suecica* but other unknown loci cannot be ruled out (Lewis and Pikaard 2001; Pikaard, Preuss et al. 2005). The establishment of nucleolar dominance is developmentally regulated. NORs derived from *A. arenosa* and *A. thaliana* are both expressed during germination, but *A. thaliana*-derived NORs are subjected to selective silencing as the seedling matures (Pontes, Lawrence et al. 2007). Repressive chromatin states are important for nucleolar dominance because treatment of *A. suecica* with histone deacetylase inhibitors or DNA methyltransferase inhibitors disrupts the silencing of the *A. thaliana*-derived genes that are usually silent in the hybrid (Chen and Pikaard 1997). Studies of the epigenetic regulation of Pol I in vivo, as afforded by the *A. suecica* system, has revealed a genetic basis for roles of histone deacetylation and DNA methylation in plant rRNA gene regulation (McStay 2006; McStay and Grummt 2008).

The switch between on and off states for rRNA genes in *A. suecica* requires the histone deacetylase, HDA6 (Lawrence and Pikaard 2004; Earley, Lawrence et al. 2006). When *HDA6* is knocked down in the *A. suecica*, the association of *A. thaliana*-derived genes with heterochromatic histone modifications are lost and these genes are associated instead with active histone modifications typical of active genes, along with DNA

hypomethylation (Lawrence, Earley et al. 2004; Probst, Fagard et al. 2004). The enzyme that is involved in directing DNA methylation on *A. thaliana*-derived NORs in *A. suecica* in order to establish nucleolar dominance is DRM2, the de novo DNA methyltransferase that is involved in the heterochromatic 24nt siRNA pathway (Preuss, Costa-Nunes et al. 2008). The methylcytosine binding proteins MBD6 and MBD10 are also implicated in enforcing the repressive effects of DNA methylation (Preuss, Costa-Nunes et al. 2008). The evidence that siRNAs can target rRNA genes for DNA methylation comes from the knockdown of DCL3 and RDR2 (the Dicer and RNA dependent RNA polymerases in the 24nt siRNA pathway), which causes the loss of nucleolar dominance as well as reduced levels of 24nt siRNAs from *A. thaliana*-derived 45S rRNAs and reduced DNA methylation (Finigan and Martienssen 2008; Preuss, Costa-Nunes et al. 2008). This provides a connection between RNA directed DNA methylation and rRNA gene silencing although Pol IV and V have not been demonstrated to be involved (Tucker, Vitins et al. 2010). New findings include the involvement of SUVH5 and SUVH6 genes in establishing nucleolar dominance; these are H3K9 methyltransferases (Pontvianne and Pikaard, unpublished; Pontvianne, Blevins et al. 2010; Rajakumara, Law et al. 2011). Currently, evidence is still lacking for genes that might be involved in activating rRNA genes that have been silenced, but candidates include DNA glycosylases and associated activities involved in active DNA demethylation (Appendix A).

Basis of RNAi

Three families of proteins are important for generating and binding small RNAs. The first is a family of RNAseIII nucleases that include Dicers and Drosha (Bernstein,

Caudy et al. 2001; Lee, Nakahara et al. 2004; Pham, Pellino et al. 2004). In the next category are dsRNA binding proteins such Loquacious, R2D2, RDE-4 and DRBs which partner with Dicer enzymes to cleave long dsRNAs into 20-30nt fragments (Tabara, Yigit et al. 2002; Denli, Tops et al. 2004; Forstemann, Tomari et al. 2005; Hiraguri, Itoh et al. 2005). Last but not least, is the Argonaute (AGO) family of proteins that contain PAZ and Piwi domains which bind small RNAs and direct cleavage of target RNAs respectively (Morel, Godon et al. 2002; Liu, Carmell et al. 2004; Baumberger and Baulcombe 2005). Small RNA pathways in eukaryotes can be quite extensive, and many pathways have been described since RNAi was first described over ten years ago (Chapman and Carrington 2007; Hutvagner and Simard 2008; Ghildiyal and Zamore 2009; Czech and Hannon 2011).

Triggering RNAi

When RNAi was first discovered, relatively few dsRNA molecules were found to be capable of inducing RNAi in *C. elegans*, suggesting that there is an amplification mechanism at work (Fire, Xu et al. 1998). Several years prior, plant biologists had observed that transgenes inserted in the genome in complex repeated or inverted arrays could induce gene silencing in what was called co-suppression (Jorgensen 1990; Napoli, Lemieux et al. 1990; Jorgensen, Cluster et al. 1996). Similar findings were also reported in fungi, flies, worms and mammals (Fire, Albertson et al. 1991; Romano and Macino 1992; Pal-Bhadra, Bhadra et al. 1997; Dernburg, Zalevsky et al. 2000). Subsequently, an RNA-dependent RNA polymerase (RDR) mechanism was suggested to amplify abberant RNAs (Dalmay, Hamilton et al. 2000) and RDR activity had been experimentally verified

for an RDR protein in tomato (Schiebel, Pelissier et al. 1998). Similar RDR proteins were identified in genetic screens in worms, plants and fungi (Cogoni and Macino 1999; Dalmay, Hamilton et al. 2000; Mourrain, Beclin et al. 2000; Smardon, Spoerke et al. 2000) but homologs have not been identified in flies or mammals.

RNA produced by DNA-dependent RNA polymerases has been implicated as a source of dsRNA upon transcription of inverted repeat loci, resulting in both sense and sense anti-sense transcripts. In fission yeast, Pol II transcrips are acted upon by the RDC complex (which contains the RNA-dependent RNA polymerase Rdp1), generating dsRNAs that are diced to generate siRNAs involved in silencing centromeric repeats (Grewal and Elgin 2007; Grewal 2010). There are six RDR proteins in *Arabidopsis*, two of which are well characterized and important for amplification steps in RNA silencing pathways involved in transcriptional or post-transcriptional silencing (Wassenegger and Krczal 2006).

Slicing and dicing

siRNA involvement in RNAi was discovered in plants where complementary small RNAs were observed to correlate with post-transcriptional silencing (Hamilton and Baulcombe 1999). Biogenesis of siRNAs was then shown to be due to the action of Dicer endonucleases which are found in most model systems other than budding yeast (Elbashir, Harborth et al. 2001; Elbashir, Lendeckel et al. 2001; Elbashir, Martinez et al. 2001). RNase III-like Dicer activity was first purified from fly lysates and shown to be effective for RNAi in vitro (Bernstein, Caudy et al. 2001; Hammond, Caudy et al. 2001). A mechanistic link between microRNAs and siRNAs began to emerge when researchers

realized that both are generated by Dicer proteins (Grishok, Pasquinelli et al. 2001; Hutvagner, McLachlan et al. 2001).

Argonaute proteins had already been implicated in development in worms and plants, but the purification of RNA-induced silencing complexes (RISC) provided clear evidence that Argonaute proteins are key components as the engines of small RNA delivery in RNAi (Rivas, Tolia et al. 2005) (Jacobsen, Running et al. 1999; Tabara, Sarkissian et al. 1999). Most organisms other than fission yeast make use of multiple Argonaute family proteins, which include AGOs from the PIWI clade (which are missing in plants) (Chapman and Carrington 2007; Ghildiyal and Zamore 2009; Czech and Hannon 2011).

Arabidopsis encodes four Dicer (DCL) proteins and ten Argonaute (AGO) proteins (Liu, Feng et al. 2009; Mallory and Vaucheret 2010). The DCL proteins have a rather complicated relationship but can be distinguished by their size classes and precursor processing (Blevins, Rajeswaran et al. 2006; Margis, Fusaro et al. 2006; Chapman and Carrington 2007). DCL1 is required for the accumulation of 21nt miRNAs (Schauer, Jacobsen et al. 2002). DCL2, DCL3 and DCL4 overlap in their functions to some extent, but DCL3 produces 24nt siRNAs and is the primary Dicer protein that cleaves the products produced by the combined actions of Pol IV and RDR2 (Kasschau, Fahlgren et al. 2007). DCL2 products are 22nt and DCL4 generates 21nt tasiRNAs and 21nt siRNAs from inverted repeat loci (Xie, Allen et al. 2005). All four DCL proteins are involved during defense against viral attacks and act preferentially on different dsRNA substrates in a hierarchical manner (Blevins, Rajeswaran et al. 2006; Deleris, Gallego-Bartolome et al. 2006).

Of the 10 *Arabidopsis* AGO proteins, only five have been extensively studied; AGO1, AGO4, AGO6, AGO7 and AGO9 (Vaucheret 2008). The AGO proteins bind the small, predominantly 21, 22 or 24nt siRNA pools, depending on 5' end sequence as well as sequence siRNA length (Mi, Cai et al. 2008). AGO1 mostly associates with miRNAs (Bohmert, Camus et al. 1998; Baumberger and Baulcombe 2005), AGO7 with tasiRNAs (Adenot, Elmayan et al. 2006) and AGO4/AGO6/AGO9 with 24nt siRNAs (Zilberman, Cao et al. 2003; Zilberman, Cao et al. 2004; Havecker, Wallbridge et al. 2010). The AGO/siRNA RISC are then targeted to the nucleic acid sequences complementary to the associated siRNAs or to interacting proteins with GW repeats, via the AGO PIWI domain. Subsequent transcriptional or post-transcriptional silencing is brought about by these interactions.

Pol IV and V specific pathways

Mutants in catalytic subunits of Pol IV and Pol V were first studied as a result of forward and reverse genetic screens (Herr, Jensen et al. 2005; Kanno, Huettel et al. 2005; Onodera, Haag et al. 2005). Forward genetic screens were designed to find mutants that cause the de-repression of a silenced transgene. One of the mutants from one of the screens, *sde4* (silencing defect 4) was later identified as *NRPD1* (at the time named *NRPD1a*), the largest subunit of Pol IV (Herr, Jensen et al. 2005). Two other mutants identified in a different laboratory, *drd2* and *drd3*, turned out to be *NRPD2a* (the gene encoding the 2nd largest subunit of Pol IV and V) and *NRPE1* (*NRPD1b* at the time), the largest subunit of Pol V (Kanno, Huettel et al. 2005). A reverse genetic approach in the Pikaard lab characterized the mutants of the largest and 2nd largest subunits of Pol IV/V

and revealed their partnership (Onodera, Haag et al. 2005). Production of 24nt siRNAs is wholly dependent on Pol IV activity. Deep sequencing of small RNA libraries generated from the *nrpd1* and *nrpe1* mutants confirmed that about 94% of *Arabidopsis* siRNAs are dependent on Pol IV, but and not Pol V (Zhang, Henderson et al. 2007; Mosher, Schwach et al. 2008).

The forward genetic screens that identified Pol IV and V catalytic subunits were based on the de-repression of silenced transgenes, illustrating the role of Pol IV and V in silencing exogenous DNA sequences. It was already known that a pathway exists for small RNA-mediated methylation of complementary DNA in *Arabidopsis* requiring de novo DNA methylation via DRM2, thus Pol IV and V were found to be components of this pathway (Zilberman, Cao et al. 2004). Plants methylate cytosines in symmetric CG and CHG (where H stands for A, C or T) contexts as well as in asymmetric CHH contexts. DRM2 is able to initiate methylation in all of these sequence contexts but is the only enzyme implicated in methylating CHH sites (Cao and Jacobsen 2002; Cao and Jacobsen 2002; Cao, Aufsatz et al. 2003). The loss of 24nt siRNA from the *NRPD1* and *NRPD2* mutants along, with other mutants of this RNA directed DNA (RdDM) methylation pathway, causes the loss of CHH methylation.

RNA directed DNA Methylation (RdDM)

The RNA directed DNA methylation (RdDM) pathway is thought to be triggered by Pol IV. The Pol IV transcripts are amplified by the RNA dependent RNA polymerase, RDR2, in a process which might be a coupled with Pol IV transcription because RDR2 co-immunoprecipitates with Pol IV (Haag and Pikaard, unpublished), with the

cooperation of a putative chromatin remodeler CLASSY1 (Figure 1a) (Smith, Pontes et al. 2007). The dsRNA products of RDR2 are diced by DCL3 to generate 24nt siRNAs in a Cajal body-like siRNA/miRNA processing center in the nucleolus (Li, Pontes et al. 2006; Pontes, Li et al. 2006). HEN1 methylation of the 24nt siRNAs at the 2'-OH positions of their 3' end stabilizes the 24nt siRNAs, which are loaded into AGO4, or its surrogate AGO6 and AGO9 (Havecker, Wallbridge et al. 2010).

AGO4-siRNAs complexes are then recruited to complementary genomic loci and direct DNA methylation and heterochromatin modifications at the 'effector step' of RdDM in a series of events that remain poorly understood (Figure 1b) (Li, Yang et al. 2005; Li, Pontes et al. 2006; Qi, He et al. 2006; Zheng, Zhu et al. 2007). Transcription by Pol V produces nascent RNAs to which 24nt siRNAs bind in the mechanism by which AGO4 is recruited in this pathway (Wierzbicki, Haag et al. 2008). A chromatin remodeling complex, known as DDR, which contains the SMC hinge domain protein DMS3, a SWI2/SNF2-family protein, DRD1 and a novel methylcytosine binding protein, RDM1 is required for Pol V-specific transcription. Moreover, components of the DDR complex co-purify with Pol V (Wierzbicki, Haag et al. 2008; Law, Ausin et al. 2010). An SGS3-like coiled-coil RNA helicase protein, IDN2 might mediate the AGO4 interaction with nascent Pol V transcripts (Ausin, Mockler et al. 2009). AGO4 can also interact with an WG Ago-hook containing protein, known as KTF1 or SPT5-LIKE, due to its homology to the yeast SPT5 transcription elongation factor (Bies-Etheve, Pontier et al. 2009; He, Hsu et al. 2009; He, Hsu et al. 2009; Kanno, Bucher et al. 2010). However, KTF1 does not appear to facilitate Pol V transcription (He, Hsu et al. 2009) but its RNA binding activity and AGO4 interaction may help recruit AGO4 to Pol V transcripts. An

IWR1-like transcription regulator, RDM4 (also known as DMS4), is able influence the effector steps leading to DNA methylation, but unlike the DDR complex, RDM4 is not required for Pol V transcription (He, Hsu et al. 2009; Kanno, Bucher et al. 2010). AGO4 RISC complexes associate with Pol V transcription complexes, recruit DRM2, the de novo DNA methyltransferase, to methylate complementary DNA sequences. Because RDM1 interacts with both AGO4 and DRM2, it is presumably important for this process (Wierzbicki, Ream et al. 2009; Henderson, Deleris et al. 2010).

The mode of action for DRM2 and DRM3 will be discussed in the next section of the introduction, as well as the activities of DNA glycosylases ROS1 and ROS3, a novel RNA binding protein involved in DNA demethylation along with ROS1.

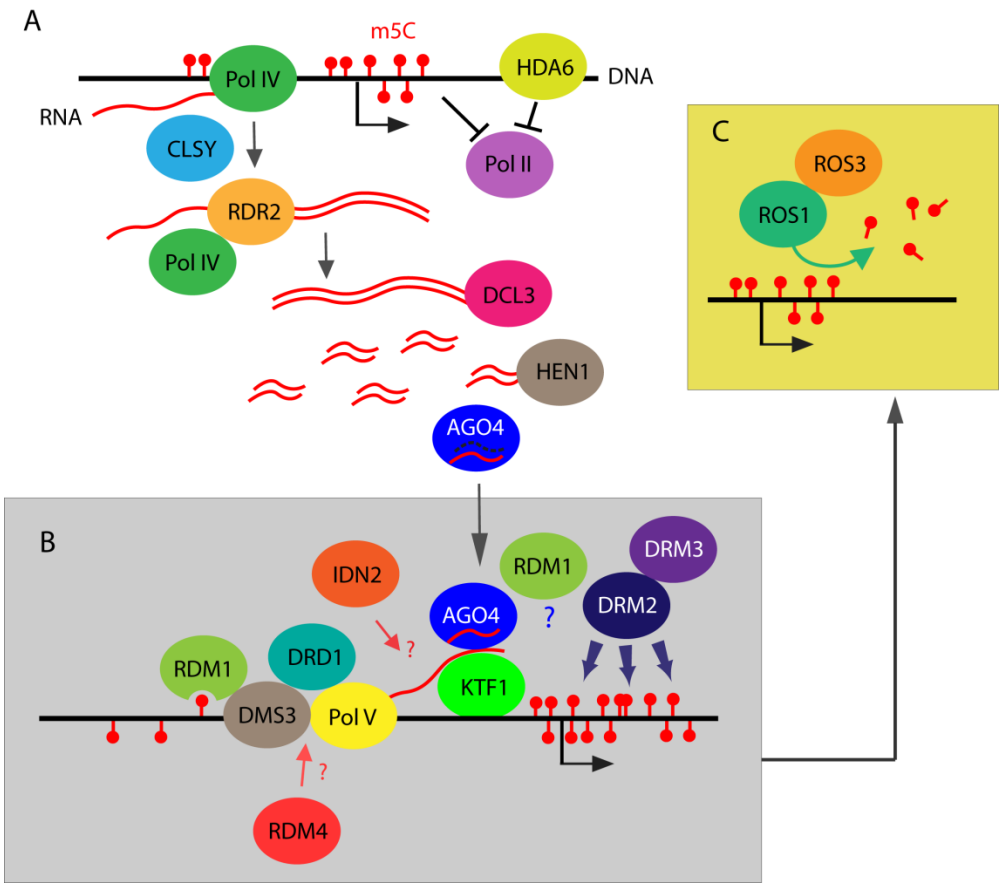


Figure 1 The RNA-directed DNA methylation and demethylation pathways

(A) Pol II is occluded from silenced promoters via HDA6-mediated histone deacetylation and DNA methylation. However, at these loci, Pol IV produces single stranded RNA for production of dsRNA by RDR2, with the help of CLSY. The dsRNAs are diced by DCL3, HEN1 methylates these siRNA duplexes, while AGO4 preferentially bind a single stranded 24nt siRNAs.

(B) Based on sequence complementarity, 24nt siRNA-containing AGO4 RISC complexes are recruited to Pol V transcript scaffolds, possibly in cooperation with KTF1 and IDN2. The DRD1, DMS3 and RDM1 (DDR) complex assists Pol V transcription, along with RDM4. DDR-assisted Pol V transcription and 24nt siRNA-associated AGO4 RISC complexes direct DNA methylation to target locus by the de novo DNA methyltransferase DRM2, which may partner with DRM3 for enhanced activity.

(C) ROS1 DNA glycosylase cuts methylated DNA, allowing its repair using unmethylated cytosines. The RNA binding protein ROS3 helps ROS1 find its sites.

Intersection between RdDM and HDA6

Mutations of *HDA6,* which encodes an RPD3-like histone deacetylase, disrupts silencing of transgene promoters and 45S rRNA genes (Aufsatz, Mette et al. 2002; Probst, Fagard et al. 2004; Earley, Lawrence et al. 2006; He, Hsu et al. 2009; Earley, Pontvianne et al. 2010). HDA6 is important for rRNA gene regulation in allotetraploid *A. suecica* plants in nucleolar dominance (discussed prior in that section), but HDA6 appears to regulate non-hybrid rRNA genes in non-hybrid *A. thaliana* as well, where it prevents spurious transcription of the 45S rRNA gene arrays by Pol II (Earley, Lawrence et al. 2006; Earley, Pontvianne et al. 2010). HDA6 is a key factor required for transcriptional gene silencing of many transposable elements, but it is not universally required to silence RdDM targets (Blevins and Pikaard, unpublished). Double mutants of *hda6* with either Pol IV and V catalytic subunits or with *drm2* exhibit pleitropic *SDC*-overexpression associated with loss of cytosine methylation in tandem arrays upstream of the *SDC*-promoter and resulting in a dwarf phenotype (Henderson and Jacobsen 2008). *SDC* encodes an F-Box protein and represents a class of genes that are regulated by siRNAs, CHG methylation, H3K9 methylation and histone deacetylation via HDA6 (Blevins and Pikaard, unpublished and (Henderson and Jacobsen 2008)).

Flowering time

Pol IV, Pol V and the RdDM pathway are involved in control of flowering time, and flowering time genes such as *FWA* and *FLC* are targets of siRNA directed silencing (Swiezewski, Crevillen et al. 2007). Mutants of Pol IV and V are significantly delayed in flowering compared to wildtype when grown under short day conditions, as are other

mutants in the RdDM pathway, including *rdr2*, *dcl3*, *ago4* and *drm2* (Chan, Zilberman et al. 2004; Onodera, Haag et al. 2005; Haag, Pontes et al. 2009). Two flowering time regulators FCA and FPA have been implicated as components of the RdDM pathway, and they could be involved in directing the establishment of silencing at genes controlling flowering time (Baurle, Smith et al. 2007).

Abiotic and biotic stress response

Stress inducible siRNA production is an important aspect of plant biology. Two classes of stress-related responses in plants are the production of abiotic and the biotic stress-inducible siRNAs, which are derived from convergently transcribed gene pairs. Abiotic stress siRNAs are often referred to as natural antisense siRNAs (natsiRNAs) (Xie and Qi 2008). Pol IV components are required for the biogenesis of these biotic and abiotic natsiRNAs but components of the miRNA pathway are involved as well (Borsani, Zhu et al. 2005) (Katiyar-Agarwal, Morgan et al. 2006). Pathogen-inducible siRNAs have been recently described and require both Pol IV and Pol V for the production of a novel class of 39-41nt long siRNAs (lsiRNA). These lsiRNAs cause the downregulation of a negative regulator of plant defense response gene (Katiyar-Agarwal, Gao et al. 2007), allowing for the defense response to proceed.

Spreading of silencing

Spreading of silencing signals occur in plants and involves Pol IV, but not Pol V components, reinforcing the role of Pol IV in the biogenesis of siRNAs that act as mobile signals (Brosnan, Mitter et al. 2007; Pikaard, Haag et al. 2008; Daxinger, Kanno et al.

2009). The mobilization of cell-to-cell siRNA signals is an important aspect of siRNA silencing and has been observed in plants and worms, both of which make use of RDR proteins for RNAi (Voinnet 2005). Using next generation sequencing, mobile 24nt signals have been found, and can direct RNA directed DNA across a graft junction in tissues defective for RdDM. These grafting experiments were also able to track movement of 24nt siRNAs from the shoots to the roots of mutants that are incapable of producing 24nt siRNAs (Martienssen 2010; Molnar, Melnyk et al. 2010; Molnar, Melnyk et al. 2011). 21nt and 22nt siRNAs are also capable of movement, and can direct RNA silencing (Dunoyer, Brosnan et al. 2010; Dunoyer, Schott et al. 2010).

Maternal imprinting of siRNA loci

The expression of 24nt siRNAs in the early endosperm has been found to be of maternal chromosome, and requires Pol IV (Mosher, Melnyk et al. 2009). In pollen, 21nt siRNAs are generated by the vegetative cell and silence corresponding loci in sperm cells (Slotkin, Vaughn et al. 2009). These forms of imprinting in plants involving siRNAs and that target transposable elements, could be part of a self, non-self recognition program initiated in the germline (Martienssen 2010). The biological role of these processes in plants are still not well understood but it is reminiscent of hybrid dysgenesis in flies, where piRNAs loading in fly egg cells results in sterility when the egg cells are fertilized by a paternal genome containing transposable elements that were not in the maternal genome (Brennecke, Aravin et al. 2007; Brennecke, Malone et al. 2008).

Paramutation

In maize, a role for Pol IV and Pol V has been found in paramutation, where meiotically heritable expression states can be inherited in trans, on homologous DNA sequences based on the chromatin states of the paramutation allele. The maize homolog of NRPD1 (largest subunit of Pol IV) as well as a gene encoding an NRPD2/E2-like protein (one of three putative second largest subunits of Pol IV and/or V in maize) are required for paramutation (Erhard, Stonaker et al. 2009; Sidorenko, Dorweiler et al. 2009). Other RdDM proteins required for paramutation in maize include homologs of RDR2 and DRD1 (Alleman, Sidorenko et al. 2006; McGinnis, Springer et al. 2006; Hale, Stonaker et al. 2007). Affinity purification and high resolution mass spectrometry analyses of the RNA polymerase complex captured via the MOP2/RMR7, an NRPD2/E2-like protein required for paramutation (Sidorenko, Dorweiler et al. 2009) shows that it is an NRPE2 subunit of Pol V (Haag et al, unpublished collaboration of the Chandler and Pikaard labs). The involvement Pol IV and Pol V in maize paramutation suggests that a similar Pol IV and Pol V-mediated RdDM mechanisms are at work in other plant species (Pikaard and Tucker 2009).

iv.

DNA METHYLATION LANDSCAPES

DNA methylation was recognized early in vertebrates as an important epigenetic modifier, especially after a developmentally important methyl-C binding protein, MeCP2, was discovered (Bird 1986; Ng and Bird 1999). This covalent modification on DNA is mostly limited to cytosine residues in eukaryotes (Bird 1992; Jacobsen 1999) but in bacteria and archaea, both cytosine and adenine nucleosides can be methylated, and are used extensively as a mark of self-DNA for genome defense, replication and DNA repair (Wion and Casadesus 2006). The study of eukaryotic epigenomes shows that although DNA methylation patterns can be variable, the correlation between promoter DNA methylation and gene silencing generally holds true (Suzuki and Bird 2008). However, gene body hypermethylation has been linked to gene activity in animals and plants, showing that DNA methylation may not just be associated with silencing (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Zemach, McDaniel et al. 2010). Some eukaryotes, such as the yeast and *C. elegans* have lost DNA methylation and the associated methyltransferase enzymes altogether, but the presence of DNA methylation in a wide range of invertebrates, vertebrates and plant species suggests that the loss of DNA methylation machineries may have been a recent evolutionary adaptation (Suzuki and Bird 2008). Plants utilize DNA methylation ubiquitously and share the mammalian DNA methylation machinery plus additional plant-specific enzymes (Meyer 2010). The discovery of the siRNA directed de novo DNA methylation pathway and the link between DNA methylation and RNA polymerases IV and V in plants illustrates the

importance of DNA methylation in plant biology (Pikaard, Haag et al. 2008; Pikaard and Tucker 2009).

There has been an impetus in mammalian DNA methylation research to elucidate the role of hydroxymethyl-cytosine (hmC), which is just beginning to emerge as a major epigenetic modification in pluripotent stem cells and is used extensively in early development (Ficz, Branco et al. 2011; Williams, Christensen et al. 2011; Wossidlo, Nakamura et al. 2011; Wu, D'Alessio et al. 2011). The TET family of proteins are responsible for the oxidation of methylcytosine to produce hydroxymethyl-cytosine, which is more stable (Jin, Kadam et al. 2010). The hmC residue has been found to be involved in activating genes as well as repressing developmentally regulated genes (Ficz, Branco et al. 2011). Very recently, the hmC modification has also been found to be a target for active DNA demethylation (Guo, Su et al. 2011). The conversion of m5C to hm5C shows that epigenetic regulation of DNA is far more complex than initially thought. It remains to be shown if hmC modifications are involved in epigenetic regulation in other organisms.

DNA methyltransferases

DNA methyltransferases are conserved proteins found in all three kingdoms (Bestor and Verdine 1994). The catalytic domains of DNA methyltransferases are conserved from bacteria to eukaryotes (Goll and Bestor 2005), as are the mechanisms involved in the catalysis. DNA methyltransferase biochemistry revolves around an invariant cysteine amino acid that is found within a hyperconserved PCG motif of these enzymes (Bestor and Verdine 1994). DNA methylation catalysis also involves the

dramatic eversion (flipping out) of the target cytosine from the DNA strand, first observed in the crystal structure of the *Hha*I methyltransferase (Klimasauskas, Kumar et al. 1994; Horton, Ratner et al. 2004).

Two main classes of DNA methyltransferase (DNMT) exist in eukaryotes: maintenance and de novo methyltransferases (He, Chen et al. 2011). The semiconservative mode of DNA replication is utilized by maintenance DNMTs to target symmetric CGs that are hemimethylated, via association with m5C-binding proteins VIM1 (in plants) and UHRF (in animals) (Bostick, Kim et al. 2007; Woo, Dittmer et al. 2008). The DNMT1 maintenance DNA methyltransferase has been found to interact with the G9a H3K9 histone methyltransferase, and is loaded along with PCNA during replication in human cell-lines (Esteve, Chin et al. 2006). The second class of de novo methyltransferase is represented by the Dnmt3-family in animals, but a plant specific Chromo-methylase, CMT3 also possesses de novo methyltransferase activity, in addition to DRM2. De novo methylation is an important aspect of genome defense in plants and animals as it allows re-establishment of silencing that can be lost during replication. In plants, de novo methylation is required for the of non-CG methylation (Feng, Cokus et al. 2010).

Accessory proteins in the MBD family that bind methylated DNA also play important roles in DNA methylation dynamics (Jorgensen and Bird 2002). MeCP2 was one of the earliest methylcytosine binding protein isolated. MeCP2 plays important roles in transcriptional repression and mutations in MeCP2 can lead to neurological Rett syndrome (Nan, Cross et al. 1998). Methylcytosine binding proteins MBD6 and MBD10 in *Arabidopsis* has been implicated in recognizing DRM2 methylation patterns in order to

selectively silence a specific class of rRNA genes in nucleolar dominance (Preuss, Costa-Nunes et al. 2008; Costa-Nunes, Pontes et al. 2010). The RNA directed DNA methylation pathway also makes use of a novel methylcytosine binding protein called RDM1 in a Pol V dependent fashion (Gao, Liu et al. 2010; Law, Ausin et al. 2010). Methylcytosine binding proteins are able to act as adapters for m5C marks and guide other chromatin modifying enzymes that do not typically have specificities toward m5C.

Maintenance methylation

The mammalian maintenance methyltransferase DNMT1 is an essential gene whose absence causes defects in imprinting and differentiation (Howell, Bestor et al. 2001; Grohmann, Spada et al. 2005; Damelin and Bestor 2007; Ooi and Bestor 2008). The recent crystal structure of DNMT1 in complex with hemimethylated DNA illustrates beautifully the manner in which this enzyme is able to direct selective methylation on hemimethylated DNA, including autoinhibitory mechanism to prevent abberant methylation of unmethylated DNA (Song, Rechkoblit et al. 2011).

The *Arabidopsis* homolog of DNMT1 is MET1 (Kankel, Ramsey et al. 2003). Transgenerational effects deduced from hypomorphic alleles of *met1* in *Arabidopsis* suggest that faithful reestablishment of CG methylation in plants is dependent on MET1 activity (Mathieu, Reinders et al. 2007). DNA methylation pattern maintenance also requires the protein DDM1, a SWI2/SNF2 chromatin remodeler whose loss of function causes pervasive cytosine methylation deficient phenotypes (Kakutani, Jeddeloh et al. 1995). The recovery of genome-wide methylation upon restoration of DDM1 activity can

occur at loci subject to RdDM, involving Pol IV and V, but loci that do not produce siRNAs are not re-methylated (Teixeira, Heredia et al. 2009).

De novo methylation

Transposable elements and DNA repeats pose a threat to the genome and are often targeted for silencing by de novo DNA methylation. The plant RNA directed DNA methylation pathway involving Pol IV and V has been described in the preceding section, so the discussion here will mainly be on the de novo DNMTs. Most animal species that encode DNMT1 also encode the de novo DNMT3 DNA methyltransferases (Goll and Bestor 2005). DNMT3a and DNMT3b represent the catalytic class of these mammalian methyltransferases whereas DNMT3L has a mutation in the active site that renders the enzyme catalytically inactive (Aapola, Kawasaki et al. 2000). However, the interaction between DNMT3L and either catalytic DNMT3 enhances the activity of the DNMT3a/b in de novo DNA methylation (Ooi, Qiu et al. 2007; Zhang, Jurkowska et al. 2010; Van Emburgh and Robertson 2011). De novo DNA methylation involving piRNAs targeting transposable elements in the germline has been reported, suggesting a similarity between animal and plant RNA directed de novo methylation pathways (Aravin, Sachidanandam et al. 2008; Saito and Siomi 2010; Siomi, Sato et al. 2011).

The plant DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) protein is the primary de novo DNA methyltransferase, and is related to the mammalian DNMT3's although the active site domains for DRM2 are arranged in a different order, as the name suggests (Ashapkin, Kutueva et al. 2002; Cao and Jacobsen 2002). The active site of DRM2 is crucial for its activity (Appendix A and (Henderson, Deleris et al.

2010; Naumann, Daxinger et al. 2011)). The plant specific CHROMOMETHYLASE3, or CMT3 is also a de novo cytosine methyltransferase, but this enzyme works in tandem with a H3K9 methyltransferase KRYPTONITE (KYP or SUVH4) in a self-reinforcing mechanism by which CHG methylation and H3K9me reinforce one another in plants (Lindroth, Cao et al. 2001; Cao and Jacobsen 2002; Jackson, Lindroth et al. 2002; Cao, Aufsatz et al. 2003). The *DRM3* gene encodes a catalytically inactive DRM-like protein in *Arabidopsis*, similar to DNMT3L mammals. A mechanistic connection between DRM2 and DRM3 has recently been demonstrated, illustrating the similarities between mammalian DNMT3a/DNMT3L and plant DRM2/DRM3 in de novo DNA methylation (Henderson, Deleris et al. 2010).

Reversing DNA methylation

Active DNA demethylation is involved in plant and animal development (Ooi and Bestor 2008; Wu and Zhang 2010). Plants utilize a conserved family of bifunctional DNA glycosylase/lyase enzymes which cleave at methylated cytosines, triggering base-excision repair, allowing the incorporation of unmethylated nucleosides at the site of cleavage (Gehring, Reik et al. 2009). The active DNA demethylation mechanism is less established in animals, but DNA demethylation is thought to be carried out by base-excision repair machinery initated by the AID/Apobec deaminases that convert methylated cytosines to thymines, triggering the G/T mismatch repair pathway (Morgan, Dean et al. 2004; Boland and Christman 2008).

The best characterized DNA demethylation pathway is the one carried out by the plant DEMETER (DME) glycosylase, an essential gene that is required for maternal

imprinting in the central cell of the ovule (Choi, Gehring et al. 2002; Kinoshita, Miura et al. 2004). Many targets for DME have been identified, including polycomb genes such as *FWA*, *MEDEA* and *FIS2* (Gehring, Huh et al. 2006; Takeda and Paszkowski 2006; Bauer and Fischer 2011). The ground state of these imprinted genes are "off"/methylated and DME demethylation in the maternal germline causes maternal expression of these genes in the female gametophyte and endosperm prior to fertilization (Kinoshita, Miura et al. 2004; Huh, Bauer et al. 2008). The tissue specific expression of DME is unique. Other DME-Like (DML) glycosylases have been identified in plants, including ROS1, DML2 and DML3 (Gong, Morales-Ruiz et al. 2002). Active DNA demethylation is thought to play a crucial role in maintaining the right balance of DNA methylation densities in *Arabidopsis* (Kapoor, Agius et al. 2005; Agius, Kapoor et al. 2006; Morales-Ruiz, Ortega-Galisteo et al. 2006). Analysis of the *ros1 dml2 dml3* triple mutant methylome reveals their role in pruning DNA methylation throughout the *Arabidopsis* epigenome (Penterman, Uzawa et al. 2007; Penterman, Zilberman et al. 2007). All four DME family proteins have in vitro activity towards methylated DNA templates (Morales-Ruiz, Ortega-Galisteo et al. 2006; Ortega-Galisteo, Morales-Ruiz et al. 2008; Ponferrada-Marin, Martinez-Macias et al. 2010).

ROS3, a novel protein with RNA recognition motifs (Zheng, Pontes et al. 2008) was identified to be a DNA demethylation factor in the same screen that identified the DNA glycosylase ROS1. In *ros3* mutants, ROS1 localization and gene expression levels are reduced, suggesting that ROS3 may function upstream of ROS1, whereas ROS3 localization was not affected in the *ros1* mutant (Zheng, Pontes et al. 2008). This result

lends some support to a model whereby ROS3 targets ROS1 to sites matching RNA sequences bound by ROS3 (Figure 1c)(Chinnusamy and Zhu 2009; Zhu 2009).

Within the DNA methylation field, there are still many mysteries to decipher from the distinct roles of DNA methylation in gene silencing and gene activity, to the role of hydroxymethylated cytosine modifications. Current bisulphite sequencing technologies do not discriminated between methylated versus hydroxymethylated cytosines and MBD proteins do not bind hmC modifications on DNA (Ndlovu, Denis et al. 2011). New methodologies will need to be developed based on antibodies that recognize mCs versus hmCs. The increasing evidence for active DNA demethylation means that DNA methylomes are not static, but is amenable to the needs of the organism (Figure 1 and (Zhu 2009; Chen and Riggs 2011)). As such, current methylome mapping reflects only a steady-state of methylation and demethylation.

v.

RNA POLYMERASE IV AND V

Transcription of DNA into RNA is facilitated by evolutionarily conserved multisubunit DNA dependent RNA polymerases (RNAPs). The crystal structure of yeast Pol II has been solved and has provided a basis for the universal conservation of RNAPs between bacteria and prokaryotes. Coupled with the elucidation of *Arabidopsis* Pol IV and V subunit compositions, a spate of structural and genetic studies from archaea has bridged our understanding of RNAPs.

RNA polymerases in the three kingdoms of life

Prokaryotes encode the simplest multisubunit RNA polymerase composed of five core subunits (Yura and Ishihama 1979). Archaea on the other hand, has subunit compositions that are more akin to eukaryotic Pol II – consisting of ten to eleven subunits (Kaine, Mehr et al. 1994; Wang, Jones et al. 1998). Eukaryotic RNA polymerase II has a twelve subunit structure, while Pol I has fourteen and Pol III has seventeen (Archambault and Friesen 1993).

The functional diversification of eukaryotic RNA polymerases I, II and III is limited to the eukaryotic kingdom, as prokaryotes and archaea utilize one RNAP. Pol I is a specialist in transcribing large 40-45S ribosomal RNA (rRNA) precursor genes from a cytologically relevant chromosome structure known as the NOR (nucleolus organizer region) and the transcriptional activity of Pol I is coincident with nucleolus formation (Moss and Stefanovsky 2002; Grummt 2003). Pol II's main functions include the

transcription of messenger RNAs (mRNAs), some small nuclear RNAs (snRNAs) and microRNAs (miRNAs) (Woychik and Hampsey 2002). The 5S rRNAs, tRNAs and some snRNAs are transcribed by Pol III (Henry, Ford et al. 1998; Hernandez 2001). Plants are unique in encoding Pol II-related atypical RNA polymerases that are exclusively used for RNA silencing (Onodera, Haag et al. 2005). Molecular genetic analyses Pol IV and V catalytic mutants have revealed most of what we know about the functions of Pol IV and V but the functions of the remaining ten non-catalytic subunits is less clear, and these non-catalytic subunits are the focus of my thesis.

Structural conservation

Prokaryotic multisubunit RNA polymerases are the simplest RNAPs, as exemplified by the crystal structure of *Thermus aquaticus* polymerase, which is composed of five core subunits; the β', β, ω and two α subunits (Zhang, Campbell et al. 1999). The largest and second largest subunits, β' and β respectively, interact to form the catalytic center, the α homodimer playing roles in assembly, along with the ω subunit (Minakhin, Nechaev et al. 2001). The resulting crabclaw structure is preserved in all RNA polymerases. Within the crabclaw structure is the site where duplex DNA is projected into the catalytic center of the polymerase along the floor of the polymerase (Cramer, Armache et al. 2008; Werner 2008). The DNA duplex is bound by a DNA binding channel and the template DNA eventually encounters the wall of the polymerase, where it is deflected at right angle to the exit channel (Gnatt, Cramer et al. 2001; Gnatt 2002). The elongating RNA/DNA hybrid is perpendicularly oriented relative to downstream duplex DNA and secured by the clamp, another conserved structural element

(Werner and Grohmann 2011). Bacterial polymerases lack the polymerase stalk, which in Archaea and eukaryotes are formed by two subunits interact extensively with the extruded RNA chain (Hirtreiter, Grohmann et al. 2010).

The subunits of RNA polymerases from bacteria, archaea and eukaryotes are presented in Figure 2. The equivalent subunits are arranged in order of the bacterial homologs. Also note the re-naming of archaeal subunits (RPO stands for RNA Polymerase) to reflect a unified numbering system based on yeast subunits (Werner 2007; Werner 2008; Werner and Grohmann 2011). Archaeal RPO1 and RPO2 subunits are sometimes encoded by two genes in archaea but are unified here for clarity (Werner and Grohmann 2011). The A, B, C, D and E designation proceeding the "RP" (RNA Polymerase) or "NRP" (Nuclear RNA Polymerase) for plants reflects Pol I, II, III, IV and V respectively (Ream, Haag et al. 2009). Subunits for Pol I and III are omitted, but the *Arabidopsis* subunit compositions for these polymerases have also been elucidated and are found to have subunits homologous to those in the yeast (Ream and Pikaard, unpublished and (Werner, Thuriaux et al. 2009)).

Bacteria	Archaea	Eukaryotes			
		Sc RNAPII	At RNAPII	At RNAPIV	At RNAPV
β'	RPO1	RPB1	NRPB1	NRPD1	NRPE1
β	RPO2	RPB2	NRPB2	NRPD2	NRPD2
α	RPO3	RPB3	NRPB3a/NRPB3b	NRPB3a/NRPB3b	NRPB3a/NRPB3b
α	RPO11	RPB11	NRPB11	NRPB11	NRPB11
ω	RPO6	RPB6	NRPB6a/NRPB6b	NRPB6a/NRPB6b	NRPB6a/NRPB6b
	RPO5	RPB5	NRPB5	NRPB5	NRPE5
	RPO8*	RPB8	NRPB8a/NRPB8b	NRPB8b	NRPB8a/NRPB8b
	RPO10	ROB10	NRPB10	NRPB10	NRPB10
	RPO12	RPB12	NRPB12	NRPB12	NRPB12
	RPO4	RPB4	NRPB4	NRPD4	NRPD4
	RPO7	RPB7	NRPB7	NRPD7/NRPE7	NRPE7
		RPB9	NRPB9a/NRPB9b	NRPB9b	NRPB9a/NRPB9b
	RPO13*				

Figure 2 Subunit compositions of DNA-dependent RNA polymerases

The table shows the subunits of bacterial, archaeal RNA polymerase (RNAP) subunits and the homologous yeast Pol II (Sc RNAPII) and *Arabidopsis* Pol II, IV and V (At RNAPII, At RNAPIV, At RNAPV) subunits. At subunits encoded by the same genes are only named once for clarity. *RPO8 and RPO13 are found in some, but not all archaeal species.

Although the crystal structure of an archaeal polymerase was solve after the yeast Pol II structure, the degree of similarity was already apparent between archaeal polymerases and Pol II from the sequence alignments of DNA dependent RNA polymerases (Kaine, Mehr et al. 1994). The functional classes for RNA polymerase core subunits are conserved in archaea, bacteria and eukaryotes. The β' and β subunits (largest and second largest subunits) that form the Mg^{2+}-binding active site, bridge and trigger helix, DNA and DNA/RNA binding sites, secondary NTP entry site and loop/switch region are conserved in terms of structure and multiple sequence homologies. The Pol IV and V catalytic subunits contain all the conserved features and require the Metal A and B

Mg^{2+}-binding sites if the active site for RNA directed DNA methylation (RdDM) (Haag, Pontes et al. 2009). The α homodimer (3rd and 11th largest subunits) assembly platform subunits that initiate subunit assembly are also conserved. In eukaryotes and archaea, these subunits associate with the 10th and 12th largest subunits as well. An auxiliary subunit that is homologous to ω; this 6th largest subunit is involved in basic promoter-associated activities.

The 4th and 7th largest subunits form a stalk of the polymerase in archaeal and eukaryotic lineages (Figure 3). RPB9 is purely an eukaryotic subunit which is encoded by unique genes for Pol I, II and III (Chapter two). The 8th largest and 13th largest subunits of archaea are only found in some species (Kwapisz, Beckouet et al. 2008).

Figure 3 Cartoon of yeast Pol II and subunits

The 12 subunits of yeast RNA polymerase II are named below their associated shape and color on the left. Duplex DNA is colored white, nascent RNA is red, and these form the depicted transcription bubble, with arrows indicating the direction of the incoming and outgoing DNA duplex.

Evolution of RNA polymerases

Based on the structural conservation and the molecular mechanisms that all RNA polymerases have in common, recent review articles has postulated the hypothesis that the last universal common ancestor (LUCA) of prokarya, archaea and eukarya had a RNA polymerase that resembles the extant bacterial RNA polymerase because it represents the most simple but fully functional complex (Grohmann and Werner 2011; Werner and Grohmann 2011). This is indeed, the most parsimonious hypothesis and Werner et al went on to suggest the ‘elongation first hypothesis’ owing to the only

universally conserved transcription factor; NusG, in bacteria and SPT5, in archaea and eukaryotes (Greenblatt, Nodwell et al. 1993; Zhou, Kuo et al. 2009; Hirtreiter, Damsma et al. 2010). The additional RNA silencing polymerases from plants, Pol IV and V, do not deviate from this hypothesis as they are descendants from the plant Pol II enzyme (Luo and Hall 2007; Tucker, Reece et al. 2011). Interestingly, an SPT5-like protein named KTF1 with WG AGO-hook motifs is important for RdDM (refer to preceeding section on RdDM) (Bies-Etheve, Pontier et al. 2009; He, Hsu et al. 2009; Wang and Dennis 2009). Although this elongation factor does not appear to be associated with Pol IV, association of an SPT5-like protein with Pol V is intriguing in light of this elongation first hypothesis.

Beyond RNA polymerase I, II and III

A paradox of transcriptional silencing via small RNAs is the generation of small RNAs from these condensed, supposedly transcriptionally silent domains (Grewal and Elgin 2007). The best studied pathway is currently from fission yeast, where Pol II is important for the maintenance of the centromeric repeats, including the formation of the precursor siRNAs and scaffold transcripts to which they bind (Djupedal, Portoso et al. 2005; Schramke, Sheedy et al. 2005). There is evidence that microRNA biogenesis evolved along with Pol II transcription and the splicing machinery (miRtrons, for instance) (Kim, Han et al. 2009; Allen and Howell 2010). The discovery of Pol IV and V as dedicated RNA polymerases for silencing transgenes, repetitive DNA and transposable elements via the 24nt siRNA directed DNA methylation pathway suggests that RNA

polymerases are able to transcribe silent DNA, creating a feedback loop that reinforces the effects of transcriptional gene silencing (Beisel and Paro 2011).

In 2005, three papers announced the functions of Pol IV and V (Pol IVa and IVb as they were known then) as atypical polymerases involved in RNA silencing (Herr, Jensen et al. 2005; Kanno, Huettel et al. 2005; Onodera, Haag et al. 2005). It wasn't until 2009 that the full composition of Pol IV and V were elucidated (Ream, Haag et al. 2009). Other clues for a connection between Pol V and Pol II were established by the purification of the Pol IVb complex from cauliflower (Huang, Jones et al. 2009), identification of the Pol IV and V specific 4th largest subunit in a genetic screen, and a report on the the Pol V specific 5th largest subunit (He, Hsu et al. 2009; Lahmy, Pontier et al. 2009). However, to appreciate the different subunit compositions and the evolutionary history between Pol II, IV and V, the full subunit compositions is required (Ream, Haag et al. 2009).

Insights from subunit composition

From an evolutionary standpoint, the emergence of a Pol IV-like largest subunit is thought to have occurred in the last common ancestor of land plants and the Charales, a complex group of algae known as stoneworts (Luo and Hall 2007). The largest subunit, NRPD1 can be detected in Charales, but not the 2nd largest subunit NRPD2 (shared by Pol IV and V in *Arabidopsis*). This suggests that this Pol IV subunit in Charales likely pair up with the Pol II 2nd largest subunit NRPB2 and other Pol II subunits. The NRPD1 subunit was duplicated from the NRPB1 gene, as shown by multiple shared intron/exon boundaries (Luo and Hall 2007). The NRPE1 subunit arose by duplication of the NRPD1

gene in angiosperms, and the substitution rates between NRPE1 and NRPD1 suggests that they are still evolving rapidly (Luo and Hall 2007), with distinct subunit compositions and functions in RdDM as well as other related small RNA pathways in *Arabidopsis* (Pikaard, Haag et al. 2008; Tucker, Reece et al. 2011). In the next section, discussions on the known functions of non-catalytic subunits of yeast Pol II and early work on non-catalytic subunits of *Arabidopsis* Pol II, IV and V are presented, along with the scope of this thesis.

SCOPE OF THESIS

Ninth largest subunits

An important subunit of eukaryotic RNAP not found in archaea is the 9th largest subunit (Figure 2). In yeast, the deletion of this Pol II subunit results in a *Δrpb9* strain that is sensitive to low and high temperatures (Woychik, Lane et al. 1991). This was the first Pol II subunit mutant isolated, providing an opportunity to study RPB9 functions. Since RPB9 and RPB5 sit at the jaw domain of the RNA polymerase and are thought to interact with the incoming DNA duplex (Cramer, Armache et al. 2008), a myriad of interactions have been implicated for RPB9 including associations with general transcription factors for initiation (Hull, McKune et al. 1995; Sun, Tessmer et al. 1996; Ziegler, Khaperskyy et al. 2003) and transcript elongation (Awrey, Weilbaecher et al. 1997; Hemming, Jansma et al. 2000); RPB9 also has a role in proofreading to ensure transcriptional fidelity *in vivo* (Nesser, Peterson et al. 2006; Walmacq, Kireeva et al. 2009). However, in vitro fidelity of Pol II is more dependent on TFIIS than RPB9 (Koyama, Ito et al. 2007). The TFIIS cleavage factor is a dissociable protein that is able to bind along RNAP and extend into the active site for target RNA cleavage when the polymerase is stalled or encounters mismatches (Christie, Awrey et al. 1994). A homologous factor, TFS, is used by archaea for this function (Hausner, Lange et al. 2000). The intrinsic cleavage activity of RPB9 is weak compared to the 9th largest subunits in Pol I and III, which do not contain TFIIS-like homologs (Ruan, Lehmann et al. 2011). In all, this suggests that TFIIS arose as a cleavage factor in the common

ancestor of archaea and eukaryotes, whereas the 9th largest subunits came about independently in the eukaryotic lineage prior to Pol I, II and III diversification. The 9th subunits have homology to TFIIS, so may have evolved by duplication of TFIIS. The Pol I and III 9th largest subunits display strong TFIIS-like transcript cleavage activity without associated TFIIS. Other activities associated with RPB9 include transcription coupled nucleotide excision repair (Li, Ding et al. 2006) and RPB1 degradation in response to UV damage (Chen, Ruggiero et al. 2007).

There are two identifiable *NRPB9* subunit genes in *Arabidopsis,* NRPB9a and NRPB9b. Pol II and V can associate with either NRPB9a or NRPB9b, as shown by mass spectrometry studies, whereas Pol IV only associates with NRPB9b. Although NRPB9a and NRPB9b are 93% identical in amino acid sequence, they play discrete roles in the RNA directed DNA methylation pathway.

Chapter two will explore the various functions of the 9th largest subunits from the Pol II, IV and V polymerases in *Arabidopsis*.

Fourth largest and seventh largest subunits

Like the viability of the *Δrpb9* deletion in yeast, *Δrpb4* is also viable although a double mutant of *Δrpb9 Δrpb4* is lethal (Woychik and Young 1989; Woychik, Lane et al. 1991; Maillet, Buhler et al. 1999). The deletion of RPB7 is not tolerated in yeast (McKune, Richards et al. 1993). The RPB4/RPB7 heterodimer makes up the stalk of the polymerase adjacent to the RNA exit channel (Figure 3 and (Bushnell and Kornberg 2003)). The architecture of this substructure is also apparent in the Rpo4/Rpo7 (previously RpoE/ RpoF) stalk observed the archaea crystal structure (Hirata, Klein et al.

2008). The dissociability of this subcomplex has been shown in yeast, but archaeal and plant 4/7 subcomplexes might be more stably associated (Choder 2004; Armache, Mitterweger et al. 2005; Grohmann, Hirtreiter et al. 2009).

The purification of yeast Pol II complexes from *Δrpb4* strains showed concomitant loss of RPB7 from the complex (Edwards, Kane et al. 1991) while the overexpression of RPB7 suppresses the *Δrpb4* phenotypes, which indicates a distinct role for RPB7 (Sheffer, Varon et al. 1999). RPB6 is involved in anchoring the 4/7 subcomplex to the polymerase, in close agreement with structural studies (Gnatt, Cramer et al. 2001; Tan, Prysak et al. 2003; Armache, Mitterweger et al. 2005; Sampath, Balakrishnan et al. 2008).

Based on *Δrpb4* studies, RPB4 has been found to be involved in promoter dependent transcription (Edwards, Kane et al. 1991; Orlicky, Tran et al. 2001; Hirtreiter, Grohmann et al. 2010) involving TFIIF (Chung, Craighead et al. 2003). FRET studies have also shown that nascent RNA comes into close proximity to the 4/7 subcomplex in both yeast and archaea (Chen, Chang et al. 2009; Grohmann, Klose et al. 2010). The main RNA interaction interface is thought to be RPB7 (Meka, Werner et al. 2005; Ujvari and Luse 2006). The close association of RPB4 and RPB7 throughout the transcription cycle is also supported by their interaction with RPB1 CTD factors (Kimura, Suzuki et al. 2002; Mitsuzawa, Kanda et al. 2003). Genomewide profiling using affinity tagged-RPB4 shows no difference in occupancy of RPB4 relative to the profile of a tagged catalytic subunit of Pol II (Jasiak, Hartmann et al. 2008; Runner, Podolny et al. 2008).

A potential role for RPB4/RPB7 in the cytoplasm during stress is something quite unusual for subunits of nuclear RNA polymerases (Farago, Nahari et al. 2003; Goler-

Baron, Selitrennik et al. 2008). RPB4 and RPB7 localize to P-bodies and stimulate the de-adenylation of mRNAs (Lotan, Bar-On et al. 2005; Lotan, Goler-Baron et al. 2007). Nucleocytoplasmic shuttling of RPB4/RPB7 is dependent on transcription and suggests regulatory roles for RNA polymerase subunits, beyond the nuclear compartment (Selitrennik, Duek et al. 2006).

The NRPB4/NRPB7 subunits of *Arabidopsis* Pol II have previously been described as AtRPB19.5 ad AtRPB15.9 respectively and the two proteins form a soluble, heterodimeric complex in renaturation experiments (Larkin and Guilfoyle 1998). Purification of the Pol II complex from *Arabidopsis* suggests that the NRPB4 and NRPB7 might be associated stably (do not dissociate) with the core complex, similar to the archaeal polymerase (Ulmasov and Guilfoyle 1992). A forward genetic screen identified an atypical NRPB4-like subunit in RNA directed DNA methylation (He, Hsu et al. 2009) and this subunit was later determined by proteomic means to be NRPD4 and is shared between Pol IV and V (Ream, Haag et al. 2009). Pol II NRPB4 and NRPB7 subunits do not associate with Pol IV and V. Instead, Pol IV and V share a 4th largest subunit NRPD4/E4, as noted earlier. Pol V associates only with NRPE7 while Pol IV associates primarily with NRPD7, with some interaction with NRPE7 peptides, at lower frequency (Table 1 and (Ream, Haag et al. 2009)). Phylogenetic studies show that the 7th largest subunits of Pol IV and V diverged from NRPB7 via a cDNA retrotranposition, which occurred prior in an ancestor of moss and higher plants. The 4th largest subunits of Pol IV and V can be found beginning with the angiosperm lineage (Tucker, Reece et al. 2011).

Chapter three explores the roles of the 4th and 7th largest subunits of Pol II, IV and V and hints at their specialized functions in RNA directed DNA methylation as well as other transcriptional processes.

Other non-catalytic subunits

Other subunits that are shared between Pol II, IV and V are the 3rd largest, 6th largest and 8th largest subunits (each of which have two isoforms encoded by different genes). The 5th largest subunits are particularly interesting in that Pol V encodes a specialized NRPE5 subunit (Huang, Jones et al. 2009; Lahmy, Pontier et al. 2009; Ream, Haag et al. 2009) while Pol I, II, III and V utilizes the same NRPB5 subunit. The same NRPB5 subunit is also used by Pol I and III in *Arabidopsis* (Ream and Pikaard, unpublished (Saez-Vasquez and Pikaard 1997; Larkin, Hagen et al. 1999)). The evolutionary history of NRPE5 is similar to that of NRPD4, being found only in angiosperms and not found in the moss genome.

Preliminary results on the 6th largest subunit usage in Pol IV and V are discussed in Chapter four.

Pol V biochemistry and C-terminal domain

Biochemical purification of bacterial RNA polymerases, yeast Pol II and archaeal polymerases (Valenzuela, Bell et al. 1978; Zhang, Campbell et al. 1999; Werner and Weinzierl 2002) have been seminal steps for the deep biochemical understanding of each polymerase. The purification of *Arabidopsis* Pol IV and V is a worthy challenge for researchers. In vitro activity for both Pol IV and Pol V isolated from affinity-

immunoprecipitated complexes with a RNA/DNA hybrid template has now been shown in the Pikaard lab (Haag and Pikaard, unpublished). The series of non-catalytic subunit mutants that have been isolated in chapters two, three and four can be used to test biochemical properties of Pol IV and V to see if their biochemical activities are affected in these mutants. The Pol V largest subunit, NRPE1 has an extended C-terminal domain which is required for Pol V activity in vivo (Appendix C).

Chapter four also provides a framework for a forward genetic screen to isolate interactors with this important domain in Pol V.

REFERENCES CITED

Aapola, U., K. Kawasaki, et al. (2000). "Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family." Genomics **65**(3): 293-298.

Adenot, X., T. Elmayan, et al. (2006). "DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7." Curr Biol **16**(9): 927-932.

Agger, K., J. Christensen, et al. (2008). "The emerging functions of histone demethylases." Curr Opin Genet Dev **18**(2): 159-168.

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

Alleman, M., L. Sidorenko, et al. (2006). "An RNA-dependent RNA polymerase is required for paramutation in maize." Nature **442**(7100): 295-298.

Allen, E. and M. D. Howell (2010). "miRNAs in the biogenesis of trans-acting siRNAs in higher plants." Semin Cell Dev Biol **21**(8): 798-804.

Aravin, A. A., R. Sachidanandam, et al. (2008). "A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice." Mol Cell **31**(6): 785-799.

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

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

Ashapkin, V. V., L. I. Kutueva, et al. (2002). "The gene for domains rearranged methyltransferase (DRM2) in Arabidopsis thaliana plants is methylated at both cytosine and adenine residues." FEBS Lett **532**(3): 367-372.

Atanassov, B. S., E. Koutelou, et al. (2010). "The role of deubiquitinating enzymes in chromatin regulation." FEBS Lett.

Aufsatz, W., M. F. Mette, et al. (2002). "HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA." EMBO J **21**(24): 6832-6841.

Ausin, I., T. C. Mockler, et al. (2009). "IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana." Nat Struct Mol Biol **16**(12): 1325-1327.

Awrey, D. E., R. G. Weilbaecher, et al. (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.

Banaszynski, L. A., C. D. Allis, et al. (2010). "Histone variants in metazoan development." Dev Cell **19**(5): 662-674.

Bauer, M. J. and R. L. Fischer (2011). "Genome demethylation and imprinting in the endosperm." Curr Opin Plant Biol.

Baumberger, N. and D. C. Baulcombe (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.

Baurle, I., L. Smith, et al. (2007). "Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing." Science **318**(5847): 109-112.
Beisel, C. and R. Paro (2011). "Silencing chromatin: comparing modes and mechanisms." Nat Rev Genet **12**(2): 123-135.
Beisson, J. (2008). "Preformed cell structure and cell heredity." Prion **2**(1): 1-8.
Beisson, J. and T. M. Sonneborn (1965). "Cytoplasmic Inheritance of the Organization of the Cell Cortex in Paramecium Aurelia." Proc Natl Acad Sci U S A **53**: 275-282.
Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.
Bernstein, E., A. A. Caudy, et al. (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.
Bestor, T. H., V. L. Chandler, et al. (1994). "Epigenetic effects in eukaryotic gene expression." Dev Genet **15**(6): 458-462.
Bestor, T. H. and G. L. Verdine (1994). "DNA methyltransferases." Curr Opin Cell Biol **6**(3): 380-389.
Bies-Etheve, N., D. Pontier, et al. (2009). "RNA-directed DNA methylation requires an AGO4-interacting member of the SPT5 elongation factor family." EMBO Rep **10**(6): 649-654.
Bird, A. (1992). "The essentials of DNA methylation." Cell **70**(1): 5-8.
Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.
Blevins, T., R. Rajeswaran, et al. (2006). "Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing." Nucleic Acids Res **34**(21): 6233-6246.
Bohmert, K., I. Camus, et al. (1998). "AGO1 defines a novel locus of Arabidopsis controlling leaf development." EMBO J **17**(1): 170-180.
Boland, M. J. and J. K. Christman (2008). "Characterization of Dnmt3b:thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated repair by DNA methyltransferase(s) and RNA." J Mol Biol **379**(3): 492-504.
Borsani, O., J. Zhu, et al. (2005). "Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis." Cell **123**(7): 1279-1291.
Bostick, M., J. K. Kim, et al. (2007). "UHRF1 plays a role in maintaining DNA methylation in mammalian cells." Science **317**(5845): 1760-1764.
Brennecke, J., A. A. Aravin, et al. (2007). "Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila." Cell **128**(6): 1089-1103.
Brennecke, J., C. D. Malone, et al. (2008). "An epigenetic role for maternally inherited piRNAs in transposon silencing." Science **322**(5906): 1387-1392.
Brosnan, C. A., N. Mitter, et al. (2007). "Nuclear gene silencing directs reception of long-distance mRNA silencing in Arabidopsis." Proc Natl Acad Sci U S A **104**(37): 14741-14746.
Brouns, S. J., M. M. Jore, et al. (2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." Science **321**(5891): 960-964.

Bushnell, D. A. and R. D. Kornberg (2003). "Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription." Proc Natl Acad Sci U S A **100**(12): 6969-6973.
Cao, X., W. Aufsatz, et al. (2003). "Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation." Curr Biol **13**(24): 2212-2217.
Cao, X. and S. E. Jacobsen (2002). "Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes." Proc Natl Acad Sci U S A **99 Suppl 4**: 16491-16498.
Cao, X. and S. E. Jacobsen (2002). "Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing." Curr Biol **12**(13): 1138-1144.
Carthew, R. W. (2001). "Gene silencing by double-stranded RNA." Curr Opin Cell Biol **13**(2): 244-248.
Cerutti, H. and J. A. Casas-Mollano (2006). "On the origin and functions of RNA-mediated silencing: from protists to man." Curr Genet **50**(2): 81-99.
Chan, S. W., D. Zilberman, et al. (2004). "RNA silencing genes control de novo DNA methylation." Science **303**(5662): 1336.
Chapman, E. J. and J. C. Carrington (2007). "Specialization and evolution of endogenous small RNA pathways." Nat Rev Genet **8**(11): 884-896.
Chen, C. Y., C. C. Chang, et al. (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, X., C. Ruggiero, et al. (2007). "Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage." Mol Cell Biol **27**(13): 4617-4625.
Chen, Z. J. and C. S. Pikaard (1997). "Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance." Genes Dev **11**(16): 2124-2136.
Chen, Z. X. and A. D. Riggs (2011). "DNA methylation and demethylation in mammals." J Biol Chem.
Chinnusamy, V. and J. K. Zhu (2009). "RNA-directed DNA methylation and demethylation in plants." Sci China C Life Sci **52**(4): 331-343.
Choder, M. (2004). "Rpb4 and Rpb7: subunits of RNA polymerase II and beyond." Trends Biochem Sci **29**(12): 674-681.
Choi, Y., M. Gehring, et al. (2002). "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." Cell **110**(1): 33-42.
Christie, K. R., D. E. Awrey, et al. (1994). "Purified yeast RNA polymerase II reads through intrinsic blocks to elongation in response to the yeast TFIIS analogue, P37." J Biol Chem **269**(2): 936-943.
Chung, W. H., J. L. Craighead, et al. (2003). "RNA polymerase II/TFIIF structure and conserved organization of the initiation complex." Mol Cell **12**(4): 1003-1013.
Cloos, P. A., J. Christensen, et al. (2008). "Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease." Genes Dev **22**(9): 1115-1140.

Cogoni, C. and G. Macino (1999). "Gene silencing in Neurospora crassa requires a protein homologous to RNA-dependent RNA polymerase." Nature **399**(6732): 166-169.
Cokus, S. J., S. Feng, et al. (2008). "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." Nature **452**(7184): 215-219.
Costa-Nunes, P., O. Pontes, et al. (2010). "Extra views on RNA-dependent DNA methylation and MBD6-dependent heterochromatin formation in nucleolar dominance." Nucleus **1**(3): 254-259.
Cramer, P., K. J. Armache, et al. (2008). "Structure of eukaryotic RNA polymerases." Annu Rev Biophys **37**: 337-352.
Czech, B. and G. J. Hannon (2011). "Small RNA sorting: matchmaking for Argonautes." Nat Rev Genet **12**(1): 19-31.
Dalmay, T., A. Hamilton, et al. (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.
Damelin, M. and T. H. Bestor (2007). "Biological functions of DNA methyltransferase 1 require its methyltransferase activity." Mol Cell Biol **27**(11): 3891-3899.
Daxinger, L., T. Kanno, et al. (2009). "A stepwise pathway for biogenesis of 24-nt secondary siRNAs and spreading of DNA methylation." EMBO J **28**(1): 48-57.
Deleris, A., J. Gallego-Bartolome, et al. (2006). "Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense." Science **313**(5783): 68-71.
Deltcheva, E., K. Chylinski, et al. (2011). "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Nature **471**(7340): 602-607.
Denli, A. M., B. B. Tops, et al. (2004). "Processing of primary microRNAs by the Microprocessor complex." Nature **432**(7014): 231-235.
Dernburg, A. F., J. Zalevsky, et al. (2000). "Transgene-mediated cosuppression in the C. elegans germ line." Genes Dev **14**(13): 1578-1583.
Dimario, P. J. (2004). "Cell and molecular biology of nucleolar assembly and disassembly." Int Rev Cytol **239**: 99-178.
Djupedal, I., M. Portoso, et al. (2005). "RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing." Genes Dev **19**(19): 2301-2306.
Dunoyer, P., C. A. Brosnan, et al. (2010). "An endogenous, systemic RNAi pathway in plants." EMBO J **29**(10): 1699-1712.
Dunoyer, P., G. Schott, et al. (2010). "Small RNA duplexes function as mobile silencing signals between plant cells." Science **328**(5980): 912-916.
Earley, K., R. J. Lawrence, et al. (2006). "Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance." Genes Dev **20**(10): 1283-1293.
Earley, K. W., F. Pontvianne, et al. (2010). "Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation." Genes Dev **24**(11): 1119-1132.
Edwards, A. M., C. M. Kane, et al. (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.

Eissenberg, J. C. and S. C. Elgin (2000). "The HP1 protein family: getting a grip on chromatin." Curr Opin Genet Dev **10**(2): 204-210.
Elbashir, S. M., J. Harborth, et al. (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells." Nature **411**(6836): 494-498.
Elbashir, S. M., W. Lendeckel, et al. (2001). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.
Elbashir, S. M., J. Martinez, et al. (2001). "Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate." EMBO J **20**(23): 6877-6888.
Elgin, S. C. and S. I. Grewal (2003). "Heterochromatin: silence is golden." Curr Biol **13**(23): R895-898.
Erhard, K. F., Jr., J. L. Stonaker, et al. (2009). "RNA polymerase IV functions in paramutation in Zea mays." Science **323**(5918): 1201-1205.
Esteve, P. O., H. G. Chin, et al. (2006). "Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication." Genes Dev **20**(22): 3089-3103.
Farago, M., T. Nahari, et al. (2003). "Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress." Mol Biol Cell **14**(7): 2744-2755.
Feng, S., S. J. Cokus, et al. (2010). "Conservation and divergence of methylation patterning in plants and animals." Proc Natl Acad Sci U S A **107**(19): 8689-8694.
Feng, S., S. E. Jacobsen, et al. (2010). "Epigenetic reprogramming in plant and animal development." Science **330**(6004): 622-627.
Ficz, G., M. R. Branco, et al. (2011). "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." Nature.
Finigan, P. and R. A. Martienssen (2008). "Nucleolar dominance and DNA methylation directed by small interfering RNA." Mol Cell **32**(6): 753-754.
Fire, A., D. Albertson, et al. (1991). "Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle." Development **113**(2): 503-514.
Fire, A., S. Xu, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.
Forstemann, K., Y. Tomari, et al. (2005). "Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein." PLoS Biol **3**(7): e236.
Fuchs, J., D. Demidov, et al. (2006). "Chromosomal histone modification patterns--from conservation to diversity." Trends Plant Sci **11**(4): 199-208.
Gao, Z., H. L. Liu, et al. (2010). "An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation." Nature **465**(7294): 106-109.
Gehring, M., J. H. Huh, et al. (2006). "DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation." Cell **124**(3): 495-506.
Gehring, M., W. Reik, et al. (2009). "DNA demethylation by DNA repair." Trends Genet **25**(2): 82-90.
Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nat Rev Genet **10**(2): 94-108.

Gnatt, A. (2002). "Elongation by RNA polymerase II: structure-function relationship." Biochim Biophys Acta **1577**(2): 175-190.
Gnatt, A. L., P. Cramer, et al. (2001). "Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution." Science **292**(5523): 1876-1882.
Goler-Baron, V., M. Selitrennik, et al. (2008). "Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes." Genes Dev **22**(15): 2022-2027.
Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.
Gong, Z., T. Morales-Ruiz, et al. (2002). "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." Cell **111**(6): 803-814.
Greenblatt, J., J. R. Nodwell, et al. (1993). "Transcriptional antitermination." Nature **364**(6436): 401-406.
Grewal, S. I. (2010). "RNAi-dependent formation of heterochromatin and its diverse functions." Curr Opin Genet Dev **20**(2): 134-141.
Grewal, S. I. and S. C. Elgin (2007). "Transcription and RNA interference in the formation of heterochromatin." Nature **447**(7143): 399-406.
Grishok, A., A. E. Pasquinelli, et al. (2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing." Cell **106**(1): 23-34.
Grissa, I., G. Vergnaud, et al. (2008). "CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats." Nucleic Acids Res **36**(Web Server issue): W145-148.
Grohmann, D., A. Hirtreiter, et al. (2009). "RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro." Biochem J **421**(3): 339-343.
Grohmann, D., D. Klose, et al. (2010). "RNA-binding to archaeal RNA polymerase subunits F/E: a DEER and FRET study." J Am Chem Soc **132**(17): 5954-5955.
Grohmann, D. and F. Werner (2010). "Hold on!: RNA polymerase interactions with the nascent RNA modulate transcription elongation and termination." RNA Biol **7**(3): 310-315.
Grohmann, D. and F. Werner (2011). "Cycling through transcription with the RNA polymerase F/E (RPB4/7) complex: structure, function and evolution of archaeal RNA polymerase." Res Microbiol **162**(1): 10-18.
Grohmann, M., F. Spada, et al. (2005). "Restricted mobility of Dnmt1 in preimplantation embryos: implications for epigenetic reprogramming." BMC Dev Biol **5**: 18.
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.
Guo, J. U., Y. Su, et al. (2011). "Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain." Cell.
Ha, M., J. Lu, et al. (2009). "Small RNAs serve as a genetic buffer against genomic shock in Arabidopsis interspecific hybrids and allopolyploids." Proc Natl Acad Sci U S A **106**(42): 17835-17840.

Haag, J. R., O. Pontes, et al. (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.
Hale, C. J., J. L. Stonaker, et al. (2007). "A novel Snf2 protein maintains trans-generational regulatory states established by paramutation in maize." PLoS Biol **5**(10): e275.
Hamilton, A., O. Voinnet, et al. (2002). "Two classes of short interfering RNA in RNA silencing." EMBO J **21**(17): 4671-4679.
Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-952.
Hammond, S. M., A. A. Caudy, et al. (2001). "Post-transcriptional gene silencing by double-stranded RNA." Nat Rev Genet **2**(2): 110-119.
Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-251.
Hannon, G. J., F. V. Rivas, et al. (2006). "The expanding universe of noncoding RNAs." Cold Spring Harb Symp Quant Biol **71**: 551-564.
Hausner, W., U. Lange, et al. (2000). "Transcription factor S, a cleavage induction factor of the archaeal RNA polymerase." J Biol Chem **275**(17): 12393-12399.
Havecker, E. R., L. M. Wallbridge, et al. (2010). "The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci." Plant Cell **22**(2): 321-334.
He, X. J., T. Chen, et al. (2011). "Regulation and function of DNA methylation in plants and animals." Cell Res **21**(3): 442-465.
He, X. J., Y. F. Hsu, et al. (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.
He, X. J., Y. F. Hsu, et al. (2009). "A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development." Genes Dev **23**(23): 2717-2722.
He, X. J., Y. F. Hsu, et al. (2009). "An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein." Cell **137**(3): 498-508.
Hemming, S. A., D. B. Jansma, et al. (2000). "RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo." J Biol Chem **275**(45): 35506-35511.
Henderson, I. R., A. Deleris, et al. (2010). "The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana." PLoS Genet **6**(10): e1001182.
Henderson, I. R. and S. E. Jacobsen (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. (1990). "Position-effect variegation after 60 years." Trends Genet **6**(12): 422-426.
Henry, R. W., E. Ford, et al. (1998). "Crossing the line between RNA polymerases: transcription of human snRNA genes by RNA polymerases II and III." Cold Spring Harb Symp Quant Biol **63**: 111-120.

Hernandez, N. (2001). "Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription." J Biol Chem **276**(29): 26733-26736.
Herr, A. J., M. B. Jensen, et al. (2005). "RNA polymerase IV directs silencing of endogenous DNA." Science **308**(5718): 118-120.
Hiraguri, A., R. Itoh, et al. (2005). "Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in Arabidopsis thaliana." Plant Mol Biol **57**(2): 173-188.
Hirata, A., B. J. Klein, et al. (2008). "The X-ray crystal structure of RNA polymerase from Archaea." Nature **451**(7180): 851-854.
Hirtreiter, A., G. E. Damsma, et al. (2010). "Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif." Nucleic Acids Res **38**(12): 4040-4051.
Hirtreiter, A., D. Grohmann, et al. (2010). "Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro." Nucleic Acids Res **38**(2): 585-596.
Horton, J. R., G. Ratner, et al. (2004). "Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase." Nucleic Acids Res **32**(13): 3877-3886.
Howell, C. Y., T. H. Bestor, et al. (2001). "Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene." Cell **104**(6): 829-838.
Huang, L., A. M. Jones, et al. (2009). "An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II." Nat Struct Mol Biol **16**(1): 91-93.
Huh, J. H., M. J. Bauer, et al. (2008). "Cellular programming of plant gene imprinting." Cell **132**(5): 735-744.
Hull, M. W., K. McKune, et al. (1995). "RNA polymerase II subunit RPB9 is required for accurate start site selection." Genes Dev **9**(4): 481-490.
Hutvagner, G., J. McLachlan, et al. (2001). "A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA." Science **293**(5531): 834-838.
Hutvagner, G. and M. J. Simard (2008). "Argonaute proteins: key players in RNA silencing." Nat Rev Mol Cell Biol **9**(1): 22-32.
Jackson, J. P., A. M. Lindroth, et al. (2002). "Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase." Nature **416**(6880): 556-560.
Jacob, Y., S. Feng, et al. (2009). "ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing." Nat Struct Mol Biol **16**(7): 763-768.
Jacob, Y., H. Stroud, et al. (2010). "Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases." Nature **466**(7309): 987-991.
Jacobsen, S. E. (1999). "Gene silencing: Maintaining methylation patterns." Curr Biol **9**(16): R617-619.
Jacobsen, S. E., M. P. Running, et al. (1999). "Disruption of an RNA helicase/RNAse III gene in Arabidopsis causes unregulated cell division in floral meristems." Development **126**(23): 5231-5243.

James, T. C. and S. C. Elgin (1986). "Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene." Mol Cell Biol **6**(11): 3862-3872.
Jasiak, A. J., H. Hartmann, et al. (2008). "Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex." J Biol Chem **283**(39): 26423-26427.
Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-1080.
Jin, S. G., S. Kadam, et al. (2010). "Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine." Nucleic Acids Res **38**(11): e125.
Jore, M. M., S. J. Brouns, et al. (2011). "RNA in Defense: CRISPRs Protect Prokaryotes against Mobile Genetic Elements." Cold Spring Harb Perspect Biol.
Jore, M. M., M. Lundgren, et al. (2011). "Structural basis for CRISPR RNA-guided DNA recognition by Cascade." Nat Struct Mol Biol.
Jorgensen, H. F. and A. Bird (2002). "MeCP2 and other methyl-CpG binding proteins." Ment Retard Dev Disabil Res Rev **8**(2): 87-93.
Jorgensen, R. (1990). "Altered gene expression in plants due to trans interactions between homologous genes." Trends Biotechnol **8**(12): 340-344.
Jorgensen, R. A., P. D. Cluster, et al. (1996). "Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences." Plant Mol Biol **31**(5): 957-973.
Kaine, B. P., I. J. Mehr, et al. (1994). "The sequence, and its evolutionary implications, of a Thermococcus celer protein associated with transcription." Proc Natl Acad Sci U S A **91**(9): 3854-3856.
Kakutani, T., J. A. Jeddeloh, et al. (1995). "Characterization of an Arabidopsis thaliana DNA hypomethylation mutant." Nucleic Acids Res **23**(1): 130-137.
Kankel, M. W., D. E. Ramsey, et al. (2003). "Arabidopsis MET1 cytosine methyltransferase mutants." Genetics **163**(3): 1109-1122.
Kanno, T., E. Bucher, et al. (2010). "RNA-directed DNA methylation and plant development require an IWR1-type transcription factor." EMBO Rep **11**(1): 65-71.
Kanno, T., B. Huettel, et al. (2005). "Atypical RNA polymerase subunits required for RNA-directed DNA methylation." Nat Genet **37**(7): 761-765.
Kapoor, A., F. Agius, et al. (2005). "Preventing transcriptional gene silencing by active DNA demethylation." FEBS Lett **579**(26): 5889-5898.
Karginov, F. V. and G. J. Hannon (2010). "The CRISPR system: small RNA-guided defense in bacteria and archaea." Mol Cell **37**(1): 7-19.
Kasschau, K. D., N. Fahlgren, et al. (2007). "Genome-wide profiling and analysis of Arabidopsis siRNAs." PLoS Biol **5**(3): e57.
Katiyar-Agarwal, S., S. Gao, et al. (2007). "A novel class of bacteria-induced small RNAs in Arabidopsis." Genes Dev **21**(23): 3123-3134.
Katiyar-Agarwal, S., R. Morgan, et al. (2006). "A pathogen-inducible endogenous siRNA in plant immunity." Proc Natl Acad Sci U S A **103**(47): 18002-18007.
Kim, V. N., J. Han, et al. (2009). "Biogenesis of small RNAs in animals." Nat Rev Mol Cell Biol **10**(2): 126-139.

Kimura, M., H. Suzuki, et al. (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., A. Miura, et al. (2004). "One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation." Science **303**(5657): 521-523.
Klimasauskas, S., S. Kumar, et al. (1994). "HhaI methyltransferase flips its target base out of the DNA helix." Cell **76**(2): 357-369.
Koyama, H., T. Ito, et al. (2007). "Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast." Genes Cells **12**(5): 547-559.
Kwapisz, M., F. Beckouet, et al. (2008). "Early evolution of eukaryotic DNA-dependent RNA polymerases." Trends Genet **24**(5): 211-215.
Lahmy, S., N. Bies-Etheve, et al. (2010). "Plant-specific multisubunit RNA polymerase in gene silencing." Epigenetics **5**(1): 4-8.
Lahmy, S., D. Pontier, et al. (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 T. J. Guilfoyle (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., G. Hagen, et al. (1999). "Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5." Gene **231**(1-2): 41-47.
Law, J. A., I. Ausin, et al. (2010). "A protein complex required for polymerase V transcripts and RNA- directed DNA methylation in Arabidopsis." Curr Biol **20**(10): 951-956.
Lawrence, R. J., K. Earley, et al. (2004). "A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance." Mol Cell **13**(4): 599-609.
Lawrence, R. J. and C. S. Pikaard (2004). "Chromatin turn ons and turn offs of ribosomal RNA genes." Cell Cycle **3**(7): 880-883.
Lee, Y. S., K. Nakahara, et al. (2004). "Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways." Cell **117**(1): 69-81.
Lewis, M. S. and C. S. Pikaard (2001). "Restricted chromosomal silencing in nucleolar dominance." Proc Natl Acad Sci U S A **98**(25): 14536-14540.
Li, C. F., O. Pontes, et al. (2006). "An ARGONAUTE4-containing nuclear processing center colocalized with Cajal bodies in Arabidopsis thaliana." Cell **126**(1): 93-106.
Li, J., Z. Yang, et al. (2005). "Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis." Curr Biol **15**(16): 1501-1507.
Li, S., B. Ding, et al. (2006). "Evidence that the transcription elongation function of Rpb9 is involved in transcription-coupled DNA repair in Saccharomyces cerevisiae." Mol Cell Biol **26**(24): 9430-9441.
Lindroth, A. M., X. Cao, et al. (2001). "Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation." Science **292**(5524): 2077-2080.

Lister, R., R. C. O'Malley, et al. (2008). "Highly integrated single-base resolution maps of the epigenome in Arabidopsis." Cell **133**(3): 523-536.
Liu, J., M. A. Carmell, et al. (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-1441.
Liu, Q., Y. Feng, et al. (2009). "Dicer-like (DCL) proteins in plants." Funct Integr Genomics **9**(3): 277-286.
Lotan, R., V. G. Bar-On, et al. (2005). "The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs." Genes Dev **19**(24): 3004-3016.
Lotan, R., V. Goler-Baron, et al. (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., A. W. Mader, et al. (1997). "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature **389**(6648): 251-260.
Luo, J. and B. D. Hall (2007). "A multistep process gave rise to RNA polymerase IV of land plants." J Mol Evol **64**(1): 101-112.
Maillet, I., J. M. Buhler, et al. (1999). "Rpb4p is necessary for RNA polymerase II activity at high temperature." J Biol Chem **274**(32): 22586-22590.
Makarova, K. S., Y. I. Wolf, et al. (2009). "Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements." Biol Direct **4**: 29.
Mallory, A. and H. Vaucheret (2010). "Form, function, and regulation of ARGONAUTE proteins." Plant Cell **22**(12): 3879-3889.
Margis, R., A. F. Fusaro, et al. (2006). "The evolution and diversification of Dicers in plants." FEBS Lett **580**(10): 2442-2450.
Marmorstein, R. and R. C. Trievel (2009). "Histone modifying enzymes: structures, mechanisms, and specificities." Biochim Biophys Acta **1789**(1): 58-68.
Marraffini, L. A. and E. J. Sontheimer (2010). "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea." Nat Rev Genet **11**(3): 181-190.
Martienssen, R. (2010). "Molecular biology. Small RNA makes its move." Science **328**(5980): 834-835.
Martienssen, R. A. (2003). "Maintenance of heterochromatin by RNA interference of tandem repeats." Nat Genet **35**(3): 213-214.
Martienssen, R. A. (2010). "Heterochromatin, small RNA and post-fertilization dysgenesis in allopolyploid and interploid hybrids of Arabidopsis." New Phytol **186**(1): 46-53.
Mathieu, O., J. Reinders, et al. (2007). "Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation." Cell **130**(5): 851-862.
Matzke, M., T. Kanno, et al. (2009). "RNA-mediated chromatin-based silencing in plants." Curr Opin Cell Biol **21**(3): 367-376.
McClintock, B. (1934). "The relation of a particular chromosomal element to the development of the nucleoli in Zea mays." Cell and Tissue Research **21**(2): 294-326.
McGinnis, K. M., C. Springer, et al. (2006). "Transcriptionally silenced transgenes in maize are activated by three mutations defective in paramutation." Genetics **173**(3): 1637-1647.

McKune, K., K. L. Richards, et al. (1993). "RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability." Yeast **9**(3): 295-299.
McStay, B. (2006). "Nucleolar dominance: a model for rRNA gene silencing." Genes Dev **20**(10): 1207-1214.
McStay, B. and I. Grummt (2008). "The epigenetics of rRNA genes: from molecular to chromosome biology." Annu Rev Cell Dev Biol **24**: 131-157.
Meka, H., F. Werner, et al. (2005). "Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II." Nucleic Acids Res **33**(19): 6435-6444.
Meyer, P. (2010). "DNA methylation systems and targets in plants." FEBS Lett.
Mi, S., T. Cai, et al. (2008). "Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide." Cell **133**(1): 116-127.
Minakhin, L., S. Nechaev, et al. (2001). "Recombinant Thermus aquaticus RNA polymerase, a new tool for structure-based analysis of transcription." J Bacteriol **183**(1): 71-76.
Mitsuzawa, H., E. Kanda, et al. (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.
Mochizuki, K., N. A. Fine, et al. (2002). "Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena." Cell **110**(6): 689-699.
Molnar, A., C. Melnyk, et al. (2011). "Silencing signals in plants: a long journey for small RNAs." Genome Biol **12**(1): 215.
Molnar, A., C. W. Melnyk, et al. (2010). "Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells." Science **328**(5980): 872-875.
Morales-Ruiz, T., A. P. Ortega-Galisteo, et al. (2006). "DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases." Proc Natl Acad Sci U S A **103**(18): 6853-6858.
Morel, J. B., C. Godon, et al. (2002). "Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance." Plant Cell **14**(3): 629-639.
Morgan, H. D., W. Dean, et al. (2004). "Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming." J Biol Chem **279**(50): 52353-52360.
Mosher, R. A., C. W. Melnyk, et al. (2009). "Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis." Nature **460**(7252): 283-286.
Mosher, R. A., F. Schwach, et al. (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.
Moss, T. and V. Y. Stefanovsky (2002). "At the center of eukaryotic life." Cell **109**(5): 545-548.
Mourrain, P., C. Beclin, et al. (2000). "Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance." Cell **101**(5): 533-542.
Nan, X., S. Cross, et al. (1998). "Gene silencing by methyl-CpG-binding proteins." Novartis Found Symp **214**: 6-16; discussion 16-21, 46-50.

Napoli, C., C. Lemieux, et al. (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.
Naumann, U., L. Daxinger, et al. (2011). "Genetic Evidence That DNA Methyltransferase DRM2 Has a Direct Catalytic Role in RNA-Directed DNA Methylation in Arabidopsis thaliana." Genetics **187**(3): 977-979.
Ndlovu, M. N., H. Denis, et al. (2011). "Exposing the DNA methylome iceberg." Trends Biochem Sci.
Nesser, N. K., D. O. Peterson, et al. (2006). "RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo." Proc Natl Acad Sci U S A **103**(9): 3268-3273.
Ng, H. H. and A. Bird (1999). "DNA methylation and chromatin modification." Curr Opin Genet Dev **9**(2): 158-163.
Onodera, Y., J. R. Haag, et al. (2005). "Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation." Cell **120**(5): 613-622.
Onodera, Y., K. Nakagawa, et al. (2008). "Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis." Genetics **180**(1): 207-218.
Ooi, S. K. and T. H. Bestor (2008). "The colorful history of active DNA demethylation." Cell **133**(7): 1145-1148.
Ooi, S. K. and T. H. Bestor (2008). "Cytosine methylation: remaining faithful." Curr Biol **18**(4): R174-176.
Ooi, S. K., A. H. O'Donnell, et al. (2009). "Mammalian cytosine methylation at a glance." J Cell Sci **122**(Pt 16): 2787-2791.
Ooi, S. K., C. Qiu, et al. (2007). "DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA." Nature **448**(7154): 714-717.
Orlicky, S. M., P. T. Tran, et al. (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.
Ortega-Galisteo, A. P., T. Morales-Ruiz, et al. (2008). "Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks." Plant Mol Biol **67**(6): 671-681.
Pal-Bhadra, M., U. Bhadra, et al. (1997). "Cosuppression in Drosophila: gene silencing of Alcohol dehydrogenase by white-Adh transgenes is Polycomb dependent." Cell **90**(3): 479-490.
Pal-Bhadra, M., B. A. Leibovitch, et al. (2004). "Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery." Science **303**(5658): 669-672.
Penterman, J., R. Uzawa, et al. (2007). "Genetic interactions between DNA demethylation and methylation in Arabidopsis." Plant Physiol **145**(4): 1549-1557.
Penterman, J., D. Zilberman, et al. (2007). "DNA demethylation in the Arabidopsis genome." Proc Natl Acad Sci U S A **104**(16): 6752-6757.
Pham, J. W., J. L. Pellino, et al. (2004). "A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila." Cell **117**(1): 83-94.
Pikaard, C. S. (1999). "Nucleolar dominance and silencing of transcription." Trends Plant Sci **4**(12): 478-483.

Pikaard, C. S. (2000). "The epigenetics of nucleolar dominance." Trends Genet **16**(11): 495-500.
Pikaard, C. S. (2000). "Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids." Plant Mol Biol **43**(2-3): 163-177.
Pikaard, C. S., J. R. Haag, et al. (2008). "Roles of RNA polymerase IV in gene silencing." Trends Plant Sci **13**(7): 390-397.
Pikaard, C. S., S. Preuss, et al. (2005). "Detecting differential expression of parental or progenitor alleles in genetic hybrids and allopolyploids." Methods Enzymol **395**: 554-569.
Pikaard, C. S. and S. Tucker (2009). "RNA-silencing enzymes Pol IV and Pol V in maize: more than one flavor?" PLoS Genet **5**(11): e1000736.
Ponferrada-Marin, M. I., M. I. Martinez-Macias, et al. (2010). "Methylation-independent DNA binding modulates specificity of Repressor of Silencing 1 (ROS1) and facilitates demethylation in long substrates." J Biol Chem **285**(30): 23032-23039.
Pontes, O., R. J. Lawrence, et al. (2007). "Postembryonic establishment of megabase-scale gene silencing in nucleolar dominance." PLoS One **2**(11): e1157.
Pontes, O., C. F. Li, et al. (2006). "The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center." Cell **126**(1): 79-92.
Pontvianne, F., T. Blevins, et al. (2010). "Arabidopsis Histone Lysine Methyltransferases." Adv Bot Res **53**: 1-22.
Preuss, S. and C. S. Pikaard (2007). "rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch." Biochim Biophys Acta **1769**(5-6): 383-392.
Preuss, S. B., P. Costa-Nunes, et al. (2008). "Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins." Mol Cell **32**(5): 673-684.
Probst, A. V., M. Fagard, et al. (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.
Qi, Y., X. He, et al. (2006). "Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation." Nature **443**(7114): 1008-1012.
Rajakumara, E., J. A. Law, et al. (2011). "A dual flip-out mechanism for 5mC recognition by the Arabidopsis SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo." Genes Dev **25**(2): 137-152.
Ream, T. S., J. R. Haag, et al. (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 S. C. Elgin (2002). "Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects." Cell **108**(4): 489-500.
Richmond, T. J. and C. A. Davey (2003). "The structure of DNA in the nucleosome core." Nature **423**(6936): 145-150.
Rivas, F. V., N. H. Tolia, et al. (2005). "Purified Argonaute2 and an siRNA form recombinant human RISC." Nat Struct Mol Biol **12**(4): 340-349.
Romano, N. and G. Macino (1992). "Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences." Mol Microbiol **6**(22): 3343-3353.

Ruan, W., E. Lehmann, et al. (2011). "Evolution of two modes of intrinsic RNA polymerase transcript cleavage." J Biol Chem.
Runner, V. M., V. Podolny, et al. (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 C. S. Pikaard (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. and M. C. Siomi (2010). "Small RNA-mediated quiescence of transposable elements in animals." Dev Cell **19**(5): 687-697.
Sampath, V., B. Balakrishnan, et al. (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.
Schauer, S. E., S. E. Jacobsen, et al. (2002). "DICER-LIKE1: blind men and elephants in Arabidopsis development." Trends Plant Sci **7**(11): 487-491.
Schiebel, W., T. Pelissier, et al. (1998). "Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato." Plant Cell **10**(12): 2087-2101.
Schramke, V., D. M. Sheedy, et al. (2005). "RNA-interference-directed chromatin modification coupled to RNA polymerase II transcription." Nature **435**(7046): 1275-1279.
Schwartz, Y. B. and V. Pirrotta (2007). "Polycomb silencing mechanisms and the management of genomic programmes." Nat Rev Genet **8**(1): 9-22.
Selitrennik, M., L. Duek, et al. (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., M. Varon, et al. (1999). "Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4." Mol Cell Biol **19**(4): 2672-2680.
Sidorenko, L., J. E. Dorweiler, et al. (2009). "A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes." PLoS Genet **5**(11): e1000725.
Siomi, M. C., K. Sato, et al. (2011). "PIWI-interacting small RNAs: the vanguard of genome defence." Nat Rev Mol Cell Biol **12**(4): 246-258.
Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-285.
Slotkin, R. K., M. Vaughn, et al. (2009). "Epigenetic reprogramming and small RNA silencing of transposable elements in pollen." Cell **136**(3): 461-472.
Smardon, A., J. M. Spoerke, et al. (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.
Smith, L. M., O. Pontes, 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**(5): 1507-1521.

Song, J., O. Rechkoblit, et al. (2011). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." Science **331**(6020): 1036-1040.
Stevenson, D. S. and P. Jarvis (2003). "Chromatin silencing: RNA in the driving seat." Curr Biol **13**(1): R13-15.
Sun, Z. W., A. Tessmer, et al. (1996). "Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in Saccharomyces cerevisiae." Nucleic Acids Res **24**(13): 2560-2566.
Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-476.
Swiezewski, S., P. Crevillen, 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**(9): 3633-3638.
Tabara, H., M. Sarkissian, et al. (1999). "The rde-1 gene, RNA interference, and transposon silencing in C. elegans." Cell **99**(2): 123-132.
Tabara, H., E. Yigit, et al. (2002). "The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans." Cell **109**(7): 861-871.
Takeda, S. and J. Paszkowski (2006). "DNA methylation and epigenetic inheritance during plant gametogenesis." Chromosoma **115**(1): 27-35.
Tan, Q., M. H. Prysak, et al. (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.
Taverna, S. D., H. Li, et al. (2007). "How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers." Nat Struct Mol Biol **14**(11): 1025-1040.
Teixeira, F. K., F. Heredia, et al. (2009). "A role for RNAi in the selective correction of DNA methylation defects." Science **323**(5921): 1600-1604.
Tijsterman, M., R. F. Ketting, et al. (2002). "The genetics of RNA silencing." Annu Rev Genet **36**: 489-519.
Timmons, L. and A. Fire (1998). "Specific interference by ingested dsRNA." Nature **395**(6705): 854.
Tsai, M. C., O. Manor, et al. (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science **329**(5992): 689-693.
Tucker, S., A. Vitins, et al. (2010). "Nucleolar dominance and ribosomal RNA gene silencing." Curr Opin Cell Biol **22**(3): 351-356.
Tucker, S. L., J. Reece, et al. (2011). "Evolutionary History of Plant Multisubunit RNA Polymerases IV and V: Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization." Cold Spring Harb Symp Quant Biol.
Tuite, M. F. and T. R. Serio (2010). "The prion hypothesis: from biological anomaly to basic regulatory mechanism." Nat Rev Mol Cell Biol **11**(12): 823-833.
Ujvari, A. and D. S. Luse (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 T. J. Guilfoyle (1992). "Sequence of the fifth largest subunit of RNA polymerase II from plants." J Biol Chem **267**(32): 23165-23169.

Valenzuela, P., G. I. Bell, et al. (1978). "Isolation and assay of eukaryotic DNA-dependent RNA polymerases." Methods Cell Biol **19**: 1-26.
van der Oost, J. and S. J. Brouns (2009). "RNAi: prokaryotes get in on the act." Cell **139**(5): 863-865.
Van Emburgh, B. O. and K. D. Robertson (2011). "Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants." Nucleic Acids Res.
Vaucheret, H. (2008). "Plant ARGONAUTES." Trends Plant Sci **13**(7): 350-358.
Voinnet, O. (2005). "Non-cell autonomous RNA silencing." FEBS Lett **579**(26): 5858-5871.
Walmacq, C., M. L. Kireeva, et al. (2009). "Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II." J Biol Chem **284**(29): 19601-19612.
Wang, B., D. N. Jones, et al. (1998). "High-resolution structure of an archaeal zinc ribbon defines a general architectural motif in eukaryotic RNA polymerases." Structure **6**(5): 555-569.
Wang, M. B. and E. S. Dennis (2009). "SPT5-like, a new component in plant RdDM." EMBO Rep **10**(6): 573-575.
Wang, Y., W. Fischle, et al. (2004). "Beyond the double helix: writing and reading the histone code." Novartis Found Symp **259**: 3-17; discussion 17-21, 163-169.
Wassenegger, M. and G. Krczal (2006). "Nomenclature and functions of RNA-directed RNA polymerases." Trends Plant Sci **11**(3): 142-151.
Waterhouse, P. M., M. W. Graham, et al. (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.
Werner, F. (2007). "Structure and function of archaeal RNA polymerases." Mol Microbiol **65**(6): 1395-1404.
Werner, F. (2008). "Structural evolution of multisubunit RNA polymerases." Trends Microbiol **16**(6): 247-250.
Werner, F. and D. Grohmann (2011). "Evolution of multisubunit RNA polymerases in the three domains of life." Nat Rev Microbiol **9**(2): 85-98.
Werner, F. and R. O. Weinzierl (2002). "A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription." Mol Cell **10**(3): 635-646.
Werner, M., P. Thuriaux, et al. (2009). "Structure-function analysis of RNA polymerases I and III." Curr Opin Struct Biol **19**(6): 740-745.
Wierzbicki, A. T., J. R. Haag, et al. (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., T. S. Ream, et al. (2009). "RNA polymerase V transcription guides ARGONAUTE4 to chromatin." Nat Genet **41**(5): 630-634.
Williams, K., J. Christensen, et al. (2011). "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity." Nature.
Wion, D. and J. Casadesus (2006). "N6-methyl-adenine: an epigenetic signal for DNA-protein interactions." Nat Rev Microbiol **4**(3): 183-192.

Woo, H. R., T. A. Dittmer, et al. (2008). "Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis." PLoS Genet **4**(8): e1000156.
Wossidlo, M., T. Nakamura, et al. (2011). "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." Nat Commun **2**: 241.
Woychik, N. A. and M. Hampsey (2002). "The RNA polymerase II machinery: structure illuminates function." Cell **108**(4): 453-463.
Woychik, N. A., W. S. Lane, et al. (1991). "Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes." J Biol Chem **266**(28): 19053-19055.
Woychik, N. A. and R. A. Young (1989). "RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth." Mol Cell Biol **9**(7): 2854-2859.
Wu, H., A. C. D'Alessio, et al. (2011). "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells." Nature.
Wu, S. C. and Y. Zhang (2010). "Active DNA demethylation: many roads lead to Rome." Nat Rev Mol Cell Biol **11**(9): 607-620.
Xie, Z., E. Allen, et al. (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. and X. Qi (2008). "Diverse small RNA-directed silencing pathways in plants." Biochim Biophys Acta **1779**(11): 720-724.
Yura, T. and A. Ishihama (1979). "Genetics of bacterial RNA polymerases." Annu Rev Genet **13**: 59-97.
Zemach, A., I. E. McDaniel, et al. (2010). "Genome-wide evolutionary analysis of eukaryotic DNA methylation." Science **328**(5980): 916-919.
Zhang, G., E. A. Campbell, et al. (1999). "Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution." Cell **98**(6): 811-824.
Zhang, X., I. R. Henderson, et al. (2007). "Role of RNA polymerase IV in plant small RNA metabolism." Proc Natl Acad Sci U S A **104**(11): 4536-4541.
Zhang, Y., R. Jurkowska, et al. (2010). "Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail." Nucleic Acids Res **38**(13): 4246-4253.
Zheng, X., O. Pontes, et al. (2008). "ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis." Nature **455**(7217): 1259-1262.
Zheng, X., J. Zhu, et al. (2007). "Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing." EMBO J **26**(6): 1691-1701.
Zhou, K., W. H. Kuo, et al. (2009). "Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5." Proc Natl Acad Sci U S A **106**(17): 6956-6961.
Zhu, J. K. (2009). "Active DNA demethylation mediated by DNA glycosylases." Annu Rev Genet **43**: 143-166.
Ziegler, L. M., D. A. Khaperskyy, et al. (2003). "Yeast RNA polymerase II lacking the Rpb9 subunit is impaired for interaction with transcription factor IIF." J Biol Chem **278**(49): 48950-48956.
Zilberman, D., X. Cao, et al. (2003). "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." Science **299**(5607): 716-719.

Zilberman, D., X. Cao, et al. (2004). "Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats." Curr Biol **14**(13): 1214-1220.

CHAPTER 2

ARABIDOPSIS RNA POLYMERASE V SUBTYPES HAVING ALTERNATIVE NINTH SUBUNITS DIFFER IN RNA DIRECTED DNA METHYLATION ACTIVITY

Arabidopsis RNA Polymerase V subtypes having alternative ninth subunits differ in RNA-directed DNA methylation activity

Ek Han Tan[1,2], Todd Blevins[1], Thomas S. Ream[3], Craig S. Pikaard[1*]

[1]Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405

[2]Division of Biology and Biomedical Sciences, Washington University in St Louis, St Louis, Missouri 63110

[3]Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

* To whom correspondence should be addressed

Email addresses:

EHT: ehtan@wustl.edu

CSP: pikaard@indiana.edu

TB: toddblev@indiana.edu

TSR: tream@wisc.edu

** In submission to Genes and Development

Abstract

Multisubunit RNA polymerases IV and V (Pol IV and Pol V) evolved as specialized forms of Pol II that mediate RNA silencing in plants. Arabidopsis Pols II, IV and V associate with alternative ninth subunits that are 93% identical. These alternative subunits, NRPB9a and NRPB9b, are redundant for Pol II-dependent viability but not RNA-directed DNA methylation (RdDM), which is abrogated in *nrpb9b*, but not *nrpb9a*, mutants. Pol IV-dependent siRNA biogenesis and silencing is unaffected in *nrpb9b* mutants but silencing requiring Pol V is disrupted, indicating that the NRPB9b-containing Pol V subtype is specifically required for RdDM at these loci.

Introduction

Eukaryotes decode their genomes utilizing three essential nuclear DNA-dependent RNA polymerases, abbreviated as Pol I, Pol II and Pol III (Cramer et al. 2008; Werner et al. 2009). In plants, two additional multisubunit RNA polymerases, Pol IV and Pol V are nonessential for viability but play key roles in the silencing of endogenous DNA repeats and transposable elements via RNA-directed DNA methylation, a process in which 24nt short interfering RNAs (siRNAs) specify the cytosine methylation of complementary DNA sequences (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005; Pontier et al. 2005). Genetic and cytological evidence indicate that Pol IV acts early in the pathway, generating primary RNA transcripts that serve as templates for RNA-DEPENDENT RNA POLYMERASE 2. Resulting double-stranded RNAs are diced into siRNA duplexes by DICER-LIKE 3 (DCL3) and 24 nt single-stranded siRNAs are then loaded into an Argonaute protein, primarily AGO4. Independent of siRNA

biogenesis, Pol V generates transcripts at most loci that are subject to RdDM. AGO4 binds to these RNAs, presumably guided by siRNA base-pairing to the Pol V transcripts as well as by interactions with the C-terminal domain of the Pol V largest subunit. In subsequent steps that are not well understood, chromatin modifying activities are recruited, resulting in de novo cytosine methylation and the establishment of repressive histone modifications.

Subunit compositions of affinity-purified Pols II, IV and V of *Arabidopsis thaliana*, determined by mass spectrometry, show that each enzyme has 12 subunits, seven of which are encoded by the same genes. The largest subunit is unique in all three enzymes and the second-largest subunit of Pol II is distinct from the corresponding subunit of Pols IV or V, the latter two of which are encoded by the same gene. Together, the largest and second-largest subunits form the template channel and the catalytic center for RNA synthesis. The ten non-catalytic subunits play roles in the assembly or stabilization of the catalytic subunits and/or the interaction with regulators of transcription initiation, elongation or transcript processing.

In yeast, the twelve Pol II subunits are encoded by single genes, ten of which are essential. The exceptions are *rpb9* and *rpb4* deletion strains, which are viable, but temperature-sensitive (Woychik et al. 1991; Maillet et al. 1999). Deletion of both *RPB4* and *RPB9* is synthetically lethal (Li and Smerdon 2002). RPB4, together with RPB7, forms a sub-complex implicated in multiple steps of RNA elongation, termination and processing. RPB9 is implicated in multiple aspects of transcription including including transcription initiation, via interactions with general transcription factors (Hull et al. 1995; Sun et al. 1996), transcript elongation and processivity (Awrey et al. 1997;

Hemming et al. 2000) and transcript cleavage upon misincorporation of improperly template nucleotides (Nesser et al. 2006; Walmacq et al. 2009). *In vitro* experiments have shown that RPB9 possesses weak intrinsic RNA cleavage activity that is stimulated in association with transcription factor TFIIS, a protein with which RPB9 shares sequence similarity (Koyama et al. 2007). RPB9 has also been found to be involved in transcription-coupled nucleotide excision repair of DNA (Li et al. 2006) and in the degradation of the Pol II largest subunit, RPB1 in response to UV damage (Chen et al. 2007). Collectively, these studies implicate RPB9 in multiple important pathways despite the viability of *rpb9* deletion mutants under laboratory conditions.

Arabidopsis thaliana expresses two genes homologous to yeast RPB9. The encoded proteins, NRPB9a and NRPB9b differ at only eight out of 114 amino acid positions. Both NRPB9 variants are detected by mass spectrometry in affinity purified samples of RNA polymerases II, IV or V that are free of cross-contamination with each other or other RNA polymerases. Likewise, all three polymerase make use of highly similar variants for the third, sixth, or eighth subunits. The simplest hypothesis is that alternative variants for these subunits are functionally redundant. In keeping with this expectation, we show that *nrpd9a* or *nrpb9b* single mutants are viable, but the double mutant is embryo lethal, indicating that either NRPB9a or NRPB9b is sufficient for viability, attributable to their redundant functions in the context of Pol II. Surprisingly, NRPB9a and NRPB9b are not redundant with respect to RNA-directed DNA methylation (RdDM) and gene silencing; whereas *nrpb9b* mutants are disrupted in RdDM, *nrpb9a* mutants have no mutant phenotype. We show that accumulation of 24 nt siRNAs is unaffected in *nrpb9b*, suggesting that Pol IV transcription of siRNA precursor transcripts

is not impaired. Likewise, analysis of a locus (*MRD1*) at which silencing and RdDM is dependent only on Pol IV, and not on Pol V, shows that silencing and RdDM is unaffected in *nrpb9b* mutants. However, at several well-studied loci whose silencing involves RNA-directed DNA methylation (RdDM) orchestrated by both Pol IV and Pol V, *nrpb9b* mutants show a loss of RdDM. Collectively, our results indicate that the use of alternative ninth subunits has functional consequences for Pol V activity, with the NRPB9b-containing Pol V subtype being specifically required for silencing at many, if not all, loci. Because Pol V transcripts are still detected in *nrpb9b* mutants, we propose that NRPB9b is not required for Pol V transcription but mediates interactions with proteins responsible for silencing at transcribed loci.

Materials and Methods

Plant materials

nrpb9a-1 (Salk_032670) and *nrpb9b-1* (Salk_031043) mutants bearing T-DNA insertions in the coding regions were obtained from the Arabidopsis Biological Resource Center, Ohio State University. Plants were grown on soil using a 18 hour light, 6 hour dark regimen. Genotyping involved PCR using primers flanking the insertion sites (Salk032670_LP: 5’cagacaaagaacagtgtcattcc, Salk032670_RP: 5’ttctggaattgcacctctctg, Salk031043_LP: 5’gatataaaggtgcatggggatatgc, Salk031043_RP: 5’taaactcattaaattatcattccttgg) or the T-DNA left border (LBa1: tggttcacgtagtgggccatcg).

RT-PCR assays

Total RNA was isolated using Trizol (Invitrogen). Reverse transcription was performed

on 100 ng of RQ1 DNAse (Promega)-treated RNA using SuperScriptIII (Invitrogen) and gene specific primers in order to generate cDNAs. PCR was performed using HotStart Taq (Fermentas). Primers for *soloLTR, ATSN1* and *Actin* were previously described (Wierzbicki et al. 2008). Primers for *NRPB9a* and *NRPB9b* transcripts were: 9a_5'UTR: gtgattcagttttggttttggaacctaa, 9b_5'UTR: gtgaaatcaaagaagcattcaaaagctc, 9aRev: ttctctccagcgatgaccac and 9bRev: ttctctccaacggtgactacagtt.

DNA methylation assays

DNA was extracted using a Nucleon PhytoPure Genomic Extraction Kit (GE Healthcare). 1ug DNA was subjected to restriction endonuclease digestion, electrophoresis on a 0.9% TAE gel and Southern blot hybridization using a *5S* rRNA gene probe as described in (Blevins et al. 2009). For methylation-sensitive PCR, 1 ug DNA were digested overnight with the appropriate restriction enzyme (New England Biolabs) and 50 ng of DNA was then used in PCR reactions using primers that span the predicted methylation sites, using GoTaq Green polymerase (Fermentas) (Wierzbicki et al. 2008).

Small RNA Northern blots

Total RNA was fractionated using an RNeasy Kit (Qiagen) as described in (Blevins 2010). 4ug of low molecular weight RNA was subjected to electrophoresis on a 12% denaturing polyacrylamide gel and transferred onto Hybond membranes (GE Healthcare). Prehybridization and hybridization in PerfectHyb buffer (Sigma) was performed at 37°C. DNA oligonucleotides that were ^{32}P end-labeled using T4 Polynucleotide Kinase (New England Biolabs) and used for blot hybridizations were siR1003:

atgccaagtttggcctcacggtct, soloLTR: tgtcattatccatcattcatctctatccataag and miR160: tggcatacagggagccaggca.

Cloning and complementation

A genomic clone for *NRPB9b*, including the promoter region, was obtained by PCR amplification of *A. thaliana* genomic DNA using PFU Ultra DNA polymerase (Stratagene), and primers: 9bPromF: caccgcacttcaacaacccaattaca and 9bRev: ttctctccaacggtgactacagtt. PCR products were captured in the pENTR D/TOPO (Invitrogen) entry vector and recombined into pEARLEYGATE 302 using LR CLonase II (Invitrogen) thereby adding a C-terminal FLAG epitope tag in place of the stop codon (Earley et al. 2006). The resulting construct was transformed into *nrpb9b-1* homozygous mutant plants using the floral dip method for *Agrobacterium*-mediated gene transfer.

Results and Discussion

RPB9 diversity and

The *RPB9* subunit of Pol II is encoded by a single-copy gene in yeast and metazoans but has multiple paralogs in plants. For instance, *Populus trichocarpa* (poplar) and *Arabidopsis thaliana* have two *RPB9* variants in the Pol II clade and maize and rice have three identifiable *RPB9*-like genes (Figure 1; Figure S1). RPA12 and RPC11 are the RPB9-homologous subunits of Pol I and Pol III respectively, and these proteins in plants, yeast and vertebrates group together in clades distinct from the Pol II/IV/V RPB9 family.

Arabidopsis *NRPB9a* (*At3g16980*) and *NRPB9b* (*At4g16265*) share similar intron/exon structures (Figure 2a) but differ in their promoter and intron sequences. T-DNA insertion alleles of *NRPB9a* and *NRPB9b*, designated *nrpb9a-1* (Salk_032670) and *nrpb9b-1* (Salk_031043), are disrupted within introns 2 or 1, respectively (Figure 2a). Genotyping of progeny bearing insertion alleles revealed that homozygous *nrpb9a-1* or *nrpb9b-1* mutants are each viable. *NRPB9a* and *NRPB9b* are expressed at similar levels in wild-type plants (both in flowers and leaves) but their transcripts are not detected in *nrpb9a-1* or *nrpb9b-1* mutants suggesting that these are null mutants (Figure 2b). **[where are the primer pairs? Add to figure part a]**

Whereas *nrpb9a-1* mutants are indistinguishable from wild-type plants (ecotype Col-0), leaves of *nrpb9b-1* are more ovate, have shorter petioles and display less downward edge curling (Figure 2c). These morphological differences are most likely due to altered Pol II transcription profiles given that null mutants for the catalytic largest subunits of Pol IV or Pol V (*nrpd1-3* or *nrpe1-11*, respectively) do not display morphological phenotypes. Moreover, the *nrpb9b-1* mutant phenotype persists in *nrpb9b-*

1 nrpd1-3 or *nrpb9b-1 nrpe1-11* double mutants (data not shown) ruling out a positive role for Pol IV or Pol V-dependent pathways in the phenotype.

NRPB9a and NRPB9b are redundant for viability

We crossed *nrpb9a-1* and *nrpb9b-1* plants to generate F1 plants that were heterozygous at each locus and then examined their progeny in subsequent generations following self-fertilization. In siliques of plants homozygous for *nrpb9a-1* and heterozygous for *nrpb9b-1*, in which 25% of the seeds are expected to be homozygous double mutants, 30 % (55/181 analyzed) of the seeds were arrested in development and 70% developed properly (indicated by red and blue arrows, respectively, in Figure 2d). Similar results were obtained for the progeny of plants homozygous for *nrpb9b-1* but heterozygous for *nrpb9a-1*. Differential Interference Contrast microscopy revealed that in undeveloped seeds, embryos failed to develop past the globular stage (Figure 2d).

Upon sowing the seeds of plants homozygous for *nrpb9a-1* and heterozygous for *nrpb9b*-or vice versa, and genotyping their progeny, no *nrpb9a-1 nrpb9b-1* double mutants were identified, but all other expected genotypes were detected (Figure 2e). We previously demonstrated that null mutants for catalytic subunits of Pol I, II or III are lethal, whereas equivalent Pol IV or V mutations are not lethal. Therefore, we interpret the results of Figures 2d and 2e to indicate that NRPB9a and NRPB9b are redundant with respect to RNA polymerase II functions that are essential for viability, such that either gene is sufficient for embryonic and post-embryonic development but the double mutant results in embryo lethality. However, the NRPB9a and NRPB9b genes must not be

completely redundant with regard to Pol II functions, as evident from the distinctive leaf morphology phenotypes observed only in *nrpb9b* mutants.

Transmission of nrpb9a-1 and nrpb9b-1 mutations via male or female gametophytes

Among the progeny of a heterozygote carrying a recessive allele of an essential gene, heterozygotes should outnumber homozygotes 2:1. However, among 129 genotyped progeny of a plant homozygous for *nrpb9a-1* and heterozygous for *nrpb9b-1*, 69 were found to be heterozygous for *nprb9b-1* and 60 were wild-type for *NRPB9b* (Figure 2e). Likewise, among the progeny of a plant homozygous for *nrpb9b-1* and heterozygous for *nrpb9a-1*, 61 plants were homozygous for *nrpb9a* and 68 were homozygous wild-type (Figure 2e). The nearly 1:1 ratio of heterozygotes to homozygotes suggested a defect in the transmission of *nrpb9a-1*and *nrpb9b-1* alleles through the haploid male or female gametophyes, or both, such that fewer heterozygotes than expected are recovered. To test this hypothesis, reciprocal crosses were performed between wild-type plants and plants homozygous for *nrpb9a-1* but heterozygous for *nrpb9b-1* or plants homozygous for *nrpb9b-1* but heterozygous for *nrpb9a-1*. Genotyping of resulting F1 progeny showed reduced transmission of the mutant alleles through both the female and male gametophytes (Figure S2). Interestingly, this allele transmission behavior differs from that of null alleles for catalytic subunits of Pols I, II and III which cannot be maternally transmitted because the female gametophyte fails to develop to maturity, and thus egg cells are never fertilized. The fact that *nrpb9a nrpb9b* double mutants are transmitted maternally, albeit at reduced frequency, indicates that a functional ninth subunit is not required for Pol II functions necessary for development of

the haploid, 8-celled female gametophyte, unlike the catalytic subunits. Instead, NRPB9 is not essential until early embryogenesis. Therefore, we conclude that the ninth subunit of Pol II is partially dispensable for viability in plants, at least for gametophyte viability, in keeping with the viability of rpb9 deletion strains of yeast.

NRPB9b is required for RNA silencing

Pol IV and Pol V are required for the silencing retrotransposons and other genomic repeats. For example, soloLTR and AtSN1 retroelement expression is undetectable in wild-type plants (Col-0) using an RT-PCR assay (Figure 3A), but readily detected in *nrpd1-3* or *nrpe1-11* mutants defective for the largest subunits of Pol IV or Pol V, respectively, in agreement with prior studies (Onodera et al. 2005; Huettel et al. 2006). In *nrpb9a-1* mutants, soloLTR and AtSN1 elements remain silenced, as in wildtype (Figure 3A). By contrast, these elements are transcriptionally active in *nrpb9b-1* mutants, showing that NRPB9a and NRPB9b are not redundant with respect to transposon silencing. (Figure 3A).

Retrotransposon silencing is correlated with cytosine hypermethylation. For instance, *Alu*I sites within *soloLTR* elements and *Hae*III sites within *AtSN1* elements are subject to Pol IV and Pol V-dependent RdDM. As a result, these sites are resistant to *Alu*I or *Hae*III digestion in wild-type plants such that PCR using primers that flank the restriction sites yields a product of the expected size (Figure 3B). By contrast, in Pol IV or Pol V catalytic mutants (*nrpd1-3* or *nrpe1-11*), loss of methylation allows the enzymes to cleave, such that PCR amplification fails (Figure 3B). Using this technique, we detect

wild-type levels of cytosine methylation at *AtSN1* and *soloLTR* loci in *nrpb9a-1* mutants, but in *nrpb9b-1* mutants, methylation is lost (Figure 3B).

To determine if repeated genes, as well as transposons, lose DNA methylation, the methylation states of 5S rRNA gene repeats were also assayed. Southern blotting using genomic DNA digested with methylation sensitive restriction enzymes, followed by hybridization to a 5S gene probe were performed. Methylated DNA is resistant to digestion such that bands at the top of the gel represent 5S gene repeats that are more methylated than bands at the bottom of the gel, corresponding to 5S genes that have been digested. Using *HpaII*, which assays for symmetric CG methylation, the Southern blot shows that CG methylation is reduced in the Pol IV and V catalytic mutants, *nrpd1-3* and *nrpe1-11*, resulting in more bands at the bottom of the gel (Figure 3c). However, CG methylation of 5S genes does not change dramatically upon mutation of the *nrpb9a-1* or *nrpb9b-1* non-catalytic subunits. Repeating the Southern blot assay for asymmetric CHH methylation using *Hae*III shows reduced methylation in *nrpd1-3, nrpe1-1* and *nrpb9b-1*. These data show that CHH cytosine methylation is impaired in *nrpb9b-1*, as in mutants of Pol IV and V, whereas mutants of *nrpb9a-1* have no effect, suggesting that NRPB9b, but not NRPB9a, is required for RdDM.

Pol IV and V have different roles in siRNA biogenesis; Pol IV is required for the biogenesis of 24nt heterochromatic siRNAs, whereas Pol V is not required for the generation of these siRNAs at most loci. However, 24nt siRNA levels are reduced at some loci in Pol V largest subunit mutants (Herr et al. 2005; Kanno et al. 2005; Onodera et al. 2005). To assay the abundance of small RNAs, small RNA blots were performed to compare a panel of Pol IV and V subunit mutants (Figure 3d). As expected, small RNAs

corresponding to 5S rRNA gene clusters (siR1003) are completely lost in *nrpd1-3* mutants but are present, albeit at substantially reduced levels in *nrpe1-11* (Onodera et al. 2005). Interestingly, siR1003 accumulation in *nrpb9a-1* and *nrpb9b-1* occur similar to wildtype levels, despite the apparent loss of CHH DNA methylation at 5S genes in *nrpb9b-1*. Likewise, 24nt siRNAs from *soloLTR* accumulate to the same levels in Col-0, *nrpb9a-1* and *nrpb9b-1* despite losses of DNA methylation and silencing of this locus in *nrpb9b-1*.

Complementation of nrpb9b-1

Using a transgene containing a genomic clone of *NRPB9b,* transcribed from its native promoter, the mutant phenotypes of *nrpb9b-1* can be rescued. Figure S3a shows four independent transgenic lines that appear wildtype, as opposed to the displaying ovate leaves of the *nrpb9b-1* mutant. The transcription of the *NRPB9b* gene is also restored in mutant lines expressing the *NRPB9b* transgene, as shown using RT-PCR (Figure S3b). The reestablishment of *ATSN1* and *soloLTR* element silencing is also observed in the complemented lines, as in wildtype plants (Figure S3b). Likewise, asymmetric DNA methylation patterns are restored at the *AtSN1* and *soloLTR* loci as shown using the *Hae*III and *Alu*I methylation sensitive PCR assay (Figure S3c). Collectively, these assays indicate that the mutant phenotypes observed in *nrpb9b-1* are due to the loss of NRPB9b and not due to a second mutation elsewhere in the genome.

NRPB9b is not required for Pol IV-mediated silencing and Pol V-dependent transcription

Next, we tested the specificity of 9th subunit usage by Pol IV using a novel site, *MRD1* that is silenced specifically by Pol IV and do not require Pol V involvement. *MRD1* is not expressed in wildtype (Col-0) and the *nrpe1-11* Pol V catalytic subunit mutant, but is highly expressed in the Pol IV catalytic subunit mutant *nrpd1-3* (Figure 4a). The *nrpb9a-1* and *nrpb9b-1* mutants do not express *MRD1*, suggesting that the RdDM defects seen in *nrpb9b-1* is not Pol IV-specific. Compared to the expression of *soloLTR* and *AtSN1* retroelements, which is derepressed in either *nrpd1-3* or *nrpe1-11*, loss of retroelement silencing is observed in *nrpb9b-1*as reported earlier. Taken together with the *MRD1* expression, which is not affected in in *nrpb9b-1*, the NRPB9b subunit most likely a crucial RdDM component of Pol V and not Pol IV (Figure 4a).

Pol V has recently been shown to produce intergenic transcripts that recruit AGO4 to target DNA methylation (Wierzbicki et al. 2008; Wierzbicki et al. 2009). In the *nrpb9a-1* and *nrpb9b-1* mutants, Pol V-dependent transcription of *IGN5* is not lost while *IGN5* transcription is lost in the catalytic *nrpe1-11* mutant of Pol V largest subunit (Figure 4b). Pol V associates with NRPB9a or NRPB9b from mass spectrometry analysis (Ream et al. 2009), and is likely able to transcribe non-coding RNA at these loci with either of the two 9th subunits.

The observations that 24nt siRNA biogenesis and Pol IV-specific silencing of *MRD1* are not affected in the *nrpb9b-1* mutant suggests that the RdDM defects are not likely due to defects in Pol IV. The transcription of *IGN5* in both *nrpb9a-1* as well as *nrpb9b-1* mutants shows that non-coding Pol V transcription can be made by NRPB9a- or NRPB9b-associated Pol V. Therefore, the defects in establishing RdDM in *nrpb9b-1*

points to a distinct role of NRPB9b-associated Pol V in the recruitment of effector step proteins.

Discussion

In this report, we provided evidence that NRPB9a and NRPB9b are redundant with respect to Pol II functions, but not redundant with respect to RNA silencing involving NRPB9b-associated Pol V. We deduce that Pol II does not appear to discriminate between NRPB9a or NRPB9b for mRNA synthesis, and either protein is sufficient for viability, whereas the double mutant is lethal. However, the altered leaf morphology in *nrpb9b-1* suggests that the two variant forms of NRPB9 are not fully redundant, as other Pol IV and V subunit mutants do not exhibit any morphological phenotypes, suggesting that any phenotypes observed are due to Pol II dysfunction (Onodera et al. 2005; He et al. 2009; Ream et al. 2009).

We observe that the transmission of the doubly mutant *nrpb9a-1 nrpb9b-1* gametes occur at much at much lower frequencies in the male than the female. This observation is in contrast to mutant alleles for Pol I, II or III essential subunits, which show zero transmission through the ovules due to female gametophyte lethality prior to fertilization (Onodera et al. 2008). In contrast, the doubly mutant *nrpb9a-1 nrpb9b-1* gametophytes develop and fertilization occurs, but the embryos fail to develop fully. The male gametophytes of Pol I, II or III catalytic subunit mutants are viable but mutant alleles have reduced transmission rates due to the failure of pollen tubes to reach the most distant ovules. The *nrpb9a-1 nrpb9b-1* double mutant pollen may have similarly reduced vigor (Onodera et al. 2008). The fact that NRPB9a and NRPB9b are not as essential as

Pol I, II or III catalytic subunits for female gametophyte development suggests that they may be partially dispensable for Pol II function, similar to yeast (Woychik et al. 1991; Onodera et al. 2008). In fact, as we have observed, the homozygous *nrpb9a-1 nrpb9b-1* double mutant embryos are able to develop up to the globular stage before development stops, which is not a feature of the null mutants of Pol I, II or III subunits. Instead, the NRPB9 null mutants phenocopy hypomorphic, but not null, mutants of Pol II catalytic catalytic subunits (Onodera et al. 2008).

Comparing 5S rRNA array methylation at symmetric CG dinucleotides exposes a clear role for Pol IV and V catalytic subunits in maintaining CG methylation at these loci. However, in mutant non-catalytic subunits of Pol IV and V, CG methylation is not lost as dramatically as catalytic subunit mutants. Pol IV dependent siRNA accumulation is not affected in either *nrpb9a-1* or *nrpb9b-1* mutants and suggests that either NRPB9a- or NRPB9b-associated Pol IV are functional in terms of producing precursor siRNA transcripts. Likewise, Pol V catalytic subunit mutant siRNA accumulation is reduced, yet *nrpb9b-1* does not appear to have similar defects to *nrpe1-11*. So, the RdDM defects in *nrpb9b-1* appear to uncouple siRNA biogenesis from other functions of Pol IV or Pol V. Other non-catalytic subunit mutants such as *nrpd4/e4* and *nrpe5* also disrupt silencing without abolishing or severely reducing siRNA biogenesis (He et al. 2009; Ream et al. 2009). Collectively, these results suggest that non-catalytic subunits may be dispensable for Pol IV or V transcription, but essential for interactions with chromatin modifying activities that interact with Pol V to bring about silencing.

NRPB9a and NRPB9b differences are mapped to surface residues

How do the different amino acid sequences between NRPB9a and NRPB9b translate to their functional architecture? Alignment of the amino acid sequences of NRPB9a and NRPB9b shows that the two proteins are very similar (Figure 7a). Only 8 out of the 114 amino acids are different between NRPB9a and NRPB9b, which are 93% identical and 97% similar overall. Based on a multiple alignment with yeast RPB9, the eight amino acid residues that are different on NRPB9a and NRPB9b can be superimposed on the RPB9 subunit from the PDB:1Y1W crystal structure, which is a 12 subunit Pol II elongation complex from yeast (Kettenberger et al. 2004). Amino acids highlighted in red correspond to orthologous amino acids that differ between NRPB9a and NRPB9b (Figure 7b).

RPB9 has two zinc finger domains referred to as Zn1 and Zn2, in the N-terminal and C-terminal domain. These Zn fingers are separated by a linker region. The Zn2 domain shares some homology with the zinc finger domain of TFIIS and is thought to facilitate efficient transcription elongation in concert with TFIIS (Hemming and Edwards 2000). The 9th largest subunits from Pol I and III possess stronger intrinsic transcript cleavage activities, compared to RPB9 from Pol II and are not dependent on TFIIS (Ruan et al. 2011). RPB9 is thought to have lost this intrinsic cleavage capability in the Zn2 region in order to mediate other Pol II-specific functions, such as promoter gene regulation, 3'-processing of RNAs and transcription termination (Ruan et al. 2011). It is intriguing that the majority of the amino acid differences between NRPB9a and NPRB9b appear to be within the Zn2 domain, in particular amino acids 77, 82 and 109. We are currently developing tools to test if these amino acids confer functional differences

between NRPB9a and NRPB9b with regards to RdDM. In addition, the general differences between NRPB9a and NRPB9b appear to be among the amino acids that are exposed on the outer surface of the protein. We speculate that amino acid differences between NRPB9a and NRPB9b mediate different interactions with other proteins.

Affinity purification of the DDR complex, consisting of DRD1, DMS3 and a novel methyl-DNA binding protein RDM1, resulted in the co-purification of Pol V that contains NRPB9a, but not NRPB9b (Kanno et al. 2004; Kanno et al. 2008; Gao et al. 2010; Law et al. 2010). Therefore, it is possible that although Pol V can associate with both NRPB9a and NRPB9b, perhaps only the NRPB9b-associated form functions in RNA-directed DNA methylation.

Based on our observations, we conclude that only one form of Pol V is functional for RdDM, the NRPB9b-associated form. The role of NRPB9a and NRPB9b are redundant in terms of Pol IV because siRNA accumulation and the silencing of a Pol IV-specific locus were not affected. The fact that silencing is disrupted in *nrpb9b-1* and has no consequence on Pol V-dependent transcription supports the hypothesis that Pol V functions in RNA silencing include events that are separable from non-coding RNA transcription.

Figures legends

Figure 1 Phylogenetic tree of Pol I, II, III, IV and V subunits homologous to the 9th largest subunit of yeast RNA polymerases II, RPB9

Proteins homologous to yeast RPA12 (Pol I), RPB9 (Pol II) and RPC11 (Pol III) in human, fly, zebrafish, *Chlamydomonas*, *Arabidopsis*, poplar, maize and rice were subjected to multiple alignment using MUSCLE (Edgar 2004) under standard parameters and imported into Geneious 5.3.6 (http://www.geneious.com/) to display the phylogenetic tree shown here. Bootstrap values are indicated at each branchpoint. Most organisms encode single genes for Pol I, II and III 9th largest subunits. However, the RPB9 subunit of Pol II is represented by multiple genes in plant species. The 9th largest subunits from Pol I, II and III form distinct clades, with the 9th largest subunits that are associated with Pol IV and V also associating with Pol II. The aligned amino acid sequences can be found in Figure S1.

Figure 2 Single mutants of *nrpb9a-1* and *nrpb9b-1* are viable but the double mutant is lethal

(A) Positions of the T-DNA insertions in the *nrpb9a-1* and *nprb9b-1* alleles are indicated with a triangle. Filled boxes are exons and lines represent introns.

(B) Homozygous recessive mutants of *nrpb9a-1* and *nrpb9b-1* are viable. *nrpb9b-1* mutant plants exhibit altered morphological phenotypes, including leaves that have shorter petioles, are more ovate in shape and have less edge curling compared to *nrpb9a-1* or wildtype plants (ecotype Col-0).

(C) Seeds developing in a silique in which 25% of embryos should be *nrpb9a-1 nrpb9b-1* double mutants. The parent plant genotype was homozygous for *nrpb9a-1* and heterozygous for *nrpb9b-1*. Blue arrows point to fully developed seeds with properly developed cotyledons. Red arrows point to embryos that are arrested at the globular stage and are presumed to be the *nrpb9a-1 nrpb9b-1* double mutants that do not survive to yield viable progeny.

(D) Homozygous *nrpb9a-1 nrpb9b-1* double mutants are not recovered among the progeny of plants that are homozygous for the *nrpb9a-1* or *nrpb9b-1* alleles and heterozygous for the other allele. Heterozygotes are also under-represented, suggesting reduced transmission of mutant alleles via the male or female gametophytes that are doubly mutant.

Figure 3 The *nrpb9b-1* mutant is defective in RNA silencing and DNA methylation

(A) Semi quantitative RT-PCR analysis of *soloLTR* and *AtSN1* retroelement expression in wildtype (Col-0) and Pol IV and V mutants. Catalytic subunits of Pol IV and V, *nrpd1-3* and *nrpe1-11* show expression of *soloLTR* and *AtSN1* elements, as does *nrpb9b-1*. However, in *nrpb9a-1* both *soloLTR* and *AtSN1* retorelements remain silenced, as in wild-type plants.

(B) Retrotransposon methylation was assayed by PCR using genomic DNA after digestion with methylation sensitive restriction enzymes. Decreased asymmetric CHH methylation in *nrpd1-3, nrpe1-11,* and *nrpb9b-1* at *soloLTR* and *AtSN1* loci prevents PCR amplification. Methylation of these retroelements in *nrpb9a-1* is similar to wildtype (Col-0). A control locus, *At2g19920,* lacks *Hae*III sites. The

cartoon below shows relative restriction enzyme sites on the amplicon, indicated by two inward-facing arrows.

(C) Southern blot analysis of the 5S rRNA gene array from genomic DNA digested with methylation sensitive enzymes *Hpa*II (CG sites) or *Hae*III (CHH sites). Loss of CG methylation is indicated by higher mobility bands at the bottom of the gel, as observed in *nrpd1-3* and *nrpe1-11*, which are the catalytic subunit mutants of Pol IV and V. The *nrpb9a-1* and *nrpb9b-1* mutants do not show the same degree of CG methylation loss compared to Pol IV and V catalytic *nrpd1-3* and *nrpe1-11* mutants. CHH methylation assayed using *Hae*III is also significantly reduced in *nrpd1-3, nrpe1-1* and *nrpb9b-1* subunit mutants, except for *nrpb9a-1* which exhibits wildtype (Col-0) methylation patterns.

(D) Small RNA abundance in Pol IV and V mutants are assayed using RNA blots. The Pol IV largest subunit mutant, *nrpd1-3* has complete loss of siR1003 and *soloLTR* 24nt siRNAs. The Pol V largest subunit mutant, *nrpe1-11* has significant loss of siR1003 siRNAs, whereas the *nrpb9a-1* and *nrpb9b-1* mutants do not show significant changes in siR1003 or soloLTR 24nt siRNA abundance. The microRNA miR160, which is not affected by Pol IV and V mutants, serves as a loading control.

Figure 4 NRPB9b is not required for Pol V-dependent transcription and Pol IV-mediated silencing

(A) *MRD1* transcription assayed using RT-PCR shows no expression in wildtype (Col-0) and the *nrpe1-11* catalytic Pol V mutant. *MRD1* is overexpressed in

nrpd1-3, when Pol IV is catalytically inactive. The *nrpb9a-1* and *nrpb9b-1* mutants do not express *MRD1*. Retroelements *soloLTR* and *AtSN1* are usually silent in wildtype (Col-0) and derepressed by either Pol IV or Pol V subunit mutations including *nrpd1-3, nrpe1-11* and *nrpb9b-1*.

(B) RT-PCR analysis of the Pol V-dependent *IGN5* non-coding RNA shows expression in wildtype (Col-0) and *nrpd1-3* but not *nrpe1-11* as expected. The *IGN5* transcripts are present in *nrpb9a-1* and *nrpb9b-1* homozygous mutants.

Figure 5 Amino acid sequence differences between NRPB9a and NRPB9b and their predicted positions on the crystal structure of yeast RPB9

(A) Aligned NRPB9a and NRPB9b protein sequences, with their specific amino acid differences highlighted. The two proteins are 93% identical and 97% similar.

(B) The eight amino acid differences between NRPB9a and NRPB9b are mapped onto RPB9 (in green), in a space-filling rendering of the yeast Pol II crystal structure, PDB:1Y1W, using PyMOL (Version 1.2r1, http://www.pymol.org/). The red amino acids highlighted are positions corresponding to the eight amino acids that are different between NRPB9a and NRPB9b, numbered according to their amino acid positions in NRPB9a and NRPB9b. The RPB1 subunit is shown in grey, RPB2 in blue, RPB5 in gold and the DNA duplex in pink.

Figure S1 MUSCLE sequence alignment for the phylogenetic tree of RNAP 9th largest subunits shown in Figure 1

The three classes of the 9th largest subunits of Pol I, II and III are indicated. The RPB variants present in Pol IV or V are part of the Pol II clade.

Figure S2 Test cross for *nrpb9a-1 nrpb9b-1* double mutant allele transmission

Test crosses performed to measure the transmission of *nrpb9a-1 nrpb9b-1* double mutant alelle through female or male gametophytes. Transmission of doubly homozygous *nrpb9a-1 nrpb9b-1* gametes are reduced via the female or male (pollen donor), below the expected frequency.

Figure S3 Complementation of the *nrpb9b-1* mutant

(A) A transgene containing genomic *NRPB9b* under control of its native promoter rescues *nrpb9b-1* and restores the morphological phenotype to wild-type.

(B) In *nrpb9b-1* lines expressing the *NRPB9b* transgene, silencing of *AtSN1* and *soloLTR* retrotransposons are re-established.

(C) The *NRPB9b* transgene also restores asymmetric DNA methylation patterns of *soloLTR* and *AtSN1* in the *nrpb9b-1* genetic background.

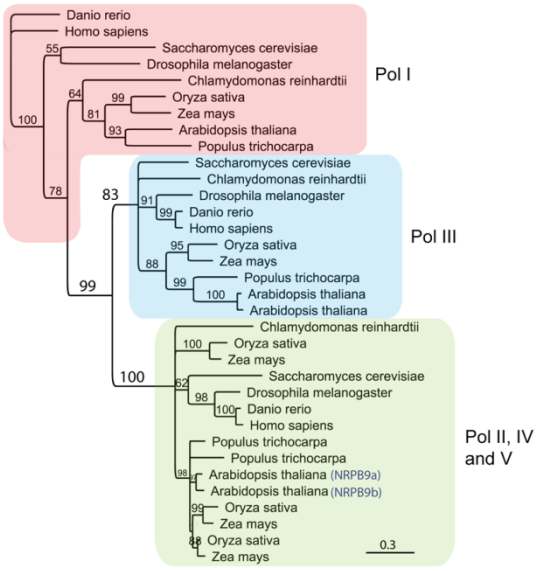
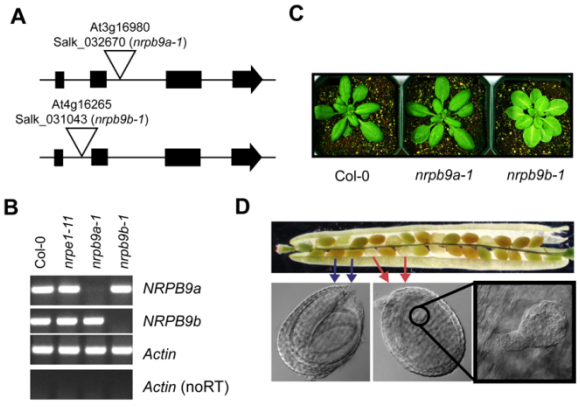

Danio rerio
Homo sapiens
Saccharomyces cerevisiae
Drosophila melanogaster
Chlamydomonas reinhardtii
Oryza sativa
Zea mays
Arabidopsis thaliana
Populus trichocarpa
Pol I
Saccharomyces cerevisiae
Chlamydomonas reinhardtii
Drosophila melanogaster
Danio rerio
Homo sapiens
Oryza sativa
Zea mays
Populus trichocarpa
Arabidopsis thaliana
Arabidopsis thaliana
Pol III
Chlamydomonas reinhardtii
Oryza sativa
Zea mays
Saccharomyces cerevisiae
Drosophila melanogaster
Danio rerio
Homo sapiens
Populus trichocarpa
Populus trichocarpa
Arabidopsis thaliana (NRPB9a)
Arabidopsis thaliana (NRPB9b)
Oryza sativa
Zea mays
Oryza sativa
Zea mays
0.3
Pol II, IV
and V

Figure 1



E

Parent genotype	Progeny genotypes and their frequency		
9a -/- , 9b +/-	9a -/- , 9b -/-	9a -/- , 9b +/-	9a -/- , 9b +/+
	0	69	60

Parent genotype	Progeny genotypes and their frequency		
9a +/- , 9b -/-	9a -/- , 9b -/-	9a +/- , 9b -/-	9a +/+ , 9b -/-
	0	61	68

Figure 2

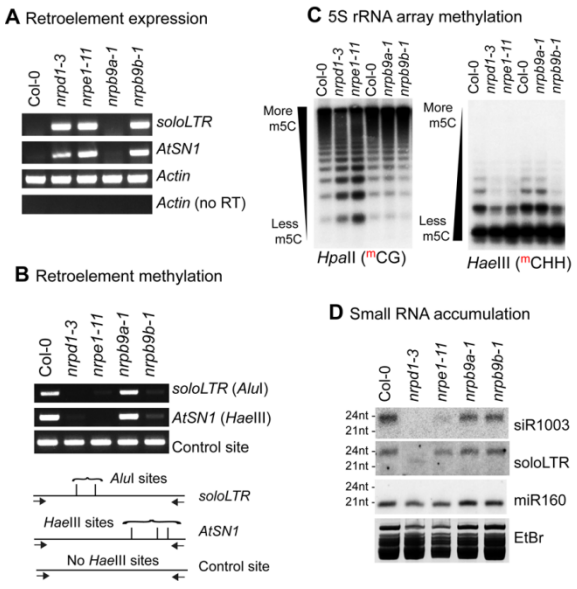

A Retroelement expression
Col-0
nrpd1-3
nrpe1-11
nrpb9a-1
nrpb9b-1
soloLTR
AtSN1
Actin
Actin (no RT)
B Retroelement methylation
soloLTR (AluI)
AtSN1 (HaeIII)
Control site
AluI sites
soloLTR
HaeIII sites
AtSN1
No HaeIII sites
Control site
C 5S rRNA array methylation
More m5C
Less m5C
HpaII (mCG)
HaeIII (mCHH)
D Small RNA accumulation
24nt
21nt
siR1003
soloLTR
miR160
EtBr

Figure 3

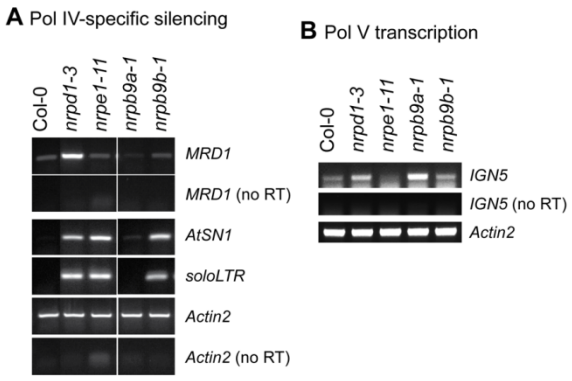

A Pol IV-specific silencing
Col-0
nrpd1-3
nrpe1-11
nrpb9a-1
nrpb9b-1
MRD1
MRD1 (no RT)
AtSN1
soloLTR
Actin2
Actin2 (no RT)
B Pol V transcription
Col-0
nrpd1-3
nrpe1-11
nrpb9a-1
nrpb9b-1
IGN5
IGN5 (no RT)
Actin2

Figure 4

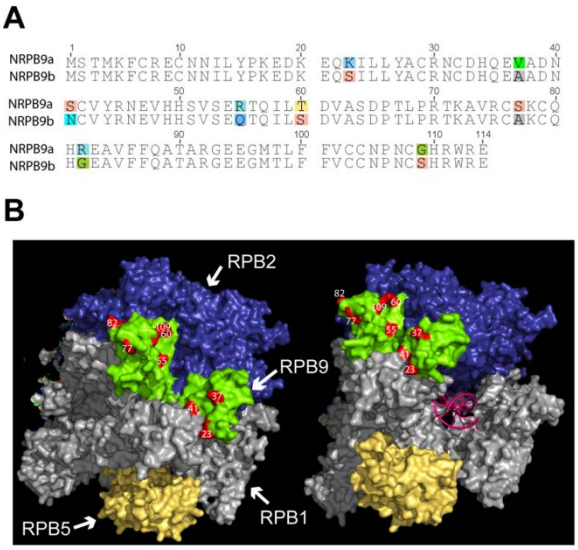

A
NRPB9a MSTMKFCRECNNILYPKEDK EQKILLYACRNCDHQEVADN
NRPB9b MSTMKFCRECNNILYPKEDK EQSILLYACRNCDHQEAADN
NRPB9a SCVYRNEVHHSVSERTQILT DVASDPTLPRTKAVRCSKCQ
NRPB9b NCVYRNEVHHSVSEQTQILS DVASDPTLPRTKAVRCAKCQ
NRPB9a HREAVFFQATARGEEGMTLF FVCCNPNCGHRWRE
NRPB9b HGEAVFFQATARGEEGMTLF FVCCNPNCSHRWRE
B
RPB2
RPB9
RPB5
RPB1

Figure 5

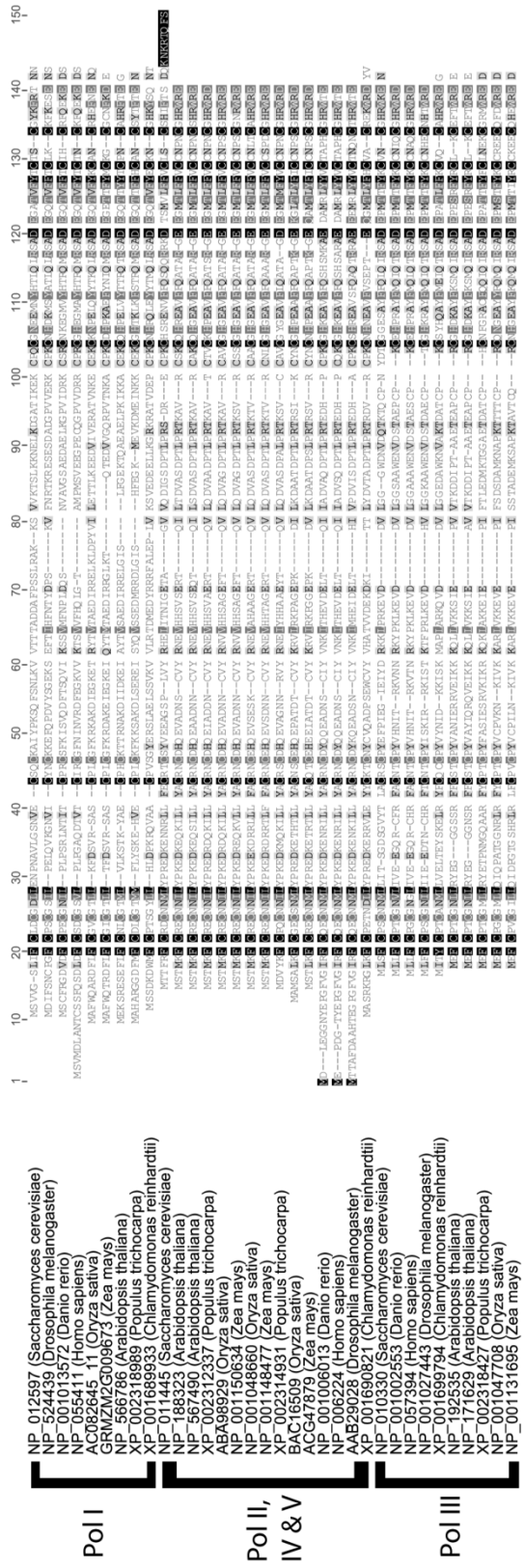

Pol I
Pol II, IV & V
Pol III
NP_012597 (Saccharomyces cerevisiae)
NP_524439 (Drosophila melanogaster)
NP_001013572 (Danio rerio)
NP_055411 (Homo sapiens)
AC082645_11 (Oryza sativa)
GRMZM2G009673 (Zea mays)
NP_566786 (Arabidopsis thaliana)
XP_002318989 (Populus trichocarpa)
XP_001689933 (Chlamydomonas reinhardtii)
NP_011445 (Saccharomyces cerevisiae)
NP_188323 (Arabidopsis thaliana)
NP_567490 (Arabidopsis thaliana)
XP_002312337 (Populus trichocarpa)
ABA98929 (Oryza sativa)
NP_001150634 (Zea mays)
NP_001048660 (Oryza sativa)
NP_001148477 (Zea mays)
XP_002314931 (Populus trichocarpa)
BAC16509 (Oryza sativa)
ACG47879 (Zea mays)
NP_001006013 (Danio rerio)
NP_006224 (Homo sapiens)
AAB29028 (Drosophila melanogaster)
XP_001690821 (Chlamydomonas reinhardtii)
NP_010330 (Saccharomyces cerevisiae)
NP_001002553 (Danio rerio)
NP_057394 (Homo sapiens)
NP_001027443 (Drosophila melanogaster)
XP_001699794 (Chlamydomonas reinhardtii)
NP_192535 (Arabidopsis thaliana)
NP_171629 (Arabidopsis thaliana)
XP_002318427 (Populus trichocarpa)
NP_001047708 (Oryza sativa)
NP_001131695 (Zea mays)

Figure S1

Parental genotypes of test crosses		F1 progeny genotypes and their frequency	
♀ (female)	♂ (male)	% carrying *9a-1* T-DNA [expect 100%]	% carrying *9b-1* T-DNA [expect 50%]
9a -/- , 9b +/-	*+/+, +/+*	100%	41.2% (35/85)
+/+, +/+	*9a -/- , 9b +/-*	100%	34.9% (30/86)

Parental genotypes of test crosses		F1 progeny genotypes and their frequency	
♀ (female)	♂ (male)	% carrying *9a-1* T-DNA [expect 50%]	% carrying *9b-1* T-DNA [expect 100%]
9a +/- , 9b -/-	*+/+, +/+*	39.8% (35/88)	100%
+/+, +/+	*9a +/- , 9b -/-*	32.2% (28/87)	100%

Figure S2

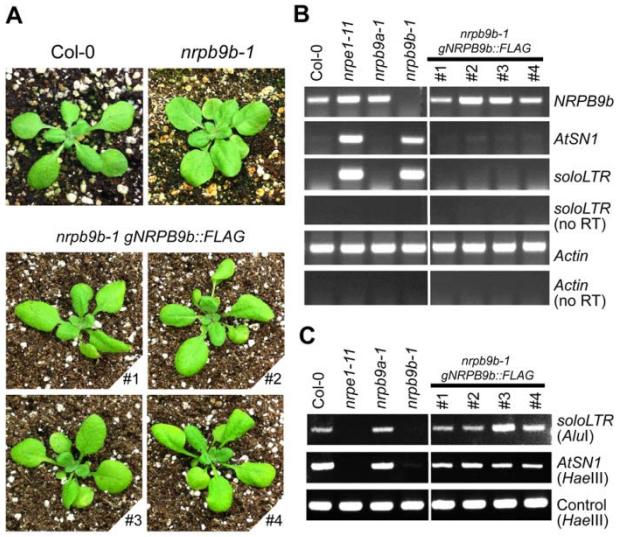

A
Col-0
nrpb9b-1
nrpb9b-1 gNRPB9b::FLAG
#1
#2
#3
#4
B
Col-0
nrpe1-11
nrpb9a-1
nrpb9b-1
nrpb9b-1 gNRPB9b::FLAG
#1
#2
#3
#4
NRPB9b
AtSN1
soloLTR
soloLTR (no RT)
Actin
Actin (no RT)
C
Col-0
nrpe1-11
nrpb9a-1
nrpb9b-1
nrpb9b-1 gNRPB9b::FLAG
#1
#2
#3
#4
soloLTR (AluI)
AtSN1 (HaeIII)
Control (HaeIII)

Figure S3

References

Awrey DE, Weilbaecher RG, Hemming SA, Orlicky SM, Kane CM, Edwards AM. 1997. Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem* 272: 14747-14754.

Blevins T. 2010. Northern blotting techniques for small RNAs. *Methods Mol Biol* 631: 87-107.

Blevins T, Pontes O, Pikaard CS, Meins F, Jr. 2009. Heterochromatic siRNAs and DDM1 independently silence aberrant 5S rDNA transcripts in Arabidopsis. *PLoS One* 4: e5932.

Chen X, Ruggiero C, Li S. 2007. Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage. *Mol Cell Biol* 27: 4617-4625.

Cramer P, Armache KJ, Baumli S, Benkert S, Brueckner F, Buchen C, Damsma GE, Dengl S, Geiger SR, Jasiak AJ et al. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* 37: 337-352.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288: 640-649.

Darst SA, Polyakov A, Richter C, 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.

Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song K, Pikaard CS. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* 45: 616-629.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792-1797.

Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S et al. 2010. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* 465: 106-109.

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

He XJ, Hsu YF, Pontes O, Zhu J, Lu J, Bressan RA, Pikaard C, Wang CS, Zhu JK. 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: 318-330.

Hemming SA, Edwards AM. 2000. Yeast RNA polymerase II subunit RPB9. Mapping of domains required for transcription elongation. *J Biol Chem* 275: 2288-2294.

Hemming SA, Jansma DB, Macgregor PF, Goryachev A, Friesen JD, Edwards AM. 2000. RNA polymerase II subunit Rpb9 regulates transcription elongation in vivo. *J Biol Chem* 275: 35506-35511.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. 2005. RNA polymerase IV directs silencing of endogenous DNA. *Science* 308: 118-120.

Hirata A, Klein BJ, Murakami KS. 2008. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451: 851-854.

Hirata A, Murakami KS. 2009. Archaeal RNA polymerase. *Curr Opin Struct Biol* 19: 724-731.

Huang L, Jones AM, Searle I, Patel K, Vogler H, Hubner NC, Baulcombe DC. 2009. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. *Nat Struct Mol Biol* 16: 91-93.

Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M. 2006. Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis. *EMBO J* 25: 2828-2836.

Hull MW, McKune K, Woychik NA. 1995. RNA polymerase II subunit RPB9 is required for accurate start site selection. *Genes Dev* 9: 481-490.

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

Kanno T, Huettel B, Mette MF, Aufsatz W, Jaligot E, Daxinger L, Kreil DP, Matzke M, Matzke AJ. 2005. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* 37: 761-765.

Kanno T, Mette MF, Kreil DP, Aufsatz W, Matzke M, Matzke AJ. 2004. Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation. *Curr Biol* 14: 801-805.

Kettenberger H, Armache KJ, Cramer P. 2004. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* 16: 955-965.

Koyama H, Ito T, Nakanishi T, Sekimizu K. 2007. Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes Cells* 12: 547-559.

Kusser AG, Bertero MG, Naji S, Becker T, Thomm M, Beckmann R, Cramer P. 2008. Structure of an archaeal RNA polymerase. *J Mol Biol* 376: 303-307.

Kwapisz M, Beckouet F, Thuriaux P. 2008. Early evolution of eukaryotic DNA-dependent RNA polymerases. *Trends Genet* 24: 211-215.

Lahmy S, Pontier D, Cavel E, Vega D, El-Shami M, Kanno T, 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: 941-946.

Law JA, Ausin I, Johnson LM, Vashisht AA, Zhu JK, Wohlschlegel JA, Jacobsen SE. 2010. A protein complex required for polymerase V transcripts and RNA-directed DNA methylation in Arabidopsis. *Curr Biol* 20: 951-956.

Li S, Ding B, Chen R, Ruggiero C, Chen X. 2006. Evidence that the transcription elongation function of Rpb9 is involved in transcription-coupled DNA repair in Saccharomyces cerevisiae. *Mol Cell Biol* 26: 9430-9441.

Li S, Smerdon MJ. 2002. Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in Saccharomyces cerevisiae. *EMBO J* 21: 5921-5929.

Luo J, Hall BD. 2007. A multistep process gave rise to RNA polymerase IV of land plants. *J Mol Evol* 64: 101-112.

Maillet I, Buhler JM, Sentenac A, Labarre J. 1999. Rpb4p is necessary for RNA polymerase II activity at high temperature. *J Biol Chem* 274: 22586-22590.

Nesser NK, Peterson DO, Hawley DK. 2006. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity in vivo. *Proc Natl Acad Sci U S A* 103: 3268-3273.

Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. 2005. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell* 120: 613-622.

Onodera Y, Nakagawa K, Haag JR, Pikaard D, Mikami T, Ream T, Ito Y, Pikaard CS. 2008. Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. *Genetics* 180: 207-218.

Pontier D, Yahubyan G, Vega D, Bulski A, Saez-Vasquez J, Hakimi MA, Lerbs-Mache S, Colot V, 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.

Ream TS, Haag JR, Wierzbicki AT, Nicora CD, Norbeck AD, Zhu JK, Hagen G, Guilfoyle TJ, Pasa-Tolic L, Pikaard CS. 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.

Ruan W, Lehmann E, Thomm M, Kostrewa D, Cramer P. 2011. Evolution of two modes of intrinsic RNA polymerase transcript cleavage. *J Biol Chem*.

Sun ZW, Tessmer A, Hampsey M. 1996. Functional interaction between TFIIB and the Rpb9 (Ssu73) subunit of RNA polymerase II in Saccharomyces cerevisiae. *Nucleic Acids Res* 24: 2560-2566.

Tucker SL, Reece J, Ream TS, Pikaard CS. 2011. Evolutionary History of Plant Multisubunit RNA Polymerases IV and V: Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization. *Cold Spring Harb Symp Quant Biol*.

Walmacq C, Kireeva ML, Irvin J, Nedialkov Y, Lubkowska L, Malagon F, Strathern JN, Kashlev M. 2009. Rpb9 subunit controls transcription fidelity by delaying NTP sequestration in RNA polymerase II. *J Biol Chem* 284: 19601-19612.

Werner M, Thuriaux P, Soutourina J. 2009. Structure-function analysis of RNA polymerases I and III. *Curr Opin Struct Biol* 19: 740-745.

Wierzbicki AT, Haag JR, Pikaard CS. 2008. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135: 635-648.

Wierzbicki AT, Ream TS, Haag JR, Pikaard CS. 2009. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat Genet* 41: 630-634.

Woychik NA, Lane WS, Young RA. 1991. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J Biol Chem* 266: 19053-19055.

Woychik NA, Liao SM, Kolodziej PA, Young RA. 1990. Subunits shared by eukaryotic nuclear RNA polymerases. *Genes Dev* 4: 313-323.

CHAPTER THREE

GENETIC ANALYSIS OF *ARABIDOPSIS* 4TH AND 7TH LARGEST SUBUNITS OF POL II, IV AND V: DISCRETE ROLES FOR POL IV AND V-SPECIFIC VARIANTS IN RNA SILENCING

My contributions to this work:

I generated the homozygous *nrpd7-1* line and the double mutant lines used in this study. The majority of the experimental work was performed by me with the exception of the Pol II activity assays, development of the native NRPB4 antibody, and transgenic lines in which the *nrpb4-1* mutant is rescued by the wildtype transgene. These tools were the work of Tom Ream and Jeremy Haag. Todd Blevins made the cross between *nrpd4-1* and *nrpd4/e4-2*. I produced all the figures and wrote the manuscript.

At the time of writing, the *nrpd4/e4-2 nrpe7-1* double mutants have yet to be isolated due to the close linkage between the two genes, estimated at 2 centiMorgans. Out of 230 F2 segregants genotyped, a single crossover event was recently detected. Provided that the homozygous *nrpd4/e4-2 nrpe7-1* double mutants are viable in the F3 generation, we anticipate being able to characterize this double mutant for studies alongside *nrpd4/e4, nrpd7-1* and *nrpe7-1* single mutants as well as the *nrpd4/e4-2 nrpd7-1* and *nrpd7-1 nrpe7-1* double mutants.

Genetic analysis of *Arabidopsis* 4th and 7th largest subunits of Pol II, IV and V: discrete roles for Pol IV and V-specific variants in RNA silencing

Ek Han Tan[1,2], Thomas S. Ream[3], Jeremy R. Haag[1], Todd Blevins[1], Craig S. Pikaard[1*]

[1]Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405

[2]Division of Biology and Biomedical Sciences, Washington University in St Louis, St Louis, Missouri 63110

[3]Department of Biochemistry, University of Wisconsin, Madison, Wisconsin 53706

* To whom correspondence should be addressed

Email addresses:

EHT: ehtan@wustl.edu

CSP: pikaard@indiana.edu

TSR: tream@wisc.edu

JRH: haagj@indiana.edu

TB: toddblev@indiana.edu

Running title: RNA silencing via the 4th and 7th largest subunits of Pol IV and V

Manuscript in early preparation

Abstract

The orthologous 4th and 7th largest subunits of RNA polymerases I, II and III, are heterodimeric complexes that form a stalk-like structure adjacent to the RNA exit channel. In addition to the 4/7 subunits specific to Pol I, II and III, *Arabidopsis thaliana* also encodes 4/7 subunits that are unique to the RNA silencing enzymes, Pol IV and V. Pol IV and V evolved as specialized forms of Pol II, but the canonical Pol II specific NRPB4 (At5g09920) and NRPB7 (At5g59180) subunits do not associate with Pol IV or Pol V. Instead, the 4th largest subunit of Pol IV and V is encoded by the same gene, NRPD4/NRPE4 (At4g15950). Two distinct 7th largest subunits, NRPD7 (At3g22900) and NRPE7 (At4g14660) associate with Pol IV or V, but not with Pol II. As in yeast, a mutant defective for the 4th largest subunit of Pol II, *nrpb4-1*, is viable in *Arabidopsis* and this mutation does not affect RNA directed DNA methylation (RdDM). In contrast, the *nrpd4/e4-2* mutants show a loss of RNA silencing akin to catalytic subunit mutants of Pol IV and V. Interestingly, Pol IV *nrpd7-1* mutants do not exhibit RdDM defects whereas Pol V *nrpe7-1* mutants are deficient for RNA silencing. Collectively, the results suggest that the 4/7 subcomplex of Pol V is essential for RNA silencing but may be less critical for Pol IV functions.

Introduction

The 4th and 7th largest subunits of yeast RNA polymerase II heterodimerize to form a subcomplex that protrudes from the enzyme core as a stalk (Bushnell and Kornberg 2003). A paralagous 4/7 substructure is also apparent in the structures of Pol I and III (Werner, Thuriaux et al. 2009) and is also formed by the orthologous F/E (Rpo4/Rpo7) subcomplex in archaeal polymerases (Hirata, Klein et al. 2008; Werner and Grohmann 2011). Research on yeast RPB4/RPB7 and the archaeal F/E subcomplex has revealed diverse and extensive roles for these subunits in various transcriptional processes from transcriptional initiation to mRNA decay (Choder 2004; Hirtreiter, Grohmann et al. 2010).

The 4th largest subunit of RNA Polymerase II in yeast was first described as a dispensable Pol II subunit because the deletion of the gene RPB4 results in slow growth and is sensitive to low and high temperatures, but is not lethal (Woychik and Young 1989; Maillet, Buhler et al. 1999). In contrast, the deletion of RPB7 is lethal in yeast (McKune, Richards et al. 1993). Purified Pol II from *Δrpb4* strains lacks the RPB7 subunit (Edwards, Kane et al. 1991), providing evidence that RPB4 is required for stable association with RPB7 in Pol II, consistent with crystal structures that show extensive RPB4/RPB7 contact with Pol II (Jensen, Meredith et al. 1998; Meka, Werner et al. 2005);(Kettenberger, Armache et al. 2004). Overexpression of RPB7 alleviates the temperature sensitive growth defects seen in *Δrpb4* strains and suggests that RPB7 can associate with Pol II in the absence of RPB4, albeit inefficiently (Sheffer, Varon et al. 1999; Tan, Li et al. 2000).

Promoter dependent transcription requires the interaction of RPB4 with different transcription factors (Edwards, Kane et al. 1991; Orlicky, Tran et al. 2001; Hirtreiter, Grohmann et al. 2010), including the transcription initiation factor, TFIIF (Chung, Craighead et al. 2003). Evidence that the 4/7 complex interacts with nascent RNA comes from Fluorescence Resonance Energy Transfer (FRET) studies, in which nascent RNAs labeled with donor dyes are able to come into close enough proximity to acceptor dyes on the 4/7 subcomplex to induce FRET, both in yeast and archaea (Chen, Chang et al. 2009; Grohmann, Klose et al. 2010). Mutagenesis studies have implicated RPB7 as the main RNA interacting protein of the 4/7 subcomplex (Meka, Werner et al. 2005; Ujvari and Luse 2006).

Genomewide profiling using RPB4 shows that RPB4 occupancy is similar to other polymerase subunits (Jasiak, Hartmann et al. 2008; Runner, Podolny et al. 2008), suggesting that the RPB4/RPB7 subcomplex does not dissociate during transcription. However, dissociability of the RPB4/RPB7 subcomplex from the rest of the enzyme can occur during Pol II purification from yeast and the 4/7 subcomplex is detectable in the cytoplasm during stress apart from the other ten subunits, an unexpected behavior for subunits of a nuclear RNA polymerase (Farago, Nahari et al. 2003). Recently, RPB4/RPB7 has been reported to play a role in mRNA decay in the cytoplasm (Goler-Baron, Selitrennik et al. 2008). The dissociability of the 4/7 subcomplex has only been observed in yeast, and the purification of archaeal and plant polymerase complexes show that the interaction between the core complexes and their respective 4/7 subcomplexes are stable (Larkin, Hagen et al. 1999; Grohmann, Hirtreiter et al. 2009). Therefore, it is not clear if the 4/7 subcomplex traffics with mRNAs in systems other than yeast.

The RPB4 and RPB7 subunits of *Arabidopsis* Pol II have previously been described as AtRPB19.5 ad AtRPB15.9, respectively (Larkin and Guilfoyle 1998) and form a heterodimer. The RPB7 subunit of *Arabidopsis* Pol II is encoded by the *NRPB7* gene (*At5g15980*) and contains 6 exons and 5 introns. Three intronless *RPB7*-like genes also exist, two of which, *NRPD7* (*At3g22900*) and *NRPE7* (*At4g14660*), associate with Pol IV or Pol V, as shown by mass spectrometry analyses of affinity purified Pol II, IV or V. A third intronless RPB7 homolog, *At4g14520*, is similar to *NRPD7* and *NRPE7* but has not been identified in association with any polymerases thus far (Ream, Haag et al. 2009; Tucker, Reece et al. 2011;Ream and Pikaard, unpublished). Although NRPD7 is only detected in Pol IV, and not Pol V, NRPE7 peptides are found associated with Pol IV at low frequency, suggesting that this protein can sometimes substitute for NRPD7 in the context of Pol IV. Pol V exclusively makes use of NRPE7 as its 7th largest subunit (Ream, Haag et al. 2009).

Evidence for a role of RPB7 in small RNA-induced silencing in *S. pombe* includes the role of RPB7 in precursor siRNA biogenesis (Djupedal, Kos-Braun et al. 2009). In *Arabidopsis*, a forward genetic screen for players in the RNA directed DNA methylation (RdDM) pathway identified NRPD4 (At4g15950), a paralog of NRPB4 (He, Hsu et al. 2009).

Using a reverse genetic approach, we tested 4th and 7th largest subunit mutants for viability and defects in RNA directed DNA methylation. We report that a null mutant of Pol II's 4th largest subunit, *nrpb4-1* is viable, similar to findings in yeast. Homozygous mutants for the 7th largest subunits of Pol IV and V are also viable. Interestingly, *nrpe7-1*

but not *nrpd7-1* is required for RNA silencing, as is the mutant of the binding partner for both of these proteins *nrpd4/e4-2*.

Materials and Methods

Plant materials

T-DNA insertion alleles *nrpb4-1, nrpd/e4-2, nrpd7-1* and *nrpe7-1* were genotyped using primers listed in Table S1. Seed stocks for these lines were ordered from the Arabidopsis Biological Resource Center (Ohio State University), with the exception of *nrpe7-1*, which was generously provided by the Martienssen lab (Cold Spring Harbor Lab).

RT-PCR and DNA methylation sensitive PCR assays

Assays for detecting *AtSN1* and *soloLTR* retrotransposon expression and DNA methylation were previously described in (Wierzbicki, Haag et al. 2008). Total RNA isolated using Trizol (Invitrogen) was DNase treated with RQ1 DNase (Promega), then subjected to reverse transcription using gene specific primers and Superscript III (Invitrogen), followed by PCR using Hotstart Taq (Fermentas). Methylation sensitive PCR assays were performed using 50ng of genomic DNA that had been digested overnight using restriction enzymes (New England Biolabs) as indicated in the figure legends.

Small RNA blot hybridization

Total RNAs were fractionated into high and low molecular weight RNAs using the method described in (Blevins 2010). 6μg of low molecular weight RNA was loaded on each lane and subjected to electrophoresis on a denaturing 12% acrylamide gel,

transferred to Hybond membrane (GE Healthcare) and hybridized to DNA oligos end-labelled using T4 Polynucleotide Kinase (New England Biolabs) and γ^{32}P-ATP at 37°C. The blots were exposed for 24-48 hours to a phosphorimaging screen and imaged on Typhoon 9400 (GE Healthcare) phosphorimager.

Protein immunoprecipitation

4g of leaf tissue was ground in liquid nitrogen and resuspended in 15ml of extraction buffer, following the protocol in (Baumberger and Baulcombe 2005). 50ul anti-FLAG resin (Sigma Aldrich) was incubated with 15ml of the leaf extract for 6-8 hours. The immunoprecipitated FLAG-tagged proteins were eluted using 50ul of 2X SDS-loading buffer at 95°C and run on 4-20% SDS polyacrylamide gels (Lonza), then transferred onto PVDF membranes (Milipore). An anti-FLAG-HRP antibody (Sigma Aldrich) was used for western blotting.

Pol II activity assays

The promoter independent RNA polymerase assay was performed as described in (Saez-Vasquez, Albert et al. 2003). Sheared salmon sperm DNA was incubated with FLAG-immunoprecipitated proteins (described above) along with α-labelled ^{32}P-CTP and unlabelled ATP, UTP and GTP. The transcription reactions were blotted on Whatman paper (3M), washed and assayed for radioactive CTP incorporation using a scintillation counter.

Results and discussion

Pol II 4th largest subunit mutant

Determination of the subunit compositions of Pol II, IV and V in *Arabidopsis* revealed that the 4th and 7th largest subunits utilized by Pol II are distinct from the orthologous subunits of Pol IV and V. We identified T-DNA insertional lines that knock-out each of these subunits and subjected these lines to further analyses. The T-DNA insertion positions in the 4th largest subunit genes are shown in Figure 1a and the 7th largest subunit insertion alleles are diagrammed in Figure 1b.

The *Arabidopsis NRPB4* gene is non-essential, as homozygous *nrpb4-1* mutants are viable. Using a native antibody specific for NRPB4, a band of the predicted size is detected upon immunoblotting wildtype but not *nrpb4-1* protein extracts, suggesting that *nrpb4-1* is a null allele (Figure S1). The homozygous *nrpb4-1* plants are small, have curled leaves but are fertile (Figure 1c). The floral organs from *nrpb4-1* also have defects, usually harboring five petals or more, as opposed to the usual four (Figure 1c).

Complementation of nrpb4-1 rescues Pol II activity

Expression of an N-terminally FLAG-tagged cDNA of *NRPB4* transgene in *nrpb4-1* mutants complements the mutation (Figure 2a). The complemented *nrpb4-1* plants have wildtype morphology and the FLAG-tagged NRPB4 protein can be detected by western blot analysis with an anti-FLAG antibody (Figure 2b). Pol II can be immunoprecipitated via the FLAG:NRPB4 subunit using anti-FLAG resin (Figure 2c), as

shown by the ability of the complex to carry out promoter-independent transcription on sheared DNA templates in an alpha-amanitin dependent fashion (Figure 2d).

Pol IV and Pol IV 4th and 7th largest subunit mutants

The *nrpd4/e4-2* allele has been previously described (He, Hsu et al. 2009). Unlike *nrpb4-1* plants, *nrpd4/e4-2* mutants have wildtype morphology (Figure 1c). A *nrpb4-1 nrpd/e4-2* double mutant was generated and this double mutant is also viable, exhibiting phenotypes similar to the *nrpb4-1* single mutant.

The *nrpd7-1* and *nrpe7-1* result from T-DNA insertions into the intronless *NRPD7* or *NRPE7* genes (Figure 1b). Homozygous mutants of *nrpd4/e4-2, nrpd7-1* or *nrpe7-1* do not have any morphological phenotypes, and resemble wildtype plants (Figure 1d). However, a double mutant of *nrpd/e4-2 nrpd7-1* is smaller than the single mutants, revealing a synthetic phenotype in the absence of both activites, which presumably form a subcomplex of Pol IV (Figure 1d). This synthetic phenotype is also exhibited by plants that are homozygous for *nrpd4/e4-2* and heterozygous for *nrpd7-1* (Figure S2).

NRPD4/E4 and NRPE7 are required for RNA silencing

AtSN1 and *soloLTR* are retrotransposons that are silenced by RNA directed DNA methylation (RdDM). Mutations of proteins involved in RdDM, such as the catalytic subunit mutants of Pol IV and V, result in the accumulation of transcripts from *AtSN1*

and *soloLTR* retroelements. Homozygous single mutants for *nrpd/e4-2, nrpd7-1* and *nrpe7-1* were subjected to RT-PCR to test for *AtSN1* and *soloLTR* expression (Figure 3a). The *nrpd/e4-2* mutant shows high levels of expression of the retroelements, similar to the *nrpb9b-1* mutant. The *nrpe7-1* but not *nrpd7-1* mutants also shows defects in silencing. DNA methylation of *AtSN1* at asymmetric CHH sites is lost in *nrpd/e4-2* and *nrpe7-1*, concomitant with the derepression of this retrotransposon (Figure 3b).

The biogenesis of 24nt siRNAs depends on transcription by Pol IV, as the complete loss of 24nt siRNAs occurs in the catalytic subunit mutants of Pol IV (Herr, Jensen et al. 2005; Onodera, Haag et al. 2005; Mosher, Schwach et al. 2008). Because NRPD7 is the only non-catalytic subunit that is unique to Pol IV (Ream, Haag et al. 2009), we sought to determine the accumulation of Pol IV-dependent siRNAs in the *nrpd7-1* mutant. A small RNA blot with the 4th and 7th largest subunit mutants was probed for siR1003, and shows that *nrpd7-1* and *nrpe7-1* plants still accumulate siRNAs (Figure 3c), unlike *nrpd1-3* mutant plants. The Pol V catalytic mutant *nrpe1-11* accumulates siRNAs at very low levels, as do *nrpd4/e4-2* mutants. The results here show that Pol IV-dependent siRNA biogenesis do not require the NRPD7 subunit, although the substitution of NRPE7 might be possible when NRPD7 is mutated. The siRNA accumulation profile of *nrpe7-1* also does not fit with Pol V catalytic subunit mutants, suggesting that the NRPE7 subunit may have a different function in RdDM that is not dependent on Pol V activity, similar to the findings reported on NRPB9b in Chapter two of this thesis.

Discussion

We attribute the *nrpb4-1* morphological phenotypes to Pol II dysregulation, since it is a Pol II specific subunit and has no RdDM defects. In yeast, *Δrpb4* mutants are slow-growing and stop growing at high or low temperatures (Woychik and Young 1989). A loss of function mutant allele of the 2nd largest subunit of Pol II, *nrpb2-3*, also exhibits pleiotropic leaf phenotypes (Zheng, Wang et al. 2009), but not the floral defects apparent in *nrpb4-1* plants. This is the first evidence for a subunit of RNA polymerase II being implicated in flower development (Krizek and Fletcher 2005). The rescue of leaf and floral phenotypes by an *NRPB4* transgene in *nrpb4-1* suggests that the *nrpb4-1* phenotypes are due to the lack of a 4th subunit in Pol II, resulting in the misregulation or developmental genes, and not due to a second mutation elsewhere. Future studies will examine the activity of Pol II from this mutant, using the promoter independent transcriptional assay. We expect that the activity of Pol II isolated from *nrpb4-1* will be lower than in wildtype.

The synthetic phenotype observed in the homozygous *nrpd4/e4-2 nrpd7-1* double mutants and the discovery that the same synthetic phenotype occurs in the *nrpd4/e4-2* homozygous *nrpd7-1* heterozygous mutant suggests that NRPD7 is haplo-insufficient without NRPD4. In an effort to confirm the synthetic phenotypes, the *nrpd7-1* allele is currently being introgressed into *nrpd4/e4-3* mutant plants bearing an independent null allele of *NRPD4/E4*. If similar synthetic phenotypes are replicated in the *nrpd4/e4-3 nrpd7-1* double mutant, we can be confident that the synthetic phenotype observed is not due to a second site mutation from the *nrpd4/e4-2* and *nrpd7-1* cross. Because single *nrpd4/e4-2* and *nrpd7-1* mutants look like wildtype plants, the phenotypic *nrpd4/e4-2*

nrpd7-1 double mutants and haplo-insufficiency displayed by heterozygous *nrpd7-1* in a homozygous *nrpd4/e4-2* mutant suggests that there might be a gain of function for Pol IV activity as opposed to a loss of function. This also suggests a novel role for the *nrpd7-1* allele outside of Pol IV activity for siRNA biogenesis and RdDM.

Figures Legends

Figure 1 Pol II, IV and V 4th and 7th largest subunit mutants

(A) T-DNA insertion alleles of *nrpb4-1* and *nrpd4/e4-2*. T-DNA insertions within exons in the coding region are shown as an open triangle.

(B) T-DNA insertion alleles of *nrpd7-1* and *nrpe7-1* with T-DNA insertions within the coding region of the single exon genes shown as an open triangle. *NRPB7* and *At4g14520* gene structures are also represented here for comparison.

(C) The *nrpb4-1* mutant phenotypes include small stature, curled leaves and floral defects, whereas *nrpd4/e4-2* resembles a wildtype plant.

(D) Comparison of *nrpd4/e4-2* and *nrpd7-1* single mutants to pleiotropic *nrpd4/e4-2 nrpd7-1* double mutants. All plants shown are the same age.

Figure 2 A NRPB4 transgene rescues the nrpb4-1 null mutants

(A) Complementation of *nrpb4-1* can be rescued by a transgene expressing a FLAG-tagged cDNA copy of *NRPB4.* Individual transformants display wildtype morphologies, unlike the *nrpb4-1* mutants.

(B) Immunoprecipitation of FLAG:NRPB4 show expression of FLAG-tagged NRPB4 in *nrpb4-1*.

(C) Immunoprecipitation of FLAG-tagged NRPB2 and NRPB4, detected with an anti-FLAG antibody.

(D) Scintillation counts for radioactive CTP incorporation from a promoter independent Pol II activity assay.

Figure 3 RNA silencing defects in *nrpd4/e4-2* and *nrpe7-1* but not *nrpd7-1* mutants

(A) Comparison of various non-catalytic subunit mutants of Pol IV and V for retrotranspon reactivation. *AtSN1* and *soloLTR* expression was monitored by RT-PCR with *Actin2* acting as the loading control. Mutants that show expression of retrolements include *nrpb9b-1, nrpd4/e4-2* and *nrpe7-1,* but not *nrpd7-1*. Col-0 ecotype is the wildtype control for *nrpb9b-1, nrpd4/e4-2* and *nrpd7-1* whereas L*er* is the wildtype control for *nrpe7-1*.

(B) *AtSN1* methylation levels in various mutants assayed using methylation sensitive PCR. CHH methylation is decreased in *nrpb9b-1, nrpd4/e4-2* and *nrpe7-1* but not *nrpd7-1*.

(C) Small RNA accumulation in Pol IV and V mutants. siR1003 24nt siRNAs are not detected in *nrpd1-3* and are depleted in *nrpe1-11* and *nrpd4/e4-2*, compared to wildtype. The control locus is miR160a, which is 21nt and not dependent on Pol IV and V. (Note: the intensity of miR160a that is slightly less in the wildtype control, Col-0, suggests that this is underloaded on this blot)

Figure S1 NRPB4 is not detected in the *nrpb4-1* null mutant

Western blot analysis on crude protein extracts of wildtype (Col-0), *nrpb4-1* and *nrpb4-1* rescued by a *FLAG:NRPB4* transgene using a native anti-NRPB4 antibody. A band of predicted size for NRPB4 is detected in the wildtype but not in *nrpb4-1*, while transgene expressing FLAG:NRPB4 is detected as larger band due to the epitope tag. Ponceau S staining is shown as loading control.

Figure S2 Haplo-insufficiency phenotypes associated with homozygous *nrpd/e4-2* lines segregating the *nrpd7-1* mutation

Comparison of morphological phenotypes from wildtype (Col-0), *nrpd4/e4-2*, *nrpd7-1* and the *nrpd4/e4-2 nrpd7-1* double mutants. The *nrpd4/e4-2 nrpd7-1* double mutants have small stature and are identifiable at 50% frequency in a segregating population from a homozygous *nrpd7-1* and heterozygous *nrpd4/e4-2* plant. In contrast, all the progeny of a homozygous *nrpd4/e4-2* and heterozygous *nrpd7-1* plant look like *nrpd4/e4-2 nrpd7-1* double mutants.

Table S1 Primers used for genotyping

The T-DNA insertion alleles associated with the 4^{th} and 7^{th} largest subunit mutants used in this study. Primers used for genotyping are listed here.

References

Baumberger, N. and D. C. Baulcombe (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.

Blevins, T. (2010). "Northern blotting techniques for small RNAs." Methods Mol Biol **631**: 87-107.

Bushnell, D. A. and R. D. Kornberg (2003). "Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription." Proc Natl Acad Sci U S A **100**(12): 6969-6973.

Chen, C. Y., C. C. Chang, et al. (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.

Chung, W. H., J. L. Craighead, et al. (2003). "RNA polymerase II/TFIIF structure and conserved organization of the initiation complex." Mol Cell **12**(4): 1003-1013.

Djupedal, I., I. C. Kos-Braun, et al. (2009). "Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA." EMBO J **28**(24): 3832-3844.

Edwards, A. M., C. M. Kane, et al. (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.

Farago, M., T. Nahari, et al. (2003). "Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress." Mol Biol Cell **14**(7): 2744-2755.

Goler-Baron, V., M. Selitrennik, et al. (2008). "Transcription in the nucleus and mRNA decay in the cytoplasm are coupled processes." Genes Dev **22**(15): 2022-2027.

Grohmann, D., A. Hirtreiter, et al. (2009). "RNAP subunits F/E (RPB4/7) are stably associated with archaeal RNA polymerase: using fluorescence anisotropy to monitor RNAP assembly in vitro." Biochem J **421**(3): 339-343.

Grohmann, D., D. Klose, et al. (2010). "RNA-binding to archaeal RNA polymerase subunits F/E: a DEER and FRET study." J Am Chem Soc **132**(17): 5954-5955.

He, X. J., Y. F. Hsu, et al. (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., M. B. Jensen, et al. (2005). "RNA polymerase IV directs silencing of endogenous DNA." Science **308**(5718): 118-120.

Hirata, A., B. J. Klein, et al. (2008). "The X-ray crystal structure of RNA polymerase from Archaea." Nature **451**(7180): 851-854.

Hirtreiter, A., D. Grohmann, et al. (2010). "Molecular mechanisms of RNA polymerase--the F/E (RPB4/7) complex is required for high processivity in vitro." Nucleic Acids Res **38**(2): 585-596.

Jasiak, A. J., H. Hartmann, et al. (2008). "Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex." J Biol Chem **283**(39): 26423-26427.

Jensen, G. J., G. Meredith, et al. (1998). "Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7." EMBO J **17**(8): 2353-2358.
Kettenberger, H., K. J. Armache, et al. (2004). "Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS." Mol Cell **16**(6): 955-965.
Krizek, B. A. and J. C. Fletcher (2005). "Molecular mechanisms of flower development: an armchair guide." Nat Rev Genet **6**(9): 688-698.
Larkin, R. M. and T. J. Guilfoyle (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., G. Hagen, et al. (1999). "Arabidopsis thaliana RNA polymerase II subunits related to yeast and human RPB5." Gene **231**(1-2): 41-47.
Maillet, I., J. M. Buhler, et al. (1999). "Rpb4p is necessary for RNA polymerase II activity at high temperature." J Biol Chem **274**(32): 22586-22590.
McKune, K., K. L. Richards, et al. (1993). "RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability." Yeast **9**(3): 295-299.
Meka, H., F. Werner, et al. (2005). "Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II." Nucleic Acids Res **33**(19): 6435-6444.
Mosher, R. A., F. Schwach, et al. (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., J. R. Haag, et al. (2005). "Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation." Cell **120**(5): 613-622.
Orlicky, S. M., P. T. Tran, et al. (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.
Ream, T. S., J. R. Haag, et al. (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.
Runner, V. M., V. Podolny, et al. (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., A. C. Albert, et al. (2003). "Purification and transcriptional analysis of RNA polymerase I holoenzymes from broccoli (Brassica oleracea) and frog (Xenopus laevis)." Methods Enzymol **370**: 121-138.
Sheffer, A., M. Varon, et al. (1999). "Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4." Mol Cell Biol **19**(4): 2672-2680.
Tan, Q., X. Li, et al. (2000). "Multiple mechanisms of suppression circumvent transcription defects in an RNA polymerase mutant." Mol Cell Biol **20**(21): 8124-8133.
Tucker, S. L., J. Reece, et al. (2011). "Evolutionary History of Plant Multisubunit RNA Polymerases IV and V: Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization." Cold Spring Harb Symp Quant Biol.

Ujvari, A. and D. S. Luse (2006). "RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit." Nat Struct Mol Biol **13**(1): 49-54.
Werner, F. and D. Grohmann (2011). "Evolution of multisubunit RNA polymerases in the three domains of life." Nat Rev Microbiol **9**(2): 85-98.
Werner, M., P. Thuriaux, et al. (2009). "Structure-function analysis of RNA polymerases I and III." Curr Opin Struct Biol **19**(6): 740-745.
Wierzbicki, A. T., J. R. Haag, et al. (2008). "Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes." Cell **135**(4): 635-648.
Woychik, N. A. and R. A. Young (1989). "RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth." Mol Cell Biol **9**(7): 2854-2859.
Zheng, B., Z. Wang, et al. (2009). "Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis." Genes Dev **23**(24): 2850-2860.

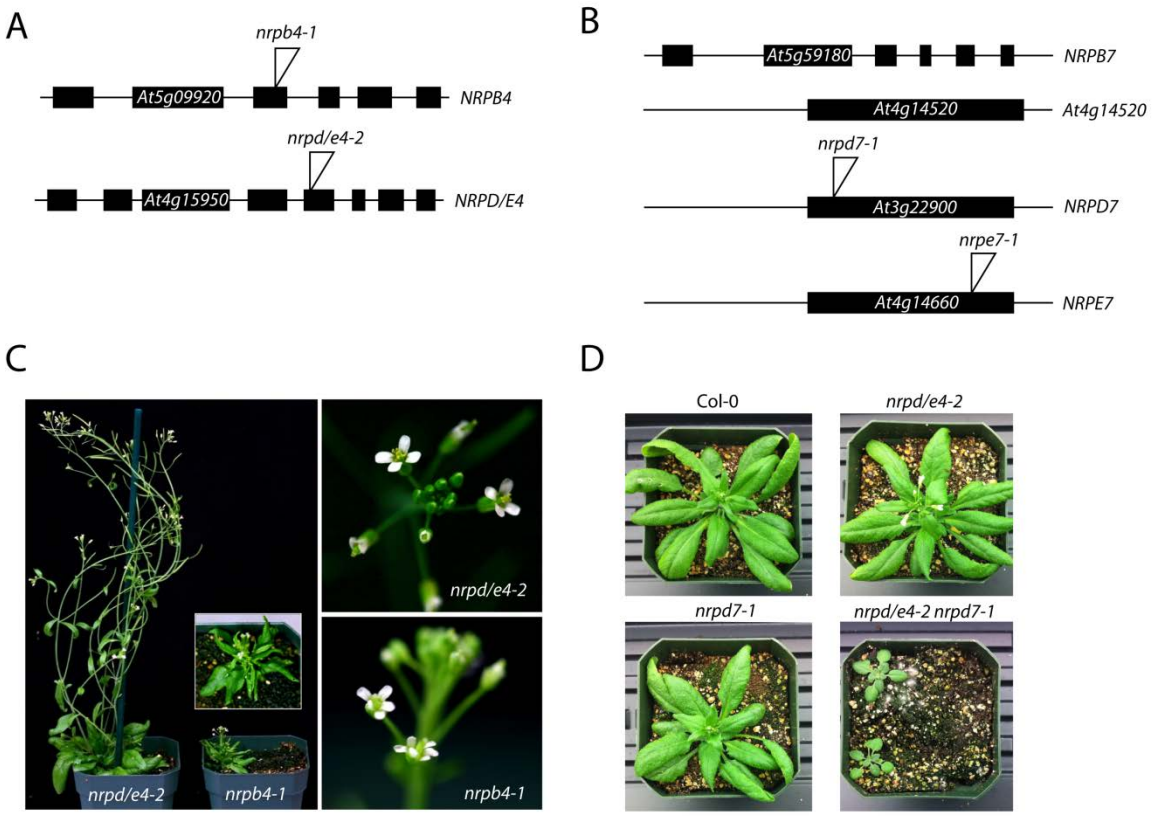

A
nrpb4-1
At5g09920
NRPB4
nrpd/e4-2
At4g15950
NRPD/E4
B
At5g59180
NRPB7
At4g14520
At4g14520
nrpd7-1
At3g22900
NRPD7
nrpe7-1
At4g14660
NRPE7
C
nrpd/e4-2
nrpb4-1
nrpd/e4-2
nrpb4-1
D
Col-0
nrpd/e4-2
nrpd7-1
nrpd/e4-2 nrpd7-1

Figure 1

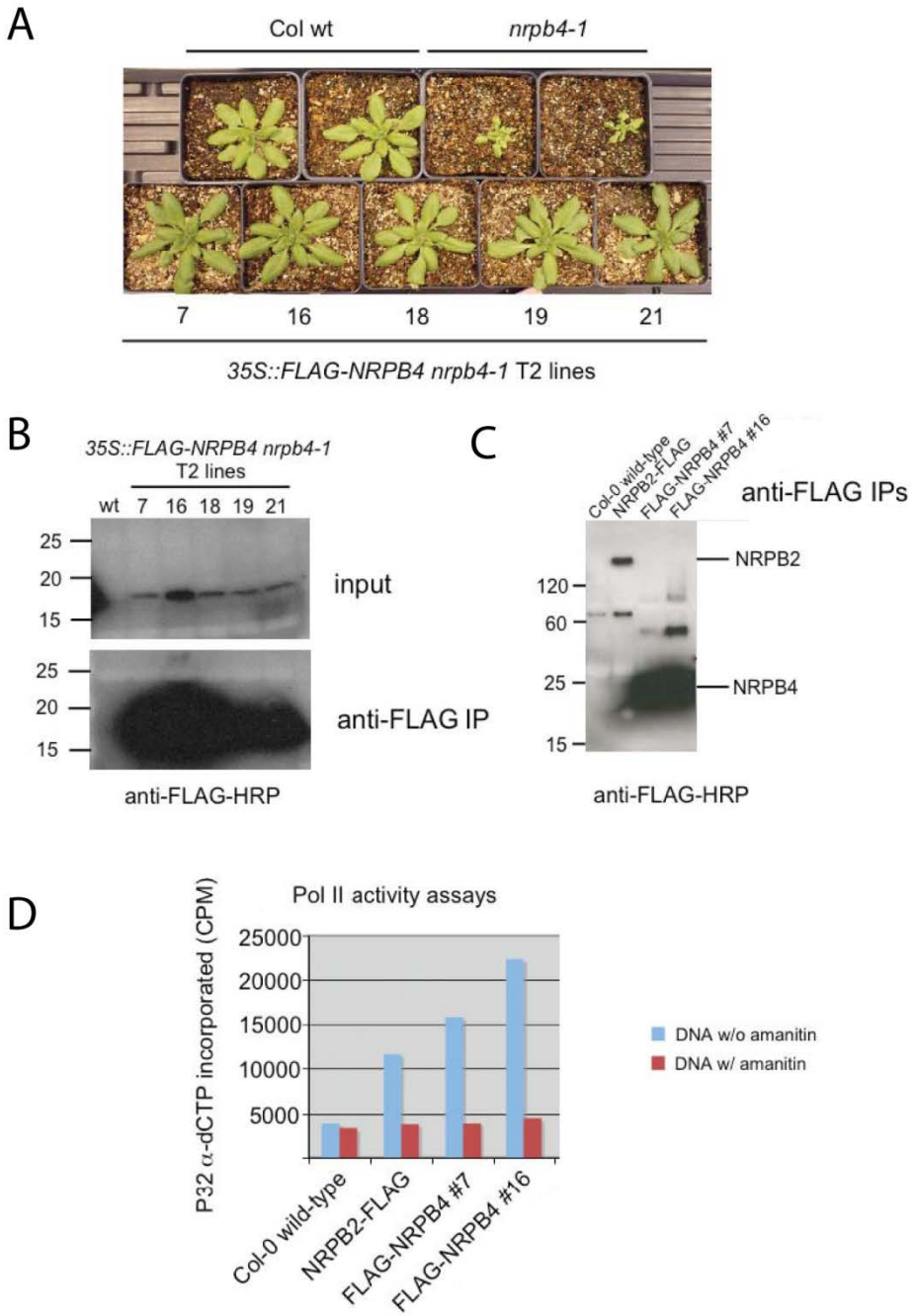

A
Col wt
nrpb4-1
7
16
18
19
21
35S::FLAG-NRPB4 nrpb4-1 T2 lines
B
35S::FLAG-NRPB4 nrpb4-1
T2 lines
wt 7 16 18 19 21
25
20
15
input
anti-FLAG IP
anti-FLAG-HRP
C
Col-0 wild-type
NRPB2-FLAG
FLAG-NRPB4 #7
FLAG-NRPB4 #16
anti-FLAG IPs
NRPB2
NRPB4
120
60
25
15
anti-FLAG-HRP
D
Pol II activity assays
P32 α-dCTP incorporated (CPM)
25000
20000
15000
10000
5000
Col-0 wild-type
NRPB2-FLAG
FLAG-NRPB4 #7
FLAG-NRPB4 #16
DNA w/o amanitin
DNA w/ amanitin

Figure 2

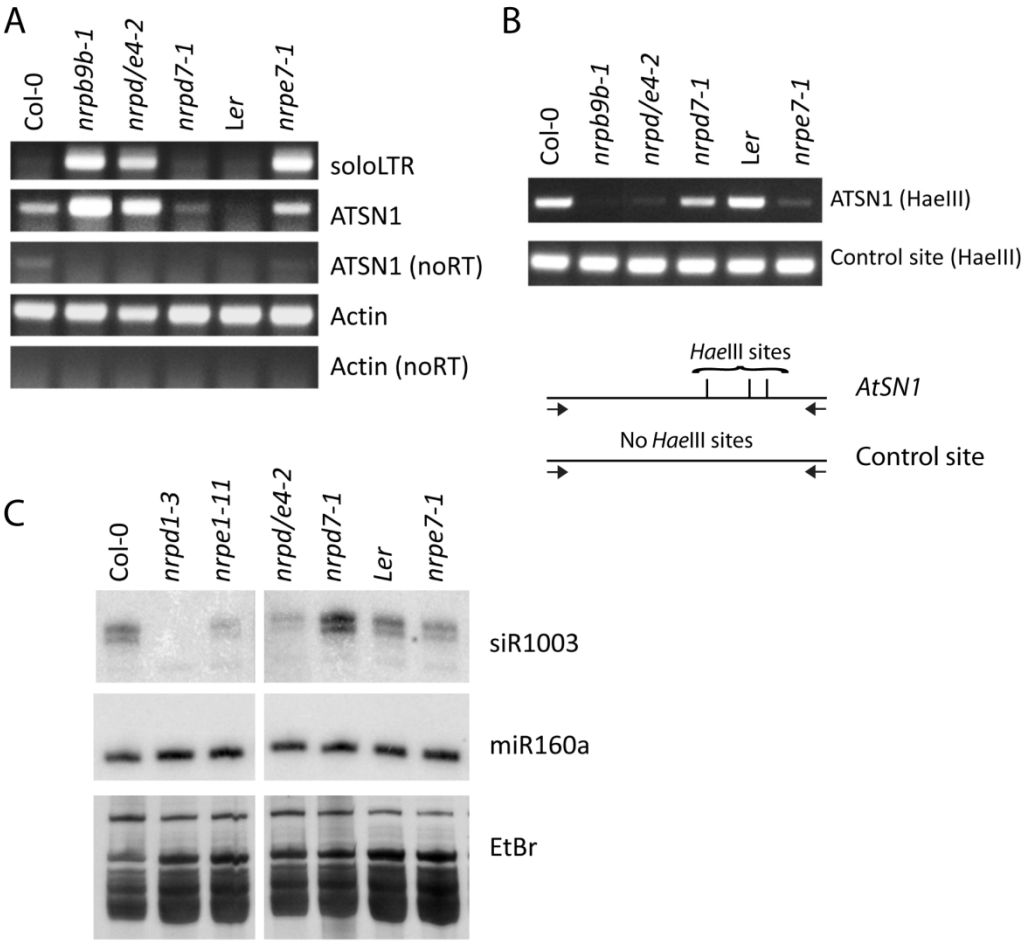

A
Col-0
nrpb9b-1
nrpd/e4-2
nrpd7-1
Ler
nrpe7-1
soloLTR
ATSN1
ATSN1 (noRT)
Actin
Actin (noRT)
B
Col-0
nrpb9b-1
nrpd/e4-2
nrpd7-1
Ler
nrpe7-1
ATSN1 (HaeIII)
Control site (HaeIII)
HaeIII sites
AtSN1
No HaeIII sites
Control site
C
Col-0
nrpd1-3
nrpe1-11
nrpd/e4-2
nrpd7-1
Ler
nrpe7-1
siR1003
miR160a
EtBr

Figure 3

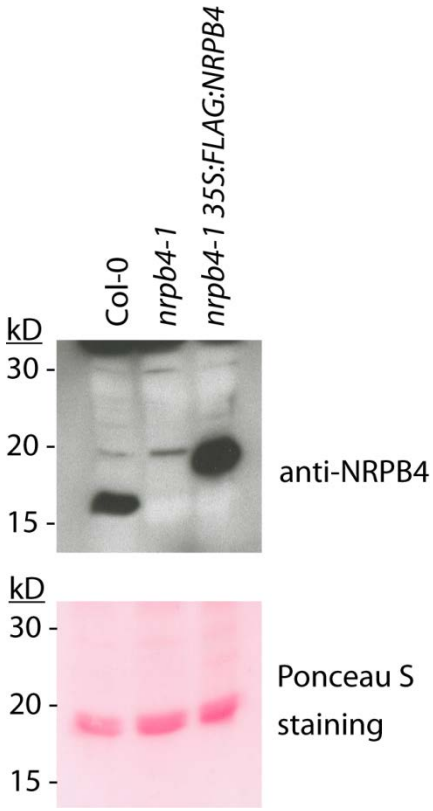

Col-0
nrpb4-1
nrpb4-1 35S:FLAG:NRPB4
kD
30
20
15
anti-NRPB4
kD
30
20
15
Ponceau S
staining

Figure S1

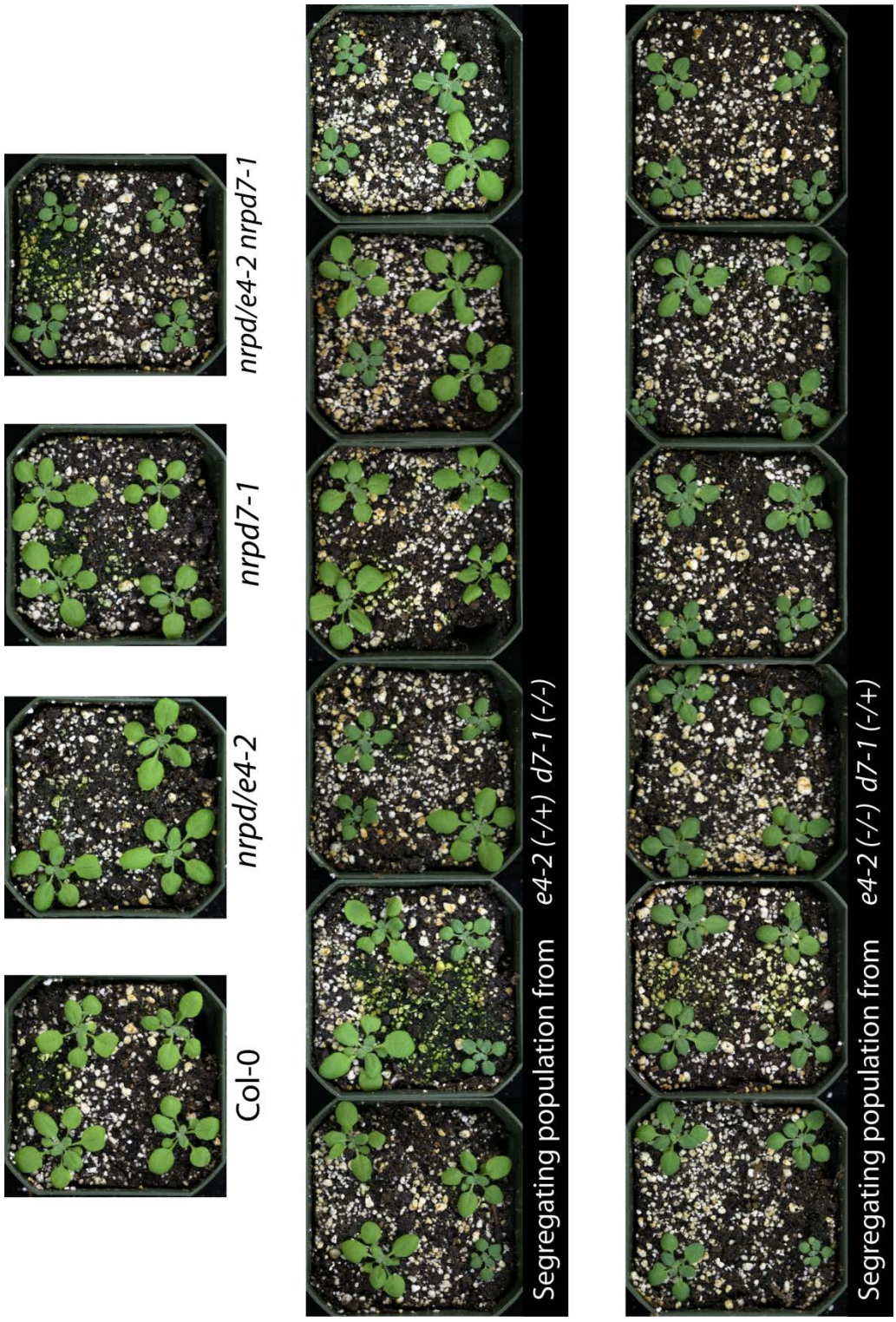

Col-0
nrpd/e4-2
nrpd7-1
nrpd/e4-2 nrpd7-1
Segregating population from e4-2 (-/+) d7-1 (-/-)
Segregating population from e4-2 (-/-) d7-1 (-/+)

Figure S2

Mutant	Gene	Insertion	Ecotype	LP	RP	Left Border
nrpb4-1	At5g09920	Salk_122761	Col-0	TTTCTTTGCCTCATCGAATTG	TTTCAGACGTGGCGAATTAAC	Lba1
nrpe4-2	At4g15950	WiscDslox_476D09	Col-0	TCAGCTTGGATAGCTCCTCCA	TCCTCTCTCAAGTCCAAAGATGGT	p745
nrpd7-1	At3g22900	WiscDsLoxHs_10206G	Col-0	TTTGGGCCTCAAACTTTGTAC	TGATTCATGAAAAACGGGTTC	L4
nrpe7-1	At4g14660	CSHL_ET13987	L*er*	TCGTTTCAGAGAAAAACCTGG	TCTCGATGGTGATGAGTGATG	

Table S1

CHAPTER FOUR

THESIS CONCLUSIONS AND FUTURE DIRECTIONS

i.

CLOSING REMARKS

My studies that are not part of the main thesis chapters, have explored two areas related to the role of DNA methylation in gene silencing. First, site-directed mutagenesis of the de novo DNA methyltransferase DRM2 in plants indicates that DRM2 is an enzymatically active enzyme that is dependent on an invariant catalytic cysteine. This activity is required at the terminal DNA methylation step for the 24nt siRNA directed silencing pathway (Appendix A). In addition, my work in allotetraploid *Arabidopsis suecica,* suggests that DNA demethylation achieved by over-expressing DNA glycosylases or their regulators can prevent the establishment of nucleolar dominance (Appendix A).

For the bulk of my thesis, I have extensively explored the roles of non-catalytic subunits of Pol II, IV and V and their involvement in RNA directed DNA methylation. In particular, the 9th largest subunits of Pol II, IV and V are very similar genes that were initially thought to be functionally redundant, but my work here shows a clear functional distinction between the subunits in RNA silencing (Chapter two). Next, I showed a difference in NRPD7 and NRPE7 usage for RdDM, which are distinct genes from the Pol II 7th largest subunit. Furthermore, genetic analysis shows that like yeast, the 4th largest subunit of Pol II in *Arabidopsis* is a dispensable subunit for viability (Chapter three). Additional studies on the 3rd largest subunits are underway at the time of this writing. In the proceeding section, I will show that the two 6th largest subunits in *Arabidopsis* may be redundant for maintaining DNA methylation via the Pol IV and V pathways. These

analyses of non-catalytic subunits presented here suggest that Pol II, IV and V functions can be dissected to some extent by mutations in catalytic versus non-catalytic subunits. Roles for these highly simiar, paralogous subunits in RNA silencing is particularly intriguing and deserving of further study.

ii.

RNA POLYMERASE II, IV AND V SUBUNITS: A NON-CATALYTIC PERSPECTIVE

Outlook on subunits isolated so far

Mutant	Gene	Insertion	Viability
nrpe1-11	At2g40030	Salk_029919	viable
nrpe2a-2	At3g23780	Salk_46208	viable
nrpd1-3	At1g63020	Salk_128428	viable
nrpe3b-1	At2g15400	Salk_099705c	viable
nrpb4-1	At5g09920	Salk_122761	viable
nrpe4-2	At4g15950	WiscDslox_476D09	viable
nrpe5-1	At3g57080	Gabi_237A08	viable
nrpe6a-1	At5g51940	SALK_013548	viable
nrpe6b-1	At2g04630	SALK_064868C	viable
nrpd7-1	At3g22900	WiscDsLoxHs_10206G	viable
nrpe7-1	At4g14660	CSHL_ET13987	viable
nrpe8a-1	At1g54250	SALK_151800	?
nrpe8b-1	At3g59600	SAIL_237_E10	?
nrpe9a-1	At3g16980	SALK_032670	viable
nrpe9b-1	At4g16265	SALK_031043	viable
nrpe10-1	At1g11475	SALK_026610	?
nrpe11-1	At3g52090	SALK_100563	lethal
nrpe12-1	At5g41010	SALK_049327	lethal

Double mutants	Viability	Morphology/phenotype
nrpe9a-1 nrpe9b-1	lethal	arrest at globular stage
nrpe5-1 nrpe9a-1	viable	wt
nrpe5-1 nrpe9b-1	viable	9b phenotype
nrpd/e4-2 nrpd7-1	viable	pleiotropic, slow, small
nrpd/e4-2 nrpe7-1	?	-
nrpd7-1 nrpe7-1	?	-
nrpd/e4-2 nrpb4-1	viable	b4-1 phenotype, pleiotropic
nrpd/e4-2 nrpe9a-1	viable	wt
nrpd/e4-2 nrpe9b-1	viable	9b phenotype
nrpe9a-1 nrpd1-3	viable	wt
nrpe9a-1 nrpe1-11	viable	wt
nrpe9b-1 nrpd1-3	viable	9b phenotype
nrpe9b-1 nrpe1-11	viable	9b phenotype
nrpb6a-1 nrpb6b-1	?	-

Figure 1 List of known Pol IV and Pol V mutants

The alleles used in this thesis in shaded in green. Most of homozygous single mutants of Pol IV and V subunits are viable, except for 11th and 12th largest subunits, which are lethal. Double mutants generated in this thesis are shaded in blue, "wt" denotes that the double mutants look like a wildtype plant.

Based on the subunit compositions of Pol II, IV and V determined by mass spectrometry following affinity purification of the enzymes (Ream, Haag et al. 2009), a collection of subunit mutants was obtained and characterized (Figure 1). Homozygous mutants used in this study are shaded in green. In addition, double mutants were also generated in this study, shaded in blue, with a focus mainly on the jaw and clamp domain subunits (9th and 5th) and the 4/7 stalk subcomplex subunits (4th and 7th). Characterization

of the single as well as double mutants is still incomplete, but some insight has been gained from study of the 9th largest subunits (Chapter two) as well as the 4th and 7th subunits (Chapter three). Homozygous single mutants from the 8th and 10th largest subunits have not been isolated, although the T-DNA insertion alleles have been obtained. We anticipate isolating the full collection of the double mutants soon, and a few directed experiments examining small RNA accumulation will be the main focus once the double mutants are available.

		siRNA	DNA methylation					Expression		
Mutant	**Gene**	***siR1003***	***5s* CG**	***5S* CHG**	***5S* CHH**	***AtSN1***	***soloLTR***	***AtSN1***	***soloLTR***	***IGN5***
Wildtype	Col-0	+	+	+	+	+	+	-	-	+
nrpe1-11	At2g40030	---	--	--	--	--	--	++	++	none
nrpe2a-2	At3g23780	none	--	--	--	--	--	++	++	none
nrpd1-3	At1g63020	none	--	--	--	--	--	++	+	+
nrpe3b-1	At2g15400	+	+	?	+	+	+	-	-	none
nrpb4-1	At5g09920	+	?	?	?	+	+	-	-	?
nrpe4-2	At4g15950	--	-+	-	--	--	-	++	+	-
nrpe5-1	At3g57080	-	-+	-	--	--	--	+	++	+
nrpe6a-1	At5g51940	?	?	?	?	-	?	?	?	?
nrpe6b-1	At2g04630	?	?	?	?	-	?	?	?	?
nrpd7-1	At3g22900	+	?	?	?	-+	+	-	-	?
nrpe7-1	At4g14660	+	?	?	?	-	-	+	++	?
nrpe9a-1	At3g16980	+	+	+	+	+	+	-	-	++
nrpe9b-1	At4g16265	+	+	-	--	--	--	++	++	+

Figure 2 siRNA accumulation, DNA methylation and retrotranspon expression profiles in Pol IV and V mutants

The wildtype (Col-0) is included as a reference point. The '+' sign is used to indicate positive levels of a siR1003, DNA methylation or RNA expression, and '-' sign is used to denote decreased or absence of the signal.

The phenotypes of individual subunit mutants are variable, and are often not reported accurately unless a comparison is made between all the other catalytic and non-catalytic subunit mutants in the Pol IV and V pathway. An attempt to illustrate the way each individual mutant subunit affects RNA directed DNA methylation (RdDM) in terms of siRNA accumulation, DNA methylation and retroelement expression is shown in Figure 2.

The catalytic subunits of Pol IV and V, *nrpd1-3*, *nrpd2a-2* and *nrpe1-11* have reliable phenotypes with respect to siRNA accumulation, DNA methylation and retrotransposon reactivation. The only subunit that comes close to phenocopying the siRNA, DNA methylation and retrotransposon expression phenotypes of the catalytic

subunits is *nrpd4/e4-2*. Phylogenetic studies have suggested that NRPD2 and NRPD4 subunits arose in a common ancestor of all angiosperms, as they are present in monocots and dicots, but not in bryophytes (Luo and Hall 2007; Tucker, Reece et al. 2011). Compared to NRPB2 and NRPB4 of Pol II, NRPD2/E2 and NRPD4/E4 are also quite different, as the form a distinct Pol IV/V clade clearly distinguishable from the Pol II subunits in phylogenetic studies (Luo and Hall 2007; Tucker, Reece et al. 2011). The 'uniqueness' of the NRPD4/E4 subunit in Pol IV and V might reflect its essential role in RdDM in partnership with the catalytic subunits of Pol IV and V.

The *nrpe5-1*, *nrpe7-1* and *nrpb9b-1* mutants also share very similar phenotypic profiles (Figure 2), although 24nt siRNA accumulation in *nrpe5-1* is slightly defective compared to *nrpe7-1* or *nrpb9b-1* which accumulate siRNAs at close to wildtype levels. NRPE5 and NRPE7 are Pol V specific subunits, so it stands to reason that these subunits have similar profiles and the loss of the individual subunits have similar effects on Pol V. Double mutants (Figure 1) will be important to analyze to determine if there are additive effects. For instance, the double mutants of *nrpe5-1 nrpb9b-1* should result in Pol V complexes that, if stable, are devoid of the clamp and jaw domains, and the characterization of these double mutant's RdDM profile may inform our understanding of Pol V architecture as it relates to RNA silencing.

Disentangling the functional identity of paralogous non-catalytic subunits subunits of Pol II, III, IV and V

The 3rd largest subunits of Pol II, IV and V

Arabidopsis thaliana has two genes that are homologous to yeast RPB3. NRPB3a (At2g15430) is associated with Pol II, IV and V whereas NRPE3b (At2g15400) associates primarily with pol V (Ream, Haag et al. 2009). NRPB3a and NRPE3b are very tightly linked, separated only by two other genes on the same chromosome arm, and are 88% identical and 94% similar. The T-DNA insertional mutant for *nrpb3b-1* is viable and preliminary assays have been conducted to characterize this mutant, but we are currently awaiting the availability of T-DNA alleles for *nrpb3a* for a complete analysis.

The two sixth largest subunits of Pol II, III, IV and V

The two sixth largest subunit NRPB6a (At5g51940) and NRPB6b (At2g04630) are shared by Pol II, III, IV and V in *Arabidopsis* (Ream, Haag et al. 2009). Surprisingly, we have not found NRPB6a or NRPB6b peptides in affinity purified Pol I (Ream and Pikaard, unpublished). This is a deviation from yeast and human, in which RPB6 associates with Pol I, II and III (Woychik, Liao et al. 1990). The 6th largest subunit is the eukaryotic homolog of the bacterial ω subunit, and in bacteria and yeast, plays a role in RNA polymerase assembly (Minakhin, Bhagat et al. 2001). RPB6 in yeast is a crucial interactor with the RPB4/RPB7 subcomplex, providing the essential interaction site between the 10 subunit core polymerase and the 4/7 subcomplex. A single Q100R point mutation in RPB6 is able to abrogate the 4/7 subcomplex association with the core

polymerase in yeast (Tan, Prysak et al. 2003). For this reason, the NRPB6 variant mutants might be of interest for functions that involve the 4/7 subcomplexes of Pol IV and V (Chapter three). Homozygous *nrpb6a-1* and *nrpb6b-1* single mutants exhibit decreased DNA methylation at *AtSN1* (Figure 4) but have no phenotypic abnormalities. Double mutants are currently being isolated but there is a possibility that the double mutant is lethal, since these are subunits that are shared between other essential RNA polymerases. The more sensitive RT-PCR assay for detection of retrotransposon expression should be conducted see if these mutants have defects in RNA silencing.

A forward genetic screen to identify intragenenic and extragenic factors involved in NRPE1 C-terminal domain functions in RNA-directed DNA methylation

The largest subunit of Pol V, encoded by the NRPE1 gene, has an extended C-terminal domain consisting of various identifiable motifs. Deletion and overexpression assays, done with Pikaard lab member Jeremy Haag, show that multiple subdomains of the CTD have functions relevant to RNA directed DNA methylation (RdDM) (Appendix C). The overexpression of just the NRPE1-CTD (E1C) in wildtype plants has a dominant suppressor effect on RdDM, disrupting 24nt siRNA biogenesis, RNA silencing and DNA methylation: very similar to a catalytic subunit mutant of Pol IV or V, and most like *nrpd2/e2* mutants which is essentially a Pol IV/V double mutant.

It should be possible to take advantage of the dominant suppressor effect for a forward genetic screen to look for suppressors of the E1C dominant negative phenotype, thereby identifying intragenic and extragenic factors that are important for CTD functions. Postdoc Todd Blevins in the Pikaard lab has observed that double mutants between *hda6* with Pol IV or V catalytic mutants exhibit phenotypes that are reminiscent of a *drm1 drm2 cmt3* mutant (Figure 4) (Henderson and Jacobsen 2008). This phenotype, where the plants are dwarfed and exhibit curled leaves, is due to the overexpression of an endogenous F-box containing protein known as *SDC* due to a the loss of RNA silencing of repetitive elements in the *SDC* gene promoter (Henderson and Jacobsen 2008). The silencing of *SDC* requires two out of three of the following components: assymetric methylation via 24nt siRNAs (Henderson and Jacobsen 2008), CHG methylation via CMT3/KRYPTONITE (Henderson and Jacobsen 2008) and histone deacetylation via

HDA6 (Blevins and Pikaard, unpublished). Because the dominant suppressor phenotypes of E1C phenocopies defects in siRNA directed CHH methylation (Appendix C), we tested whether E1C overexpression could participate in *SDC* misregulation in cooperation with *hda6, cmt3* or *kyp* mutants. Transformation of the EIC overexpressor transgene constructs into mutants of *axe1-5, cmt3* or *kyp-6* shows that it does (Figure 5).

In transformants that are dwarfed and have curled leaves, the *SDC* gene is overexpressed compared to a wildtype transformed lines (Figure 6). I propose an EMS forward genetic screen using plants overexpressing E1C in the Col-0 ecotype *cmt3* mutant background to isolate suppressors of the *SDC* phenotype. The *SDC* phenotype is robust when E1C is overexpressed in *cmt3* and these plants are fertile. Moreover, there are L*er* lines (a different *A. thaliana* strain) that have the *cmt3* mutation such that outcrossing mutants recovered in the Col-0 genetic background to L*er cmt3* will speed up the mapping of mutant loci in the mapping population, using bulk segregant analysis and deep sequencing to identify the suppressor loci (Lukowitz, Gillmor et al. 2000).

This genetic approach was conceived by Todd Blevins and me. Currently, a graduate student is assisting us with the initial preparation leading up to EMS mutagenesis. Controls to eliminate false positives include an *sdc-1* mutant line that has also been transformed with the same E1C overexpression construct. This line will be used for crosses when complementation groups are assigned, ruling out mutations in the *SDC* genes and allowing for the discovery of novel mutations.

Figures and figure legends

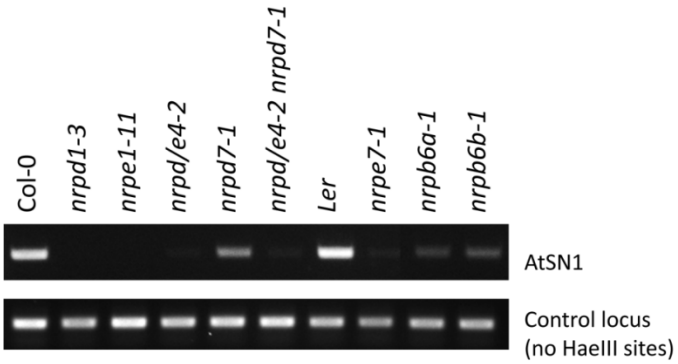


Figure 3 CHH DNA methylation of *AtSN1* assayed using methylation sensitive PCR

Robust PCR signal from *AtSN1* in the wildtype (Col-0) after digestion with *Hae*III, while decreased CHH methylation fom this locus results in no amplification from *nrpd1-3, nrpe1-11, nrpd4/e4-2* and *nrpe7-1*. CHH methylation in *nrpb6a-1* and *nrpb6b-1* are decreased compared to wildtype.

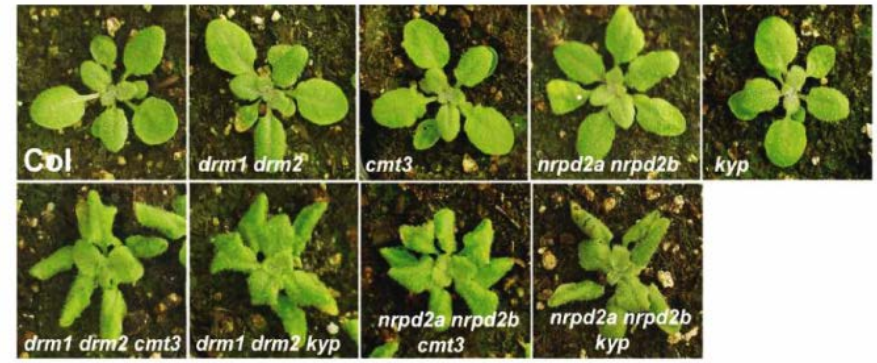


Figure 4 Plants showing wildtype (Col-0) leaf phenotypes or curled leaves in mutants such as *drm1 drm2 cmt3* (*Figure is from (Henderson and Jacobsen 2008))

At the top panel, single mutants of *drm2, cmt3, nrpd2a* and *kyp* have leaves that are wildtype (Col-0). Bottom panel shows mutant combinations that causes curling leaves, due to *SDC*-overexpression.

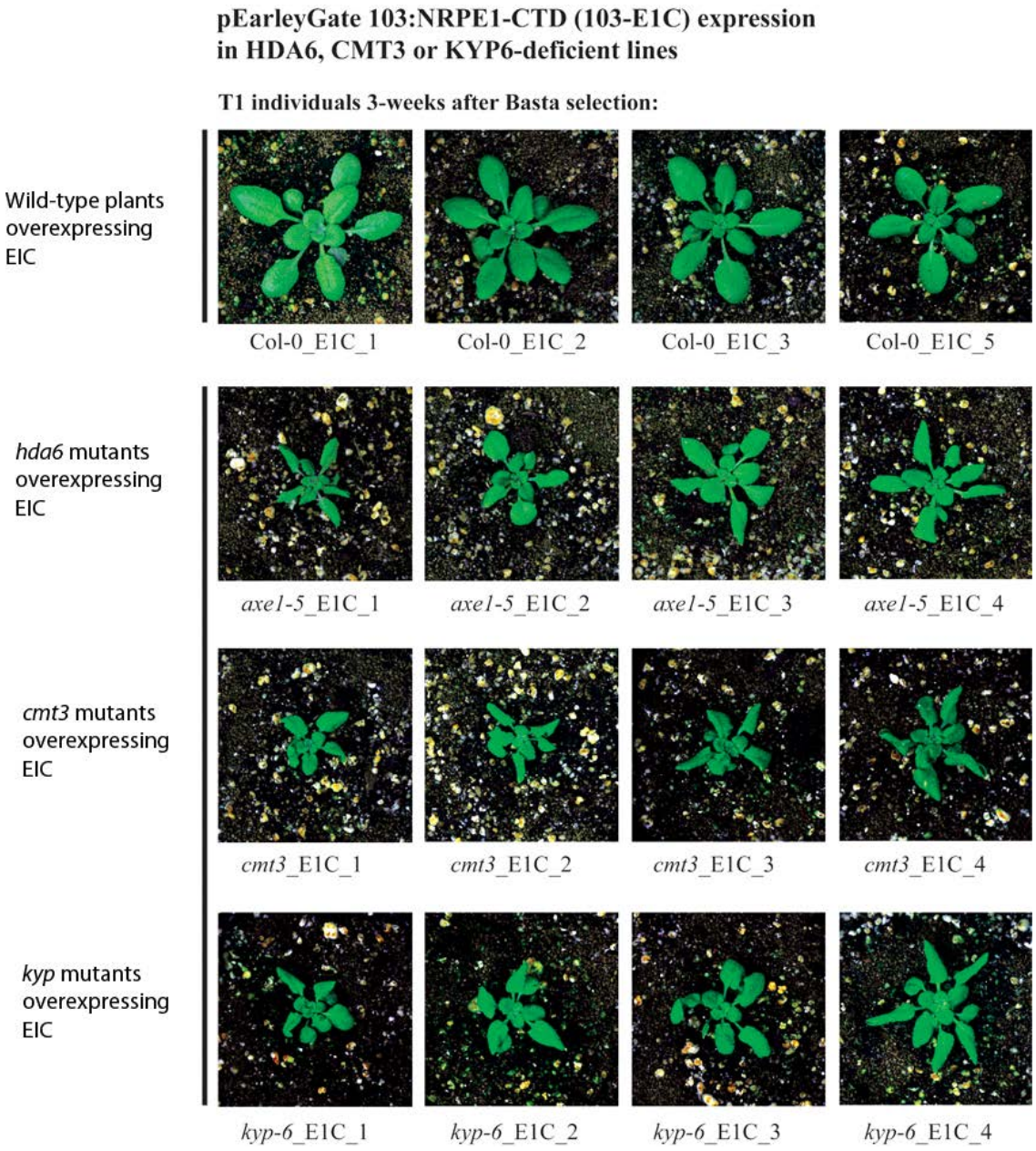


Figure 5 Phenotypes of Pol V largest subunit CTD (E1C) overexpression in wildtype (Col-0), *hda6* or mutants defective in CHG methylation

Wildtype (Col-0) overexpressing E1C have flat, wildtype leaves whereas *hda6*, *cmt3* and *kyp* mutants overexpressing E1C look like *SDC*-overexpressing mutants, with curled, downward facing leaves.

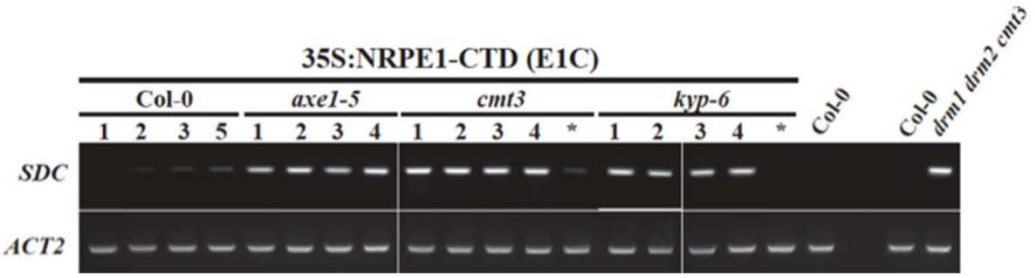


Figure 6 RT-PCR of *SDC* expression in the mutant lines overexpressing EIC

Overexpression of EIC in wildtype (Col-0) alone does not cause *SDC* overexpression, compared to the *drm1 drm2 cmt3* triple mutant shown on the far right. *SDC* overexpression is observed when EIC is overexpressed in *hda6* (*axe1-5*), *cmt3* and *kyp* mutants. Asterisk denote low *SDC* overexpression, and these plants do not have the curled leaf phenotypes.

REFERENCES CITED

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

Lukowitz, W., C. S. Gillmor, et al. (2000). "Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you." Plant Physiol **123**(3): 795-805.

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

Minakhin, L., S. Bhagat, et al. (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.

Ream, T. S., J. R. Haag, et al. (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.

Tan, Q., M. H. Prysak, et al. (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.

Tucker, S. L., J. Reece, et al. (2011). "Evolutionary History of Plant Multisubunit RNA Polymerases IV and V: Subunit Origins via Genome-Wide and Segmental Gene Duplications, Retrotransposition, and Lineage-Specific Subfunctionalization." Cold Spring Harb Symp Quant Biol.

Woychik, N. A., S. M. Liao, et al. (1990). "Subunits shared by eukaryotic nuclear RNA polymerases." Genes Dev **4**(3): 313-323.

APPENDIX A

DNA METHYLATION AND DEMETHYLATION IN *ARABIDOPSIS*

ii.

DRM2 ACTIVE SITE REQUIREMENTS FOR DE NOVO DNA METHYLATION

Post replicative de novo DNA methylation is of great interest in both the mammalian and plant fields as the deposition of DNA methylation is important for silencing repetitive DNA and for epigenetic reprogramming during development (Aravin and Hannon 2008; Law and Jacobsen 2010). The plant de novo methyltransferase, DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2) is the enzyme that is responsible for DNA methylation for RNA directed silencing, as shown by both forward and reverse genetic screens. (Cao and Jacobsen 2002; Henderson, Deleris et al. 2010; Naumann, Daxinger et al. 2011).

The conserved domains of DRM2, as its name suggests, are rearranged compared to other DNA methyltransferases from bacteria and mammals, but DRM2 still retains all the necessary motifs for catalysis (Henderson and Jacobsen 2007). The N-terminal region of DRM2 includes ubiquitin binding domains (UBA) which are unique features of plant DRM proteins. These UBA domains were recently reported to be essential for DRM2 catalysis in vivo (Henderson, Deleris et al. 2010).

Active DNA methyltransferases have invariant PCG amino acids, in motif IV, that are used for catalysis. The thiol group on the cysteine residue covalently attaches to the cytosine base, producing an intermediate which allows the transfer of the methyl group of Ado-Met to the cytosine (Goll and Bestor 2005). The P586 residue of the PCG motif in DRM2 has been reported to be important based on a mutation at this position isolated in a forward genetic screen (Naumann, Daxinger et al. 2011), and a reverse

genetic approach showed that mutating C587 to an alanine abrogates DRM2 function in vivo (Henderson, Deleris et al. 2010).

Prior to the publications mentioned above (Naumann, Daxinger et al. 2011;Henderson, Deleris et al. 2010), I had generated *drm1/2* mutant lines expressing wildtype *DRM2* transgenes and mutant *DRM2* transgenes, in which the invariant Cysteine at position 587 was mutated to serine (Figure 1a). Expression of wildtype DRM2 in the null *drm1/2* mutant background is able to rescue RdDM deficiencies in *drm1/2* mutants, restoring retroelement silencing, whereas expression of the C587S DRM2 mutant does not rescue the *drm1/2* mutants (Figure 1b). Likewise, DNA methylation is restored in *drm1/2* plants expressing wildtype DRM2, but is not restored when the C587S DRM2 is expressed (Figure 1c). Protein expression of both wildtype and C587S DRM2 can be detected by on immunoblots, which suggests that the C587S mutation on DRM2 does not affect expression, but is enzymatically inactive (Figure 1d). These results concur with the findings of the other groups (Naumann, Daxinger et al. 2011; Henderson, Deleris et al. 2010), but the C597S mutation reported here is novel.

Chromatin IP (ChIP) was performed to determine if loss of RNA silencing in *drm1/2* mutants can be correlated to the loss of heterochromatin modifications at the target loci (Figure 2). The housekeeping gene, Actin, is used as a control for active chromatin modifications assayed using an antibody against acetylated histone H3 (AcH3). In all samples tested, Actin is enriched in the AcH3 IPs. In the wildtype (WS) background, retroelements (*AtSN1* and *IG/LINE*) and DRM2 targets (*IGN5, SDC*, G755 and *480*), are primarily associated with H3K9me2 and not with AcH3, consistent with their silenced state. The *drm1/2* mutant rescued with wildtype DRM2 has the same

patterns seen wildtype plants. However, *drm1/2* and *drm1/2 C857S DRM2* mutants have increased AcH3 at retroelements and DRM2-target genes, as a consequence of the loss of RNA silencing in *drm1/2* mutants. Attempts to ChIP wildtype DRM2 or C587S DRM2 to the target loci was unsuccessful (Figure 2). It is also possible that the interaction of DRM2 could be mediated by DRM3, an enzymatically inactive DRM protein which is missing the catalytic motif IV domains, but has recently been shown to mediate DNA methylation by DRM2 (Henderson, Deleris et al. 2010). The DRM2/DRM3 interaction is reminiscent of the Dnmt3/Dnmt3L interaction that is crucial for the de novo activity of mammalian de novo methyltransferase (Jia, Jurkowska et al. 2007). Cross-linking using formaldehyde might not preserve this DRM2/DRM3 interaction due to their large sizes and as a result, an IP via the FLAG-tag on DRM2 alone is not able to preserve this interaction.

In an attempt to study in vitro biochemical activity of DRM2, I designed an assay using short dsDNA oligos that would detect DNA methylation by virtue of the resistance of the methylated DNA to methylation sensitive restriction endonucleases (Figure 3b). The DNA sequence used is from the *AtSN1* retroelement which is a major target of DRM2 (Figure 3a). Using *Hae*III Methyltransferase as a positive control, the dsDNA oligo can be methylated, conferring resistance to the *Hae*III restriction endonuclease, which is methylation sensitive (Figure 3c). This reaction is dependent on S-adenosyl-methionine (SAM) co-factor, which is used by both bacterial and mammalian DNA methyltransferase for catalysis. This in vitro DNA methylation assay was conducted using DRM2:FLAG immunoprecipitated using anti-FLAG resin from plant extracts as well as purified HIS:DRM2 which was recombinantly produced in *E. coli*. However, no

SAM-dependent DNA methylation activity was detected in both these experiments (Figure 3c). Presumably, other proteins or chromatin templates are needed for DRM2 activity.

The mechanistic details of RNA directed DNA methylation is still not clearly defined. For instance, the current model proposes the targeting by 24nt siRNA of AGO4 to Pol V transcription sites, leading to subsequent of de novo DNA methylation. However, how DRM2 methylation is recruited and methylation accomplished remains a mystery.

Figure and figure legends

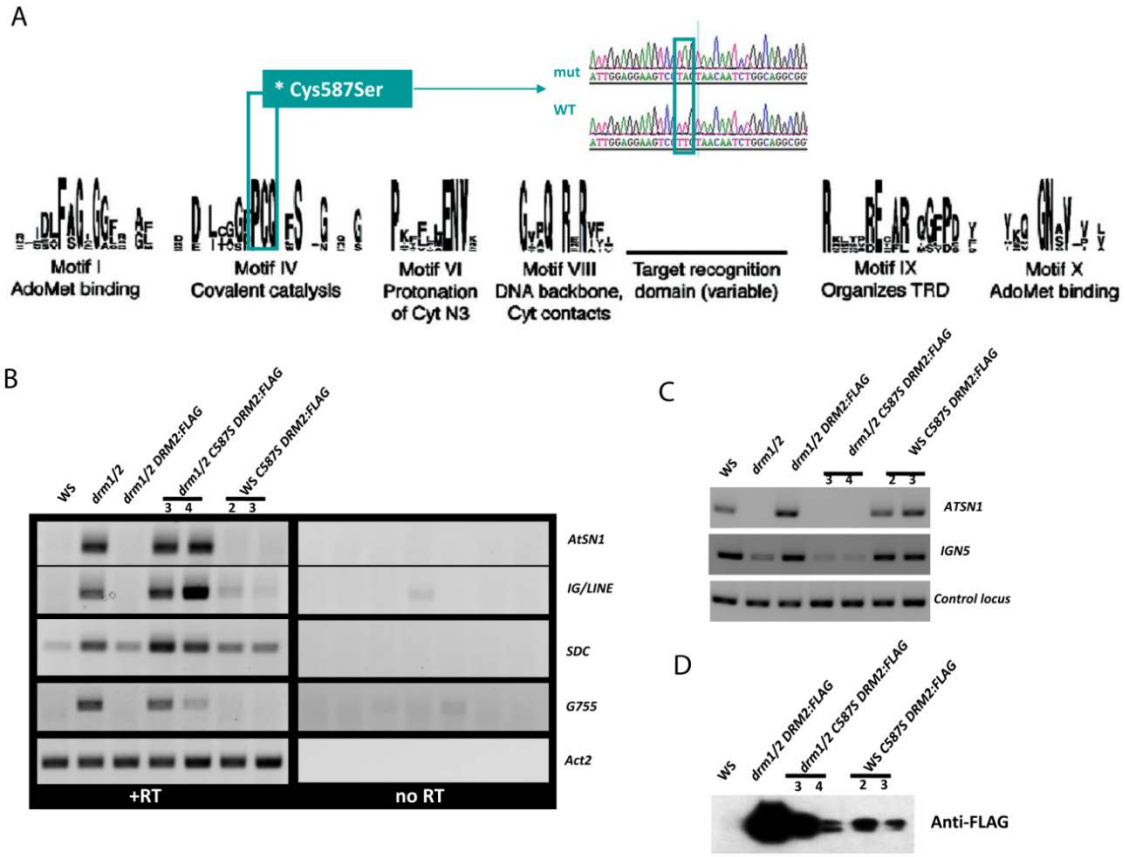


Figure 1 Characterization of a C587S active site mutation in DRM2

(A) Mutagenesis of the invariant Cysteine amino acid at position 587 to Serine in DRM2 is shown in the sequence alignment. DNA methyltransferase active site motif is from (Goll and Bestor 2005).

(B) RT-PCR assays of DRM2-dependent RdDM targets *AtSN1*, *IG/LINE*, *SDC* and *G755* performed in the mutants of *DRM2* and *drm1/2* expressing wildtype *DRM2* or mutant *C587S DRM2* transgenes.

(C) Assay for assymetric DNA methylation of DRM2 target loci, using *Hae*III restriction digests, followed by PCR in mutants of *DRM2* and *drm1/2* expressing wildtype *DRM2* or mutant *C587S DRM2* transgenes. The control locus contain no *Hae*III restriction sites.

(D) Western blot using an anti-FLAG antibody to detect protein expression of FLAG-tagged wildtype *DRM2* or mutant *C587S DRM2* transgenes in *drm1/2* mutants and wildtype (WS).

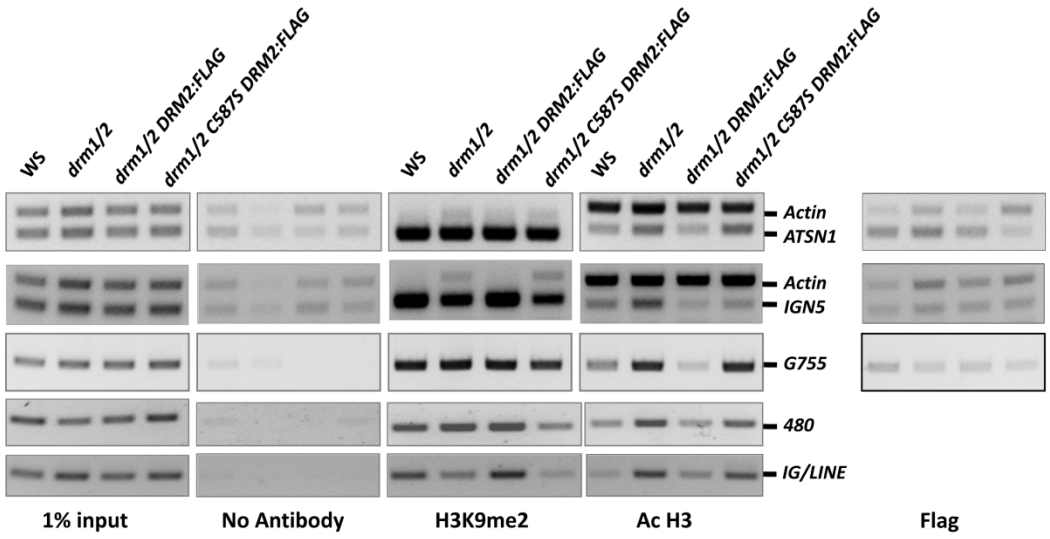


Figure 2 Chromatin IP of *drm1/2* mutants complemented with either wildtype DRM2 or C587S DRM2

Chromatin isolated from wildtype (WS), *drm1/2*, *drm1/2* expressing the FLAG-tagged wildtype *DRM2* or FLAG-tagged mutant *C587S DRM2* were immunoprecipitated with antibodies against H3K9me2, acetylated histone H3 or FLAG. PCR was performed from purified DNA obtained from Chromatin IPs, using primers to various targets of RdDM or Actin.

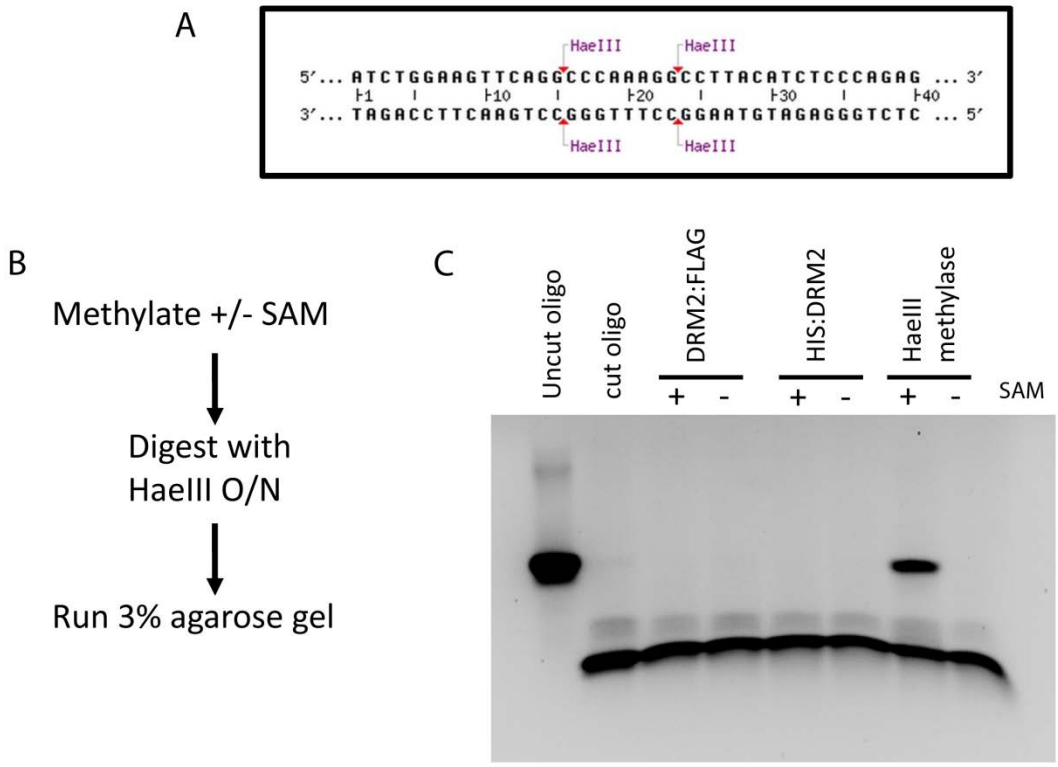


Figure 3 In vitro methylation assay using *Hae*III and DRM2 DNA methyltransferases

(A) Sequence of the DNA oligo used for the in vitro methylation assay, with *Hae*III sites marked

(B) Scheme for in vitro assay – Enzyme is incubated with or without the S-adenosyl-methionine (SAM) co-factor, digested with the *Hae*III restriction enzyme overnight (O/N) and the oligos are electrophoretically separated on a 3% agarose gel

(C) Visualizing the dsDNA oligo on an Ethidium Bromide stained gel. Methylation by the *Hae*III methylase prevents *Hae*III restriction digest when SAM is present. DRM2 proteins do not methylated the dsDNA oligo.

ii.

OVEREXPRESSION OF ROS3, AN RNA BINDING PROTEIN INVOLVED IN DNA DEMETHYLATION, DISRUPTS UNIPARENTAL rRNA GENE SILENCING IN *ARABIDOPSIS SUECICA*

My contributions to this work:

The pENTR plasmid clone containing ROS3 cDNA kindly provided by Jian-Kang Zhu. I cloned the ROS1, DML2 and DML3 cDNAs and generated all the overexpression constructs. Stable transformation into *Arabidopsis suecica* was also performed by me as were the subsequent experimental assays. The graphical representation of the Cymate output from bisulphite sequencing was generated with the help of Todd Blevins. I made all the figures and prepared the manuscript. Craig Pikaard provided valuable editorial assistance.

Overexpression of ROS3, an RNA binding protein involved in DNA demethylation, disrupts uniparental ribosomal RNA silencing in nucleolar dominance

Ek Han Tan[1,2], Jian-Kang Zhu[3], Craig S. Pikaard[1*]

[1]Department of Biology and Department of Molecular and Cellular Biochemistry, Indiana University, Bloomington, Indiana 47405

[2]Division of Biology and Biomedical Sciences, Washington University in St Louis, St Louis, Missouri 63110

[3]Department of Botany and Plant Sciences, Institute for Integrative Genome Biology, University of California, Riverside, California 92521

* To whom correspondence should be addressed

Email addresses:

EHT: ehtan@wustl.edu

CSP: pikaard@indiana.edu

JKZ: jian-kang.zhu@ucr.edu

Running title: DNA demethylation by ROS3 disrupts nucleolar dominance

Article in preparation

Abstract

In plants, a class of DNA glycosylases related to the embryonic imprinting demethylase, DEMETER orchestrates the removal of methylated cytosines across the genome. The activity of the DEMETER and the DEMETER-Like glycosylases; ROS1, DML2 and DML3 are thought to counteract the effects of RNA directed DNA methylation, allowing for epigenomic plasticity. An RNA binding protein, ROS3 has also been implicated in the DNA demethylation pathway and the targeting of ROS1 to sites of action. Using nucleolar dominance in *Arabidopsis suecica* as a model system, we find that the overexpression of ROS3, DML2 or DML3 is capable of disrupting the developmentally-regulated establishment of silencing of the *Arabidopsis thaliana*-derived rRNA genes. The spacer promoter regions in the intergenic spacers of the *Arabidopsis thaliana*-derived rRNA genes were found to be demethylated when ROS3 was overexpressed. The ability of ROS3, DML2 and DML3 overexpression to demethylate rRNA genes, combined with their partial localization in the nucleolus may play a role in reversing or preventing the silencing of rRNA genes by epigenetic mechanisms involving cytosine methylation.

Introduction

Nucleolar dominance is occurs in interspecific hybrids and describes an epigenetic phenomena in which ribosomal RNA (rRNA) genes from one parent are selectively silenced, a phenomena demonstrated in flies, frogs, animals, plants and other eukaryotes (Chen and Pikaard 1997; Pikaard 2000; Pikaard 2000; Preuss and Pikaard 2007; Tucker, Vitins et al. 2010). This phenomenon is well characterized in *Arabidopsis suecica,* an allotetraploid hybrid of *Arabidopsis arenosa* and *Arabidopsis thaliana* in which the rRNA genes derived from *Arabidopsis thaliana* are developmentally silenced and are transcriptionally inactive in mature plants (Pontes, Lawrence et al. 2007). *Arabidopsis suecica* is amenable to targeted gene knockdown by RNAi-inducing transgenes introduced into the genome via *Agrobacterium* assisted transformation (Lawrence and Pikaard 2003). This has led to identification of activities required for nucleolar dominance including the histone deacetylase, HDA6 and HDT1, which are involved in the repression of inactive rRNA genes (Earley, Lawrence et al. 2006; Earley, Pontvianne et al. 2010). Players in the facultative heterochromatic 24nt siRNA pathway in plants, chiefly the de novo DNA methyltransferase DRM2, RDR2 and DCL3 are also involved in maintaining rRNA gene silencing in *A. suecica* (Preuss, Costa-Nunes et al. 2008).

The methylation of symmetric cytosines in the CG dinucleotide context is a common feature in eukaryotic genomes and is often associated with gene repression (Bird 1986). After replication, maintenance DNA methytransferases (DNMT1 in mammals, MET1 in plants) recognize hemimethylated CG sites and direct the deposition of a methyl group on the unmethylated CG (Yoder, Soman et al. 1997; Kankel, Ramsey et al.

2003; Song, Rechkoblit et al. 2011). Plants also have robust DNA methylation machineries to methylate DNA in CHG (H standing for either C, A or T) and CHH contexts (Wassenegger and Pelissier 1998), accomplished by the CMT3 and DRM2 de novo DNA methyltransferases (Lindroth, Cao et al. 2001; Cao and Jacobsen 2002; Cao and Jacobsen 2002; Cao, Aufsatz et al. 2003). RNAi-mediated knockdown of DRM2 but not MET1 or CMT3 in *Arabidopsis suecica* results in the loss of nucleolar dominance, indicating that DRM2 is the key methyltransferase required to establish nucleolar dominance (Preuss, Costa-Nunes et al. 2008).

The Demeter (DME) DNA glycosylase/lyase is an enzyme that actively removes methylated cytosines in the central cell and endosperm of the developing embryo, specifically demethylating the maternal allele of the *Medea* gene and facilitating its expression (Choi, Gehring et al. 2002). DME-Like DNA glycosylases that include DML2, DML3 and Repressor of Silencing 1 (ROS1) are not similarly implicated in imprinting mechanisms but are expressed throughout the life cycle of the plants in order to maintain a balanced methylome (Gong, Morales-Ruiz et al. 2002; Agius, Kapoor et al. 2006; Penterman, Uzawa et al. 2007; Penterman, Zilberman et al. 2007). These enzymes have been shown biochemically to cleave methylated DNA substrates in vitro (Morales-Ruiz, Ortega-Galisteo et al. 2006; Ortega-Galisteo, Morales-Ruiz et al. 2008) and are thought to recruit DNA repair machinery for the substitution of unmethylated cytosines for methylcytosine (Gehring, Reik et al. 2009). A recent discovery in the field of DNA demethylation hints at an RNA component in active DNA demethylation. Specifically, Repressor of Silencing 3 (ROS3), a gene identified in the same genetic screen that identified the DNA glycosylase ROS1, contains an RRM domain, RNA binding motif

that binds small RNAs in vitro and in vivo (Zheng, Pontes et al. 2008), suggesting that small RNAs may guide ROS3 to sites of demethylation.

Localization of ROS1 and ROS3 indicate their presence in the nucleolar compartment as well as the nucleus (Zheng, Pontes et al. 2008), very much like the localization of Histone Deacetylase6 (HDA6) localization (Earley, Lawrence et al. 2006), which is involved in maintaining nucleolar dominance. In this report, we show that active DNA demethylation via overexpression of DNA gylcosylases DML2 and DML3 can prevent the establishment of nucleolar dominance in *Arabidopsis suecica*. ROS3 overexpression also disrupts nucleolar dominance but surprisingly, the overexpression of ROS1 does not abrogate silencing of *Arabidopsis thaliana* NORs. We discuss some possible roles for ROS3 and ROS1 based on our current understanding of these two enzymes and bring to light some effects of DNA demethylation on Pol I transcription.

Materials and Methods

Cloning and generation of transgenic plants

The pENTR clone containing the *ROS3* cDNA was obtained from Prof. Jian Kang Zhu. PCR amplification of of *ROS1, DML2* and *DML3* was performed using primers listed in Table S1, using reverse transcribed total RNA from wildtype *Arabidopsis thaliana* (Col-0 ecotype). Full length cDNA PCR products were introduced into pENTR D-TOPO (Invitrogen) and recombined into pEARLEYGATE 202, 104 or 103 (Earley, Haag et al. 2006) using LR Clonase II (Invitrogen). Clones were verified by sequencing before transforming into *Agrobacterium tumefaciens* GV3101. Stable transformation of *Arabidopsis suecica* LC1 was performed as previously described (Lawrence and Pikaard 2003).

RT-PCR assays

Detection of *A. arenosa* and *A. thaliana* rRNA transcripts from *A. suecica* was performed using RT-PCR, followed by restriction digestion, to differentiate the polymorphism between the two species in the ITS1 region (Lewis and Pikaard 2001), an assay referred to as ITS1 CAPS (cleaved amplified polymorphic sequence). The primers used to detect overexpression of ROS1, ROS3, DML2 and DML3 are provided in Table S1. Briefly, 1μg of total RNA was treated with RQ1 DNAse (Promega) before being reversed transcribed with random $d(N)_6$ primers using Superscript III (Invitrogen) reverse transcriptase. ITS1 PCR was performed on 10ng of reverse transcribed RNA followed by

overnight digestion with *Hha*I restriction endonuclease (New England Biolabs) and separated on a 2.5% gel by electrophoresis.

Bisulphite sequencing

Bisulphite sequencing of the spacer promoter regions as well as the gene promoter regions of *A.thaliana* rRNA genes in *A. suecica* was performed using primers and the protocol described in (Preuss, Costa-Nunes et al. 2008). The resultant clones were analyzed using the Cymate program (Hetzl, Foerster et al. 2007). The outputs from Cymate were parsed using a custom Perl script that collates the methylation frequency on each methylcytosine site. These frequency were then plotted in Microsoft Excel. The number of clones used: gene promoter (n=30 per genotype), spacer promoter (n=20 per genotype)

Results

Localization of ROS1 and ROS3 in Arabidopsis suecica

Stably transformed *A. suecica* harboring transgenes expressing YFP:ROS1 and ROS3:GFP show that ROS1 and ROS3 are nuclear localized, based on their detection usin fluorescence microscopy of root tip cells (Figure 1). Discrete speckles and nucleolar localization is observed as well, in agreement with ROS1 and ROS3 localization patterns describe previously in *A. thaliana* (Zheng, Pontes et al. 2008).

ROS3 overexpression disrupts nucleolar dominance

RT-PCR using the ITS1 CAPS assay (Lewis and Pikaard 2001) on *A.thaliana, A. suecica* and *A. arenosa* show that in *A. suecica*, only the *A. arenosa* rRNA genes are expressed, as the *A. thaliana* specific band at around 250bp is not detected. However, in *A. suecica* expressing the *35S::FLAG::ROS3* transgene, *A. thaliana* rRNA genes are expressed in *A. suecica*, which we interpret as the failure to establish nucleolar dominance during development (Figure 2a). The overexpression of just the ROS3 RNA-binding motif (ROS3-RRM) does not affect nucleolar dominance (Figure S1).

DML2 and DML3, but not ROS1, overexpression also disrupts nucleolar dominance

Ectopic overexpression of DNA glycosylases DML2 and DML3 also prevents *A. thaliana*-derived rRNA gene repression in *A. suecica* (Figure 1a). However, nucleolar dominance is not disrupted in *A. suecica* when ROS1 is overexpressed (Figure 1b).

Demethylation of the intergenic spacer promoter region of A. thaliana-derived rRNA genes in A. suecica

Using bisulphite sequencing, the cytosine methylation positions in the intergenic spacer as well as gene promoter regions of *Arabidopsis thaliana*-derived rRNA genes in *A. suecica* were determined. Genomic DNA was isolated from a line overexpressing *ROS3,* which is disrupted in nucleolar dominance, and compared to wildtype *A. suecica* (LC1). At the intergenic spacer promoter, DNA methylation is reduced in the line overexpressing ROS3 compared to wildtype (Figure 3a), while at the gene promoter region assayed here, there was no difference (Figure 3b). The total methylation in the intergenic spacer promoter was reduced about 50% (Figure 3c) between the ROS3 overexpressing line compared to wildtype, while the total methylation at the gene promoter regions were unchanged (Figure 3d).

Discussion

Active DNA demethylation has important developmental functions in both plants and mammals (Feng, Jacobsen et al. 2010). Plant DNA glycosylases such as DME, ROS1, DML2 and DML3 can recognize methylated cytosine and cleave the phosphodiester bond, which recruits base excision repair machinery to the site, replacing the methlyated cytosine with an unmethylated cytosine (Morales-Ruiz, Ortega-Galisteo et al. 2006; Ortega-Galisteo, Morales-Ruiz et al. 2008). Overexpression of DME has been shown to demethylate some of its target genes (Ohr, Bui et al. 2007), so we sought to determine if ROS1 and ROS3 are able to target rRNA genes, since they are localized to the nucleolus. At the same time, DML2 and DML3 were also tested to determine their overexpression phenotypes.

We show that ROS3 but not ROS1 causes a failure in establishing nucleolar dominance when overexpressed. DML2 and DML3 overexpression also prevents nucleolar dominance establishment, suggesting that ROS3, DML2 and DML3 may play roles in the epigenetic regulation of rRNA genes (Figure 4). The regulation of ribosomal RNA genes is thought to involve as an on/off switch (Lawrence, Earley et al. 2004; Lawrence and Pikaard 2004) in which concerted changes in DNA methylation and chromatin modifications occur. The ectopic expression of ROS3, DML2 and DML3 can prevent the establishment of DNA methylation on the *A. thaliana*-derived rRNA in *A. suecica,* presumably preventing the establishment of heterochromatic histone modifications and silencing.

One of the caveats of these experiments is that loss of nucleolar dominance is caused by virtue of the overexpression of DNA demethylase machinery, which may

reflect a gain of function phenotype not necessarily reflecting the true in vivo regulatory process. When ROS3, DML2 and DML3 were knocked down using artificial microRNAs, nucleolar dominance establishment was normal in all cases. Unregulated methylation and silencing of all rRNA genes would be lethal. Therefore, if ROS3, DML2 and DML3 play roles in regulating the number of active rRNA genes by balancing de novo methylation with demethylation, other mechanisms must exist to prevent runaway methylation and rRNA silencing in their absence.

Figure legends

Figure 1 Nuclear localization of YFP::ROS1 and ROS3::GFP in *Arabidopsis suecica*

(A) ROS1-YFP localizes to the nucleus in stably transformed YFP::ROS1 *A. suecica*.

(B) ROS3-GFP is similarly localized to the nucleus from stably transformed ROS3:GFP plants.

Figure 2 Overexpression of ROS3, DML2 and DML3 but not ROS1 disrupts nucleolar dominance

(A) RT-PCR on individual plants transformed with the transgene containing *ROS3* cDNA driven by the strong 35S promoter, show a loss in nucleolar dominance in four to five T1 plants, where the *A.thaliana* specific band at around 250bp are expressed compared to the *A. suecica* (LC1) wildtype controls. The overexpression of the *DEMETER-Like* (*DML*) proteins *DML2* and *DML3* also results in the derepression of the silent *A. thaliana* rRNA genes in all independent transgenic *A. suecica* lines.

(B) *ROS1* overexpression from an N-terminally FLAG tagged construct or N-terminally YFP -agged construct does not disrupt nucleolar dominance.

(C) Phylogenetic tree generated using a multiple alignment of protein sequences from Demeter (DME), ROS1, DML2 and DML3. DML2 and DML3 group together but ROS1 and Demeter (DME) are grouped separately.

Figure 3 Reduced methylation of *A. thaliana* rRNA intergenic spacer sequences in *A. suecica* overexpressing ROS3

(A) The methylation of the spacer promoter, located upstream of the transcription start site, and within in the intergenic spacer of *A. thaliana* rRNA genes, is demethylated between -70 and +90, relative to the spacer promter transcription start site, defined as +1, in *A. suecica* overexpressing ROS3 compared to wildtype.

(B) The methylation of the *A. thaliana* rRNA gene promoter, between -400 and +1, is not affected when ROS3 is overexpressed in *A. suecica*, compared to wildtype plants.

(C) Bar graphs show that the total methylation in the intergenic spacer promoter in plants overexpressing ROS3 is reduced by at least 50% from all the sites assayed. Mostly CG and CHH sites are effected.

(D) Bar graphs show that total methylation is unchanged in plants overexpressing ROS3 compared to wildtype in the gene promoter.

Figure 4 A model for ribosomal RNA gene regulation

Concerted changes in chromatin modifications and DNA methylation result in a proposed on/off switch regulating rRNA genes in plants. To date, knockdown screens in *Arabidopsis suecica* have highlighted repressive chromatin modifications that maintain the silencing of *Arabidopsis thaliana* rRNA genes. Here, we report that ROS3, an RNA binding protein implicated in DNA demethylation, can function to positively influence rRNA gene expression, while the DNA demethylation activity from DML2 and DML3

activity can also prevent the establishment of silencing on *A. thaliana* rRNA genes in *A. suecica*.

Figure S1 Overexpression of the RRM domain of ROS3 alone does not disrupt nucleolar dominance

RT-PCR using the ITS1 CAPS assay shows that nucleolar dominance is not affected when just the RRM domain of ROS3 is overexpressed.

Table S1 Primers used for cloning and RT-PCR assays

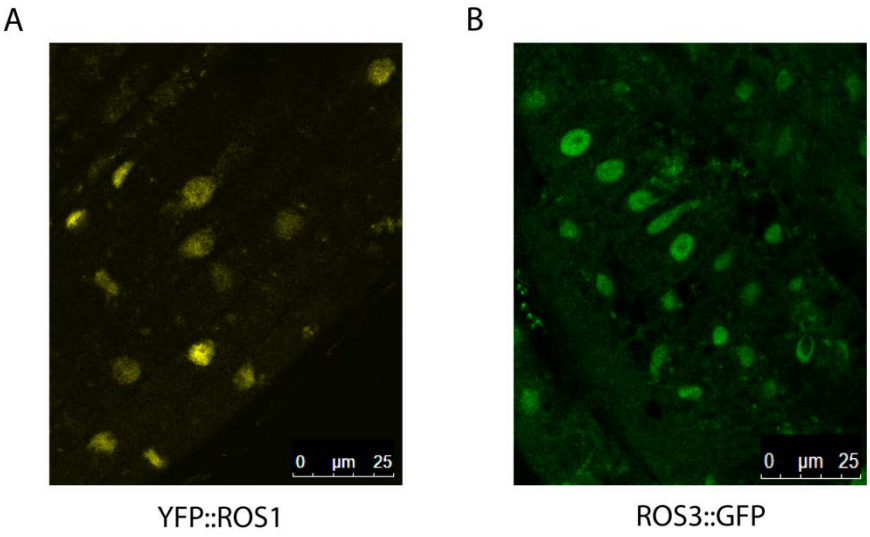


Figure 1

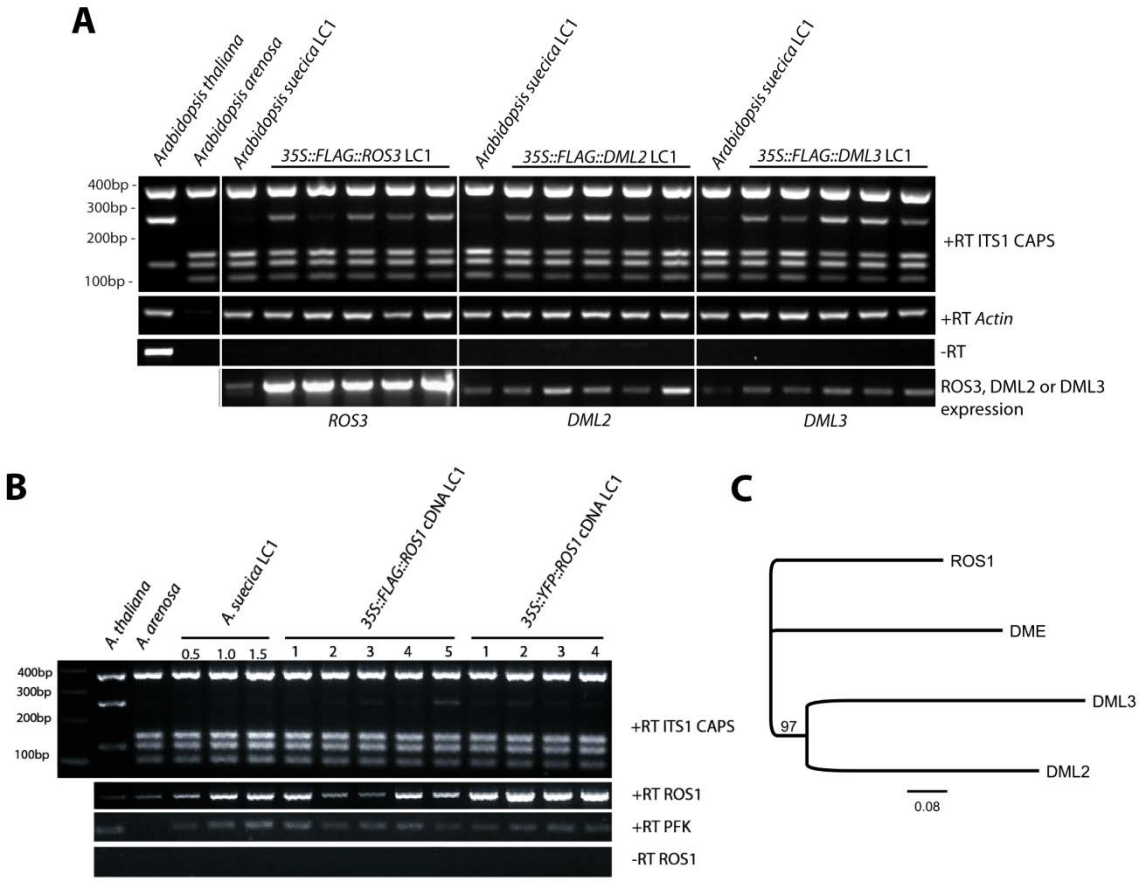

A
Arabidopsis thaliana
Arabidopsis arenosa
Arabidopsis suecica LC1
35S::FLAG::ROS3 LC1
Arabidopsis suecica LC1
35S::FLAG::DML2 LC1
Arabidopsis suecica LC1
35S::FLAG::DML3 LC1
400bp
300bp
200bp
100bp
+RT ITS1 CAPS
+RT Actin
-RT
ROS3, DML2 or DML3 expression
ROS3
DML2
DML3
B
A. thaliana
A. arenosa
A. suecica LC1
35S::FLAG::ROS1 cDNA LC1
35S::YFP::ROS1 cDNA LC1
0.5 1.0 1.5 1 2 3 4 5 1 2 3 4
400bp
300bp
200bp
100bp
+RT ITS1 CAPS
+RT ROS1
+RT PFK
-RT ROS1
C
ROS1
DME
DML3
97
DML2
0.08

Figure 2

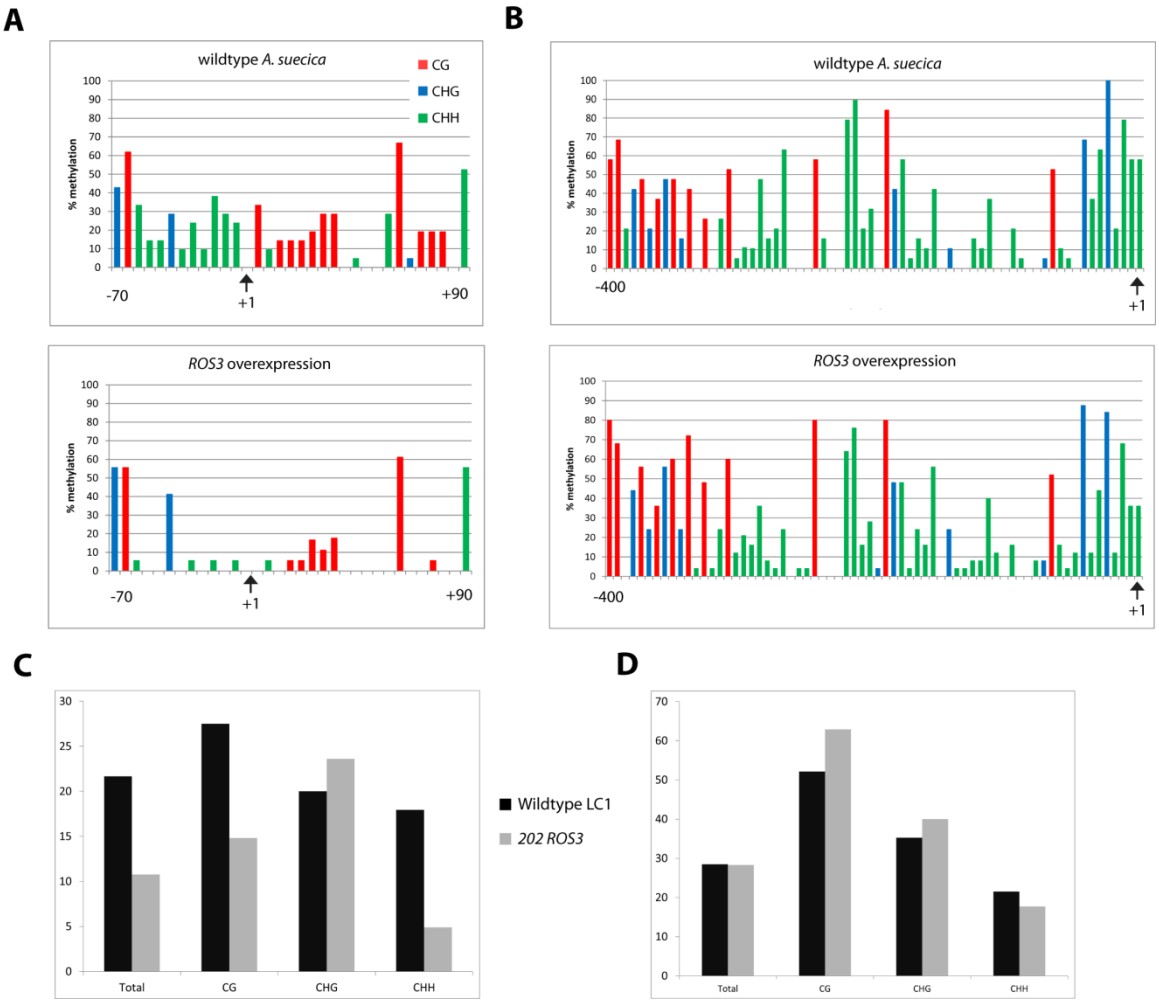

A
wildtype A. suecica
CG
CHG
CHH
% methylation
-70
+1
+90
B
wildtype A. suecica
% methylation
-400
+1
ROS3 overexpression
% methylation
-70
+1
+90
ROS3 overexpression
% methylation
-400
+1
C
Total
CG
CHG
CHH
Wildtype LC1
202 ROS3
D
Total
CG
CHG
CHH

Figure 3

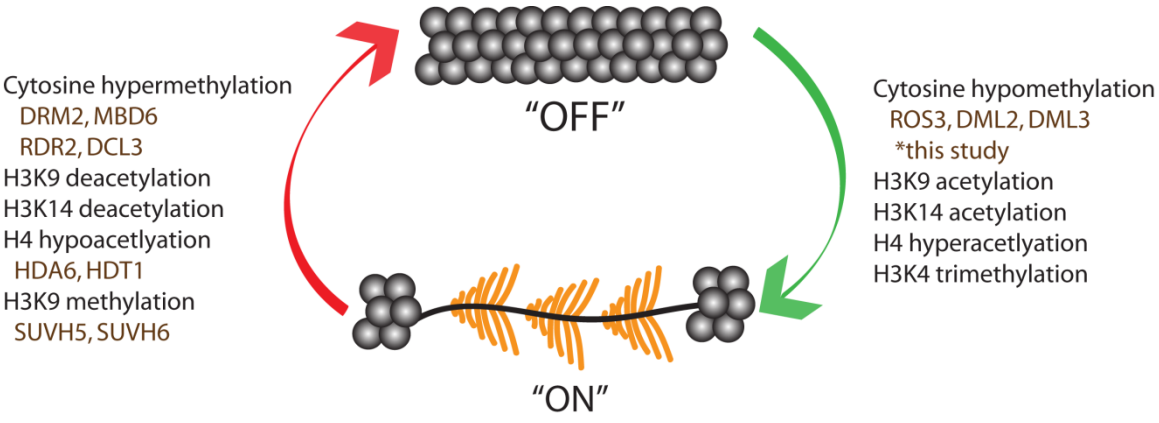

Cytosine hypermethylation
DRM2, MBD6
RDR2, DCL3
H3K9 deacetylation
H3K14 deacetylation
H4 hypoacetlyation
HDA6, HDT1
H3K9 methylation
SUVH5, SUVH6
"OFF"
"ON"
Cytosine hypomethylation
ROS3, DML2, DML3
*this study
H3K9 acetylation
H3K14 acetylation
H4 hyperacetlyation
H3K4 trimethylation

Figure 4

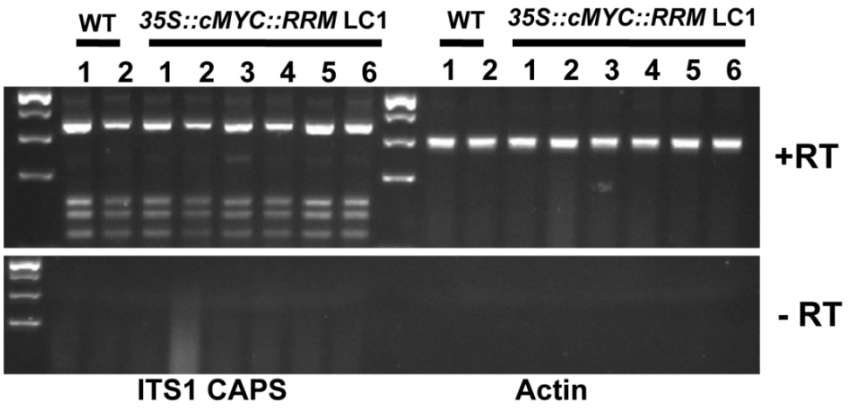

WT
35S::cMYC::RRM LC1
WT
35S::cMYC::RRM LC1
1 2 1 2 3 4 5 6
1 2 1 2 3 4 5 6
+RT
- RT
ITS1 CAPS
Actin

Figure S1

Primers used for cloning full length cDNAs

ROS1 Fwd	CACCATGGAGAAACAGAGGAGAGAAG
ROS1 Rev	TTAGGCGAGGTTAGCTTGTTGTC
DML2 Fwd	CACCATGGAAGTGGAAGGTGAAGTG
DML2 Rev	TCATTCCTCTGTCTTCTCTTTAGTTCTG
DML3 Fwd	CACCATGTTGACAGATGGTTCACAACAC
DML3 Rev	CTATATATCATCATCACTCATAAACTTTGGCC

Primers used for RT-PCR

ROS13F	TGGAAGGGATCCGTCGTGGATTCT
ROS13R	CCCGCGACTCTTGATTGTTTCAGCAACTT
DML2F	ACCCGGAGAGTACCATTCAGACAC
DML2R	TCAGGAGGAACATGTGTTAGCCACTCTAA
DML3F	GCCAAATCGCAAGAAGGTAAGGA
DML3R	GACGTTGCTGTAGATATGAC
ROS3 Fwd	CACCATGGAGGAAAAAAGCAGCGG
ROS3 RRM Rev	TCATCTCGCTTTAAGCGAGCTAG

Table S1

iii.

REFERENCES CITED

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

Aravin, A. A. and G. J. Hannon (2008). "Small RNA silencing pathways in germ and stem cells." Cold Spring Harb Symp Quant Biol **73**: 283-290.

Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." Nature **321**(6067): 209-213.

Cao, X., W. Aufsatz, et al. (2003). "Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation." Curr Biol **13**(24): 2212-2217.

Cao, X. and S. E. Jacobsen (2002). "Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes." Proc Natl Acad Sci U S A **99 Suppl 4**: 16491-16498.

Cao, X. and S. E. Jacobsen (2002). "Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing." Curr Biol **12**(13): 1138-1144.

Chen, Z. J. and C. S. Pikaard (1997). "Epigenetic silencing of RNA polymerase I transcription: a role for DNA methylation and histone modification in nucleolar dominance." Genes Dev **11**(16): 2124-2136.

Choi, Y., M. Gehring, et al. (2002). "DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis." Cell **110**(1): 33-42.

Earley, K., R. J. Lawrence, et al. (2006). "Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance." Genes Dev **20**(10): 1283-1293.

Earley, K. W., J. R. Haag, et al. (2006). "Gateway-compatible vectors for plant functional genomics and proteomics." Plant J **45**(4): 616-629.

Earley, K. W., F. Pontvianne, et al. (2010). "Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation." Genes Dev **24**(11): 1119-1132.

Feng, S., S. E. Jacobsen, et al. (2010). "Epigenetic reprogramming in plant and animal development." Science **330**(6004): 622-627.

Gehring, M., W. Reik, et al. (2009). "DNA demethylation by DNA repair." Trends Genet **25**(2): 82-90.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.

Gong, Z., T. Morales-Ruiz, et al. (2002). "ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase." Cell **111**(6): 803-814.

Henderson, I. R., A. Deleris, et al. (2010). "The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in Arabidopsis thaliana." PLoS Genet **6**(10): e1001182.

Henderson, I. R. and S. E. Jacobsen (2007). "Epigenetic inheritance in plants." Nature **447**(7143): 418-424.
Hetzl, J., A. M. Foerster, et al. (2007). "CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing." Plant J **51**(3): 526-536.
Jia, D., R. Z. Jurkowska, et al. (2007). "Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation." Nature **449**(7159): 248-251.
Kankel, M. W., D. E. Ramsey, et al. (2003). "Arabidopsis MET1 cytosine methyltransferase mutants." Genetics **163**(3): 1109-1122.
Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-220.
Lawrence, R. J., K. Earley, et al. (2004). "A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance." Mol Cell **13**(4): 599-609.
Lawrence, R. J. and C. S. Pikaard (2003). "Transgene-induced RNA interference: a strategy for overcoming gene redundancy in polyploids to generate loss-of-function mutations." Plant J **36**(1): 114-121.
Lawrence, R. J. and C. S. Pikaard (2004). "Chromatin turn ons and turn offs of ribosomal RNA genes." Cell Cycle **3**(7): 880-883.
Lewis, M. S. and C. S. Pikaard (2001). "Restricted chromosomal silencing in nucleolar dominance." Proc Natl Acad Sci U S A **98**(25): 14536-14540.
Lindroth, A. M., X. Cao, et al. (2001). "Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation." Science **292**(5524): 2077-2080.
McStay, B. and I. Grummt (2008). "The epigenetics of rRNA genes: from molecular to chromosome biology." Annu Rev Cell Dev Biol **24**: 131-157.
Morales-Ruiz, T., A. P. Ortega-Galisteo, et al. (2006). "DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases." Proc Natl Acad Sci U S A **103**(18): 6853-6858.
Naumann, U., L. Daxinger, et al. (2011). "Genetic Evidence That DNA Methyltransferase DRM2 Has a Direct Catalytic Role in RNA-Directed DNA Methylation in Arabidopsis thaliana." Genetics **187**(3): 977-979.
Ohr, H., A. Q. Bui, et al. (2007). "Identification of putative Arabidopsis DEMETER target genes by GeneChip analysis." Biochem Biophys Res Commun **364**(4): 856-860.
Ortega-Galisteo, A. P., T. Morales-Ruiz, et al. (2008). "Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks." Plant Mol Biol **67**(6): 671-681.
Penterman, J., R. Uzawa, et al. (2007). "Genetic interactions between DNA demethylation and methylation in Arabidopsis." Plant Physiol **145**(4): 1549-1557.
Penterman, J., D. Zilberman, et al. (2007). "DNA demethylation in the Arabidopsis genome." Proc Natl Acad Sci U S A **104**(16): 6752-6757.
Pikaard, C. S. (2000). "The epigenetics of nucleolar dominance." Trends Genet **16**(11): 495-500.
Pikaard, C. S. (2000). "Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids." Plant Mol Biol **43**(2-3): 163-177.
Pontes, O., R. J. Lawrence, et al. (2007). "Postembryonic establishment of megabase-scale gene silencing in nucleolar dominance." PLoS One **2**(11): e1157.

Preuss, S. and C. S. Pikaard (2007). "rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch." Biochim Biophys Acta **1769**(5-6): 383-392.

Preuss, S. B., P. Costa-Nunes, et al. (2008). "Multimegabase silencing in nucleolar dominance involves siRNA-directed DNA methylation and specific methylcytosine-binding proteins." Mol Cell **32**(5): 673-684.

Song, J., O. Rechkoblit, et al. (2011). "Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation." Science **331**(6020): 1036-1040.

Tucker, S., A. Vitins, et al. (2010). "Nucleolar dominance and ribosomal RNA gene silencing." Curr Opin Cell Biol **22**(3): 351-356.

Wassenegger, M. and T. Pelissier (1998). "A model for RNA-mediated gene silencing in higher plants." Plant Mol Biol **37**(2): 349-362.

Yoder, J. A., N. S. Soman, et al. (1997). "DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe." J Mol Biol **270**(3): 385-395.

Zheng, X., O. Pontes, et al. (2008). "ROS3 is an RNA-binding protein required for DNA demethylation in Arabidopsis." Nature **455**(7217): 1259-1262.

APPENDIX B

GENOMIC IMPRINTING OF POL IV-DEPENDENT SIRNAS IS REGULATED BY A NOVEL MECHANISM

My contributions to this work:

I cloned the cDNAs for ROS1, DML2 and DML3 and generated corresponding transgene constructs for overexpression. Transformation of these constructs into Col-0 and the isolation stably transformed lines were performed by me. I wrote materials and methods concerning the generation of these overexpression lines and participated in discussions with Rebecca Mosher with regards to experimental design.

Genomic imprinting of Pol IV-dependent siRNAs is regulated by a novel mechanism

Rebecca A. Mosher[1,2], Ek Han Tan[3], Juhyun Shin[4], Robert L. Fischer[4], Craig S. Pikaard[3], and David C. Baulcombe[1]*

[1] Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

[2] School of Plant Sciences, University of Arizona, Tucson, AZ 85721, United States

[3] Biology Department, Indiana University, Bloomington, Indiana 47405, United States

[4] Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, United States

* To whom correspondence should be addressed

Email addresses:

RAM: rmosher@cals.arizona.edu

EHT: ehtan@wustl.edu

JS: unowen@berkeley.edu

RLF: rfischer@berkeley.edu

CSP: cpikaard@indiana.edu

DCB: dcb40@cam.ac.uk

Article prepared for **Silence**

Abstract

Background

Small RNAs generated by RNA polymerase IV (Pol IV) are the most abundant class of small RNAs in flowering plants. In *Arabidopsis thaliana* Pol IV-dependent short interfering (p4-si)RNAs are imprinted and accumulate specifically from maternal chromosomes in the endosperm of developing seeds. Imprinted expression of protein-coding genes is controlled by differential DNA or histone methylation placed in gametes. To identify factors required for imprinting of p4-siRNAs we analyzed a series of candidate mutations, including those required for genomic imprinting of protein-coding genes.

Results

Paternal alleles of imprinted genes are marked by DNA or histone methylation placed by DNA METHYLTRANSFERASE 1 or the Polycomb Repressive Complex 2. Here we demonstrate that repression of paternal p4-siRNA expression is not controlled by either of these mechanisms. Similarly, loss of several chromatin modification enzymes, including a histone acetyltransferase, a histone methyltransferase, and two nucleosome remodeling proteins, does not affect imprinting of p4-siRNAs. Maternal alleles of all known imprinted genes are hypomethylated by DEMETER DNA glycosylase, yet uniparental expression of p4-siRNAs occurs irrespective of demethylation by DEMETER or related glycosylases.

Conclusions

Imprinted expression of p4-siRNAs from thousands of genomic loci indicates that maternal and paternal epigenetic modifications are widespread. Here we demonstrate that

differential DNA methylation, although present at many p4-siRNA loci, is not responsible for uniparental expression of p4-siRNAs. We further show that several chromatin modifications associated with epigenetic silencing are not required for genomic imprinting of p4-siRNAs. These data indicate that there are multiple layers of parent-of-origin epigenetic marks – differential DNA methylation triggering imprinted expression of genes, and an unknown epigenetic modification activating imprinted expression of p4-siRNAs.

Background

Mendelian laws of inheritance state that a genetic element behaves identically when transmitted through maternal or paternal gametes. Genetic elements that break this law by exhibiting preferential or exclusive expression when inherited from one parent are genomically imprinted. Genomic imprinting is well described only in placental mammals and flowering plants, although a number of parent-of-origin-dependent effects are observed in other organisms [1-5].

Flowering plants are characterized by double fertilization, whereby two identical haploid sperm cells in the pollen grain fertilize two cells in the female gametophyte. Fertilization of the haploid egg cell generates the diploid embryo while fertilization of the diploid central cell generates the triploid endosperm. The endosperm is functionally analogous to mammalian placenta, acting as a conduit between maternal somatic tissues and the growing embryo but not contributing genetically to the next generation. Endosperm makes up the bulk of grains such as rice, wheat, and maize, making it a critical tissue for human nutrition. With a single exception in maize [6], all characterized imprinted genes in plants display uniparental expression specifically in the endosperm and some imprinted genes affect the growth and development of this tissue [7, 8].

Each imprinted gene is associated with a region of differential parental methylation, with maternal alleles always hypomethylated regardless of which allele is expressed. In *Arabidopsis thaliana* differential methylation is established by the opposing actions of DNA METHYLTRANSFERASE 1 (MET1) in the paternal gametophyte and the DNA glycosylase DEMETER (DME) in the central cell of the female gametophyte [9, 10]. Loss of paternal DNA methylation through mutation of *MET1* causes expression of the

normally silent paternal copies of *FLOWERING WAGENINGEN* (*FWA*), *FERTILIZATION INDEPENDENT ENDOPSERM 2* (*FIS2*), and *MATERNALLY EXPRESSED PAB C-TERMINAL* (*MPC*), and reduces expression of the paternal-specific imprinted gene *PHERES* (*PHE*) [11-14]. Similarly, loss of DME activity inhibits the maternal expression of at least *FIS2* and *MPC* [13, 15], and ectopic expression of *DME* outside of the central cell is sufficient to induce expression of another maternal-specific gene, *MEDEA* (*MEA*) [16]. These observations demonstrate the importance of DNA methylation patterns in the expression of imprinted genes.

It is estimated that approximately 50 genes in *Arabidopsis* are imprinted [17]. In contrast, thousands of intergenic regions producing RNA Polymerase IV-dependent small interfering (p4-si) RNAs are imprinted and maternally expressed in the endosperm [18]. Many p4-siRNAs are produced from transposable elements, but others coincide with imprinted genes such as *FWA*, *MPC*, and *MEA* [19, 20]. Recent genome-wide analyses of DNA methylation in the endosperm further support a connection between p4-siRNA expression and imprinting of genes. Maternal chromosomes are extensively demethylated by DME at regions of p4-siRNA production [17, 21], indicating that p4-siRNAs and imprinted genes may be coordinately regulated by DME and MET1.

To examine the mechanism of p4-siRNA imprinting we investigated the genetic requirements for maternal expression and paternal silencing of p4-siRNAs. Here we show that differential DNA methylation does not explain uniparental expression of p4-siRNAs; neither do various histone modifications, including Histone H3 Lysine 27 methylation (H3K27me), establish maternal-specific expression of p4-siRNAs. Furthermore,

demethylation of maternal chromosomes by DME is dispensable for p4-siRNA expression in the endosperm.

Results

Loss of DNA methylation does not effect uniparental p4-siRNA expression in endosperm

It was previously reported that loss of DNA METHYLTRANSFERASE 1 (MET1) did not alter maternal-specific expression of p4-siRNAs in *Arabidopsis* endosperm [18]. MET1 is the primary methyltransferase in *Arabidopsis* and is responsible for maintenance of CG dinucleotide methylation [22]. Methylation at CHG sites (where H is A, T, or C) is performed by CHROMOMETHYLTRANSFERASE 3 (CMT3) and asymmetric methylation (at CHH sites) is placed by DOMAINS REARRANGED METHYLTRANFERASES (DRM1 and DRM2) [22]. To determine if non-CG DNA methylation represses paternal p4-siRNA alleles or induces expression of maternal alleles, crosses were generated between *cmt3* or *drm* mutants and wild-type plants of a differing ecotype. RNA was extracted from crossed fruits at 5 days after fertilization when p4-siRNA levels are highest, and allele-specific northern blots were performed to determine the parental origin of p4-siRNAs at locus 08002 (figure 1). Demethylation of the pollen donor did not induce p4-siRNA production paternally, indicating that CHG and CHH methylation do not repress paternal expression of p4-siRNAs. In reciprocal crosses, no change in p4-siRNA expression was detected, indicating that non-CG methylation does not promote p4-siRNA expression maternally. To determine if CHG and CHH methylation might act redundantly to repress expression, as occurs at *SUPPRESSOR OF drm1 drm2 cmt3* [23], the *drm1 drm2 cmt3* triple mutant (*ddc*) was

also used as maternal or paternal parent in inter-ecotype crosses. Maternal-specific expression was maintained even when the pollen donor lacked both CHG and CHH methylation (figure1). Although the possibility exists that MET1 acts redundantly with either CMT3 or DRM proteins when establishing uniparental DNA methylation at p4-siRNA loci, these result indicate that differential DNA methylation is not responsible for genetic imprinting of p4-siRNAs.

Various chromatin modifications do not affect p4-siRNA expression

The imprinted gene *MEDEA* (*MEA*) is a notable exception in *Arabidopsis* in that paternal expression is not repressed by MET1-mediated DNA methylation, but rather by histone H3 lysine 27 methylation (H3K27me) placed by the Polycomb Repressive Complex 2 (PRC2). Loss of PRC2 in the pollen donor triggers biparental expression of *MEA* in developing seeds [12]. To investigate the role of PRC2 in uniparental expression of p4-siRNAs, we performed crosses as above with the PRC2 mutation *fertilization independent endosperm* (*fie*). FIE is the only *Extra Sex Combs* homolog in *Arabidopsis* and this mutation lacks all potential PRC2 complexes [7]. When the *fie* allele is transmitted through pollen (from a heterozygous pollen donor) paternal *MEA* accumulates in the developing seeds [12]. However, biparental expression of p4-siRNAs was not detected when this mutation was present in the paternal lineage (figure 2).

To determine if other chromatin modifications might repress expression of p4-siRNAs from paternal chromosomes in the endosperm, several candidate genes were tested as above. *HISTONE DEACETYLASE 6* (*HDA6*) is associated with silencing of transposable elements [24] and rDNA repeats [25, 26]. *KRYPTONITE* (*KYP*) encodes a

histone methyltransferase that catalyzes dimethylation at lysine 9 of histone H3 (H3K9me), the canonical mark of silent chromatin [27]. *DECREASED DNA METHYLATION 1* (*DDM1*) and *MORPHEUS' MOLECULE 1* (*MOM1*) have similarity to SWI2/SNF2 ATPases and encode presumed nucleosome remodeling proteins. Mutations in *DDM1* eliminate DNA methylation and transcriptional silencing from transposable elements [24, 28], while loss of *MOM1* causes transcriptional reactivation of transgenes and repeated sequences without changes in DNA methylation [29, 30]. Loss of these factors did not alter the uniparental expression of p4-siRNAs in the endosperm (figure 2). Notably, these factors also were not required maternally for expression of p4-siRNAs.

Ectopic expression of DME or other family members does not induce p4-siRNA expression

Loci generating p4-siRNAs are extensively demethylated by DME in the central cell, leading to differential methylation in endosperm [17, 21]. Demethylation by DME is required for expression of *MEA*, *FWA*, and *FIS2* [10, 11, 15], and partially required for expression of *MPC* [13]. *DME* expression is also sufficient for expression of at least *MEA*, as ectopic expression of *DME* in vegetative tissue triggers *MEA* accumulation [16]. DME is part of a small family of glycosylases in *Arabidopsis* including REPRESSOR OF SILENCING (ROS1), a protein implicated in maintaining the expression of transgenes [31], and two related proteins, DEMETER-LIKE 2 (DML2) AND DEMETER-LIKE 3 (DML3) [32].

To determine if demethylation by DME or its relatives is involved in maternal

expression of p4-siRNAs, we first assayed p4-siRNA expression in transgenic lines overexpressing each glycosylase behind the nearly constitutive *35S* promoter (figure 3, figure S1) [16, 33]. Independent transgenic lines did not display ectopic expression of type I p4-siRNAs, which are normally restricted to endosperm, nor did they enhance expression of type II p4-siRNAs, which accumulate vegetatively [18]. These observations indicate that demethylation by DME or its relatives are not sufficient to trigger p4-siRNA accumulation.

To determine if demethylation acts in conjunction with endosperm-specific factors to trigger expression of p4-siRNAs, DME family overexpression lines were crossed to wild-type plants of a different ecotype and parental origin of p4-siRNAs was determined at 5 days after fertilization. If demethylation is required for expression, crosses generated with the transgenic lines as pollen donors should result in biallelic expression of p4-siRNAs. Instead, strict maternal-specific expression was detected for all crosses (figure 4), indicating that ectopic demethylation of the paternal genome by overexpression of DME family glycosylases does not induce paternal accumulation of p4-siRNAs.

DME demethylation is not required for p4-siRNA expression

To further assess the role of DME in accumulation of p4-siRNAs, we assayed p4-siRNA expression in *dme* mutant endosperm, which is not demethylated at p4-siRNA loci [17, 21]. In *dme-2* heterozygotes, seeds inheriting a maternal *dme* allele abort early in development while seeds inheriting a maternal *DME* allele develop normally. To determine if DME is required for normal accumulation of p4-siRNAs from maternal chromosomes, aborted and developed seeds were dissected from heterozygous *dme-2*

self-fertilized siliques during mid-embryo development (10-12 days post-fertilization). Unexpectedly, p4-siRNA accumulation in *dme* seeds was higher than in wild-type siblings (figure 5). To determine if this was due to lack of demethylation by DME or due to the developmental arrest of mutant seeds during an earlier period of high p4-siRNA accumulation, wild-type and *dme* seeds from the same developmental stage were analyzed. When transmitted maternally the weaker *dme-1* allele does not always trigger seed abortion, making homozygous mutant lines possible. Developing siliques from *dme-1* and wild type were collected at 5 days post anthesis and p4-siRNA accumulation was assayed (figure 5). *dme-1* siliques display slightly elevated expression of p4-siRNAs, however this may be due to over-proliferation of endosperm in the mutant compared to wild type. These results indicate that DME does not promote p4-siRNA production from maternal chromosomes in the endosperm.

Discussion

Differential methylation of maternal and paternal DNA is extensive in the endosperm of *Arabidopsis*, primarily due to DEMETER-mediated demethylation of transposable elements in the central cell [17, 21]. Many transposable elements produce p4-siRNAs, leading to the hypothesis that demethylation of these elements in the endosperm causes production of p4-siRNAs [21]. However, we show that loss of DNA methylation is insufficient for paternal p4-siRNA expression (figure 1), and loss of maternal DNA demethylation does not eliminate p4-siRNA expression (figure 5). We also demonstrate that several known histone modifications, including H3K27 and H3K9 methylation, are dispensable for p4-siRNA expression (figure 2). These data indicate that

there is an additional uniparental chromatin signal controlling p4-siRNA expression. Evidence indicates that this unknown mark is established by p4-siRNAs before fertilization because p4-siRNA expression in the female gametophyte is required for p4-siRNA expression in the developing endosperm [18].

Although there is significant overlap between regions of DME demethylation and p4-siRNA expression [17, 21], we have shown that DME demethylation is not required for p4-siRNA expression. Furthermore, p4-siRNA expression in the female gametophyte is not required for DME activity because none of the mutations that lack p4-siRNAs exhibit the developmental phenotypes associated with loss of DME activity (seed abortion and endosperm overgrowth). These data lead to the conclusion that many genomic regions, especially transposable elements, independently attract both Pol IV and DME. *FWA* is imprinted in *Arabidopsis halleri*, most likely through the action of DME at a SINE element, and yet *A. halleri FWA* lacks the tandem repeats that are required for p4-siRNA expression in *A. thaliana* [19, 34]. *A. halleri FWA* might therefore be an example of a genomic region that has recruited DME but not Pol IV. It is possible that DME and Pol IV have overlapping but independent roles in establishing parent-of-origin chromatin signatures exist across the *Arabidopsis* genome.

Parent-of-origin chromatin signals might be more prevalent than previously thought. Although imprinted expression of endogenous protein-coding genes has only been described in placental mammals and flowering plants, parent-of-origin phenomena exist throughout the animal kingdom. Some transgenes in the nematode *Caenohabditis elegans* and the zebrafish *Danio rerio* are imprinted [1, 5], and *Drosophila melanogaster* transgenes inserted near regions of heterochromatin or within the Y chromosome are also

imprinted [2, 35]. Parent-of-origin effects are not limited to uniparental gene expression. The first published case of parental “imprints” is Sciarid flies, where paternal chromosomes are eliminated from specific cell lineages [3, 36]. In coccid insects the entire paternal genome is either heterochromatinized or eliminated from somatic tissues [4], while in *C. elegans* the X chromosome adopts specific histone modifications depending on the parent of origin [37]. It seems likely that parent-of-origin chromatin signatures are widespread throughout sexual eukaryotes, and it will be interesting to discover what role small RNA-directed chromatin modification might play in establishing or responding to these signals.

Conclusions

Here we demonstrate that known modifiers of genomic imprinting, including DNA and histone methylation, do not affect the uniparental expression of p4-siRNAs in *Arabidopsis* endosperm. In particular, the DNA demethylase DME, which acts at many p4-siRNA loci, is not required for p4-siRNA expression. We therefore propose a novel mechanism controls genetic imprinting of p4-siRNAs, perhaps by establishing a genome-wide parent-of-origin chromatin signature.

Methods

Plant growth conditions and genotypes

All plants were grown under standard conditions including 16 hours of light each day. Mutant alleles were as follows. Columbia ecotype: *met1-1* [38], *drm 1-2* (SALK_031705) [39], *drm2-2* (SALK_150863) [39], *cmt3-11* (SALK_148381) [39],

hda6-9 (E. Havecker, C. Melnyk, and D. Baulcombe, unpublished allele), *ddm1-2* [40], *mom1-2* (SALK_141293) [29], and *fie* (GABI 362D08); Landsberg *erecta* ecotype: *cmt3-7* [41], and *kyp-2* [27]; Wassilewskijia ecotype: *drm1-1* [42] and *drm2-1* [42]. The *dme-1* and *dme-2* mutations were isolated in Columbia and backcrossed to Landsberg *erecta* [16]. The *drm1 drm2* double mutant contained *drm1-1* and *drm2-1*; the ddc triple mutant contained *drm1-2*, *drm2-2*, and *cmt3-11*. Wassilewskijia and C24 contain the Columbia-0 allele at locus 08002 (figure S2).

To eliminate possible self-fertilization, crosses were performed 24 hours after manual emasculation of immature flowers. For each cross, six to ten siliques were collected 5 days after fertilization. To determine the effect of the loss-of-function *dme-2* allele, *dme-2* heterozygotes were allowed to self-fertilize. The resulting seeds were dissected 10-12 days after fertilization and divided into DME+ and dme- based on development of the embryo. For analysis of the weaker *dme-1* allele, flowers were inspected daily and marked upon anthesis. Siliques were collected 5 days after anthesis.

Generation of transgenic lines

Total RNA was extracted from wild type Col-0 leaf tissue using Trizol (Invitrogen) and 1ug RNA was subjected to RQ1 DNAse (Promega) digestion for 30 minutes. First strand cDNA synthesis using Random Primers (Invitrogen) was performed using Superscript III (Invitrogen). Full length ROS1, DML2 and DML3 cDNAs were PCR amplified using Pfu Ultra (Stratagene) after reverse transcription. Individual PCR products were introduced into pENTR D-TOPO (Invitrogen) and the resulting entry vectors were recombined into pEARLEYGATE 202 (Earley, Haag et al. 2006) using LR

Clonase (Invitrogen). *Agrobacterium*-assisted transformations of the overexpression constructs into wildtype Col-0 plants were performed via floral dip (Clough and Bent 1998). Overexpression of DME-family glycosylases was verified with qRT-PCR (figure S1).

Primer sets used for cloning ROS1, DML2 and DML3 are as follows: ROS1 Fwd – CACCATGGAGAAACAGAGGAGAGAAG, ROS1 Rev – TTAGGCGAGGTTAGCTTGTTGTC; DML2 Fwd – CACCATGGAAGTGGAAGGTGAAGTG, DML2 Rev – TCATTCCTCTGTCTTCTCTTTAGTTCTG; DML3 Fwd – CACCATGTTGACAGATGGTTCACAACAC, DML3 Rev – CTATATATCATCATCACTCATAAACTTTGGCC.

RNA extraction and northern hybridizations

RNA was extracted from leaves using TRI® Reagent (Sigma-Aldrich) according to the manufacturer's protocol. RNA from crossed siliques or dissected seeds was extracted as follows: 5-6 siliques were frozen in liquid nitrogen and ground to a fine powder. 500 µL of room temperature extraction buffer (100 mM glycine pH 9.5, 10 mM EDTA, 100 mM NaCl, 2% SDS) was added and once thawed, samples were further homogenized and placed on ice. Lysates were extracted once with cold Tris-saturated phenol (pH 8.0), twice with cold 25:24:1 Tris-saturated phenol:chloroform:isoamyl alcohol, and once with cold 24:1 chloroform:isoamyl alcohol before precipitation with sodium acetate and ethanol.

Small RNA was enriched from 30-50 µg total RNA with mirVana miRNA isolation

columns (Ambion) according to the manufacturer's protocol. Small RNAs were resolved on a 7M urea/1X TBE/15% acrylamide gel (19:1 acrylamide:bisacrylamide) and transferred to Hybond N+ membrane (GE/Amersham). Membranes were UV-crosslinked before pre-hybridization in UltraHyb Oligo buffer (Ambion). Oligonucleotides were labeled with [γ-^{32}P]-ATP and T4 polynucleotide kinase and purified over an illustra MicroSpin G-25 column (GE/Amersham). After overnight hybridization with labeled oligonucleotides in UltraHyb Oligo buffer membranes were washed twice in 2X SSC, 0.1% SDS. Hybridization and washing was at 35° C. Membranes were exposed to phosphor-storage screens for detection of siRNAs.

Probe sequences are as follows (underlined bases are LNA): tRNA^{met} TCGAACTCTCGACCTCAGGAT; 08002.L1 CCCATGGTCTCAAACACATCCTCG; 08002.Ler TCAAGTGAATCTTTAGCGTATGCT; 08002.Col AGTGAATCTAGAGATTTAGCGTAT; 00687 GTTCCTCGTTCTACCCTCATACCT; 02815 CCATGTCATTCCACCCATCAAAAG; siRNA02 GTTGACCAGTCCGCCAGCCGAT; AtRep2 GCGGGACGGGTTTGGCAGGACGTTACTTAAT; Simplehat TGGGTTACCCATTTTGACACCCCTA; siRNA1003 ATGCCAAGTTTGGCCTCACGGTCT. All experiments were replicated with independent biological samples.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

RAM conceived, designed, and performed the experiments; EHT generated transgenic constructs; JS participated in experiments; RLF, CSP, and DCB designed and coordinated the experiments; RAM wrote the manuscript; all authors read and approved the final manuscript.

Acknowledgements

The authors thank the following colleagues for providing mutant or transgenic seed stocks: Ian R. Henderson (*drm1-2 drm2-2 cmt3-11*), Frédéric Berger (*fie*), Yeonee Choi (*35S::DME*), and Jian-kang Zhu (*35S::ROS1*). This work was supported by the Gatsby Charitable Foundation and the EU Sixth Framework Programme Integrated Project SIROCCO (LSHG-CT-2006-037900). RAM was supported in part by a National Science Foundation International Research Fellowship. DCB is supported as a Royal Society Research Professor.

References

1. Martin CC, Mcgowan R: **Parent-of-Origin Specific Effects on the Methylation of a Transgene in the Zebrafish, Danio-Rerio.** *Dev Genet* 1995, **17:**233-239.

2. Lloyd VK, Sinclair DA, Grigliatti TA: **Genomic imprinting and position-effect variegation in Drosophila melanogaster.** *Genetics* 1999, **151:**1503-1516.

3. Goday C, Esteban MR: **Chromosome elimination in sciarid flies.** *Bioessays* 2001, **23:**242-250.

4. Bongiorni S, Prantera G: **Imprinted facultative heterochromatization in mealybugs.** *Genetica* 2003, **117:**271-279.

5. Sha K, Fire A: **Imprinting capacity of gamete lineages in Caenorhabditis elegans.** *Genetics* 2005, **170:**1633-1652.

6. Jahnke S, Scholten S: **Epigenetic resetting of a gene imprinted in plant embryos.** *Curr Biol* 2009, **19:**1677-1681.

7. Guitton AE, Page DR, Chambrier P, Lionnet C, Faure JE, Grossniklaus U, Berger F: **Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana.** *Development* 2004, **131:**2971-2981.

8. Garnier O, Laoueille-Duprat S, Spillane C: **Genomic imprinting in plants.** *Epigenetics* 2008, **3:**14-20.

9. Xiao W, Gehring M, Choi Y, Margossian L, Pu H, Harada JJ, Goldberg RB, Pennell RI, Fischer RL: **Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase.** *Developmental Cell* 2003, **5:**891-901.

10. Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL: **DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation.** *Cell* 2006, **124:**495-506.

11. Kinoshita T, Miura A, Choi Y, Kinoshita Y, Cao X, Jacobsen SE, Fischer RL, Kakutani T: **One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation.** *Science* 2004, **303:**521-523.

12. Jullien PE, Katz A, Oliva M, Ohad N, Berger F: **Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in Arabidopsis.** *Curr Biol* 2006, **16:**486-492.

13. Tiwari S, Schulz R, Ikeda Y, Dytham L, Bravo J, Mathers L, Spielman M, Guzman P, Oakey RJ, Kinoshita T, Scott RJ: **MATERNALLY EXPRESSED PAB C-TERMINAL, a novel imprinted gene in *Arabidopsis*, encodes the conserved C-terminal domain of polyadenylate binding proteins.** *Plant Cell* 2008, **20:**2387-2398.

14. Makarevich G, Villar CB, Erilova A, Kohler C: **Mechanism of PHERES1 imprinting in Arabidopsis.** *Journal of Cell Science* 2008, **121:**906-912.

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

16. Choi Y, Gehring M, Johnson L, Hannon M, Harada JJ, Goldberg RB, Jacobsen SE, Fischer R: **DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*.** *Cell* 2002, **110:**33-42.

17. Gehring M, Bubb KL, Henikoff S: **Extensive demethylation of repetitive elements during seed development underlies gene imprinting.** *Science* 2009, **324:**1447-1451.

18. Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC: **Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*.** *Nature* 2009, **460:**283-286.

19. Chan SW-L, Zhang X, Bernatavichute YV, Jacobsen SE: **Two-step recruitment of RNA-directed DNA methylation to tandem repeats.** *PLoS Biol* 2006, **4:**1923-1933.

20. Mosher RA, Schwach F, Studholme D, Baulcombe DC: **PolIVb influences RNA-directed DNA-methylation independently of its role in siRNA biogenesis.** *Proc Natl Acad Sci U S A* 2008, **105:**3145-3150.

21. Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D: **Genome-Wide demethylation of *Arabidopsis* endosperm.** *Science* 2009, **324:**1451-1454.

22. Law JA, Jacobsen SE: **Establishing, maintaining and modifying DNA methylation patterns in plants and animals.** *Nat Rev Genet* 2010, **11:**204-220.

23. Henderson IR, Jacobsen SE: **Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading.** *Gene Dev* 2008, **22:**1597-1606.

24. Lippman Z, May B, Yordan C, Singer T, Martienssen R: **Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification.** *PLoS Biol* 2003, **1:**E67.

25. Earley K, Lawrence RJ, Pontes O, Reuther R, Enciso AJ, Silva M, Neves N, Gross M, Viegas W, Pikaard CS: **Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance.** *Gene Dev* 2006, **20:**1283-1293.

26. Earley KW, Pontvianne F, Wierzbicki AT, Blevins T, Tucker S, Costa-Nunes P, Pontes O, Pikaard CS: **Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation.** *Gene Dev* 2010, **24:**1119-1132.

27. Jackson JP, Lindroth AM, Cao X, Jacobsen SE: **Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase.** *Nature* 2002, **416:**556-560.

28. Hirochika H, Okamoto H, Kakutani T: **Silencing of retrotransposons in arabidopsis and reactivation by the ddm1 mutation.** *Plant Cell* 2000, **12:**357-369.

29. Amedeo P, Habu Y, Afsar K, Mittelsten Scheid O, Paszkowski J: **Disruption of the plant gene MOM releases transcriptional silencing of methylated genes.** *Nature* 2000, **405:**203-206.

30. Vaillant I, Schubert I, Tourmente S, Mathieu O: **MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in *Arabidopsis*.** *EMBO Rep* 2006, **in press**.

31. Gong Z, Morales-Ruiz T, Ariza RR, Roldan-Arjona T, David L, Zhu JK: **ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase.** *Cell* 2002, **111:**803-814.

32. Ortega-Galisteo AP, Morales-Ruiz T, Ariza RR, Roldan-Arjona T: **Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks.** *Plant Mol Biol* 2008, **67:**671-681.

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

34. Fujimoto R, Kinoshita Y, Kawabe A, Kinoshita T, Takashima K, Nordborg M, Nasrallah ME, Shimizu KK, Kudoh H, Kakutani T: **Evolution and control of imprinted FWA genes in the genus Arabidopsis.** *PLoS Genet* 2008, **4:**-.

35. Maggert KA, Golic KG: **The Y chromosome of Drosophila melanogaster exhibits chromosome-wide imprinting.** *Genetics* 2002, **162:**1245-1258.

36. Crouse HV: **The Controlling Element in Sex Chromosome Behavior in Sciara.** *Genetics* 1960, **45:**1429-1443.

37. Bean CJ, Schaner CE, Kelly WG: **Meiotic pairing and imprinted X chromatin assembly in Caenorhabditis elegans.** *Nat Genet* 2004, **36:**100-105.

38. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ: **Arabidopsis MET1 cytosine methyltransferase mutants.** *Genetics* 2003, **163:**1109-1122.

39. Chan SW-L, Henderson IR, Zhang X, Shah G, Chien JS-C, Jacobsen SE: **RNAi, DRD1, and histone methylation actively target developmentally important non-CG DNA methylation in *Arabidopsis*.** *PLoS Genet* 2006, **2:**0791-0797.

40. Vongs A, Kakutani T, Martienssen RA, Richards EJ: **Arabidopsis-Thaliana DNA Methylation Mutants.** *Science* 1993, **260:**1926-1928.

41. Lindroth AM, Cao XF, Jackson JP, Zilberman D, McCallum CM, Henikoff S, Jacobsen SE: **Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation.** *Science* 2001, **292:**2077-2080.

42. Cao XF, Jacobsen SE: **Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing.** *Curr Biol* 2002, **12:**1138-1144.

Figure legends

Figure 1. Loss of methylation does not induce biparental p4-siRNA production in endosperm.

Small RNAs were isolated from inter-ecotype crosses between wild type and DNA methyltransferase mutants at 5 days after fertilization; maternal parent is listed first for all crosses. Parental origin of small RNA was determined with allele-specific small RNA probes (08002.Col and 08002.Ler). 08002.L1 hybridizes to small RNAs from both alleles and is a control for small RNA production at this locus; $tRNA^{met}$ is a loading control. Small RNAs were detected specifically from maternal alleles in crosses between the wild-type ecotypes Columbia-0 (Col) and Landsberg *erecta* (Ler). Demethylation of the paternal genome through the mutations *dna methyltransferase 1* (*met1*), *chromomethyltranserase 3* (*cmt3*), and *domains rearranged methyltransferases 2* and *3* (*drm*) was not sufficient to trigger accumulation of paternal p4-siRNAs. Furthermore, loss of all non-CG methylation in the triple mutant *drm1 drm2 cmt3* (*ddc*) was insufficient to trigger paternal p4-siRNA accumulation.

Figure 2. Assorted chromatin modifications are not required for imprinted p4-siRNA production in endosperm.

Small RNAs were isolated from inter-ecotype crosses between wild type and a histone modification mutant and parental origin of small RNA was determined as described in figure 1. Accumulation of p4-siRNAs from paternal chromosomes was not induced when the Polycomb Repressive Complex 2 mutant *fertilization independent endosperm* (*fie*) was transmitted paternally. Likewise, mutations in *histone deacetylase 6*

(*hda6*), the H3K9 methyltransferase *kryptonite* (*kyp*), and the nucleosome remodeling proteins *decreased dna methylation 1* (*ddm1*) and *morpheus' molecule 1* (*mom1*) did not affect uniparental expression of p4-siRNAs.

Figure 3. *DEMETER* family glycosylases are insufficient to induce vegetative expression of p4-siRNAs.

Ectopic expression of the *DEMETER* glycosylase behind the strong, nearly constitutive *35S* promoter (*35S::DME*) does not cause ectopic accumulation of type I p4-siRNAs (00687, 02815, 08002, and siRNA 02) in leaves, nor does it alter expression of type II p4-siRNAs (AtRep2, Simplehat, and siRNA1003) in leaves. Similarly, overexpression of the related glycosylases *REPRESSOR OF SILENCING* (*35S::ROS1*), *DEMETER-LIKE 2* (*35S::DML2*), or *DEMETER-LIKE 3* (*35S::DML3*) has no affect on p4-siRNA expression. Two independent transgenic lines were assayed for each overexpression construct. *35S::ROS1* lines are in the C24 background [33]; all other lines are in the Col background [16].

Figure 4. *DEMETER* family glycosylases do not trigger paternal expression of p4-siRNAs.

Small RNAs were isolated from inter-ecotype crosses between wild type and transgenic lines and parental origin of small RNA was determined as described in figure 1. Expression of the *DEMETER* glycosylase in the male gametophyte from the strong, nearly constitutive *35S* promoter (*35S::DME*) does not trigger paternal expression of p4-siRNAs in endosperm. Similarly, overexpression of the related glycosylases

REPRESSOR OF SILENCING (*35S::ROS1*), *DEMETER-LIKE 2* (*35S::DML2*), or *DEMETER-LIKE 3* (*35S::DML3*) does not affect imprinted p4-siRNA expression in endosperm. Two independent transgenic lines were assayed for each overexpression construct. *35S::ROS1* lines are in the C24 background [33]; all other lines are in the Col background [16].

Figure 5. p4-siRNA expression in endosperm does not require DEMETER demethylation.

Left side: Developing (WT) or arrested (dme-) seeds were dissected from self-fertilized *dme-2* heterozygous fruits 10-12 days after fertilization and small RNAs were extracted. DME-deficient seeds express p4-siRNAs at levels higher than wild type, perhaps due to arrest at an earlier developmental stage or due to endosperm overgrowth. Right side: RNA was extracted from wild type and *dme-1* homozygous fruits at 5 days after anthesis and small RNAs were extracted. Mutant seeds accumulate p4-siRNAs slightly higher than wild type seeds, perhaps due to endosperm overgrowth in mutant seeds.

Supplemental Figure 1. Characterization of *DEMETER* family overexpression lines.

Transgenic lines expressing the four members of the *DEMETER* family behind the nearly constitutive *35S* promoter were assayed for transcript accumulation in leaves by quantitiative reverse transcription-PCR. Overexpression of *REPRESSOR OF SILENCING* (*ROS1*) is in the C24 ecotype [33]; all other constructs are in Columbia (Col-0) [16]. All graphs are mean values for 3 biological replicates. *35S::DME* and

35S::ROS1 lines are homozygous; *35S::DML2* and *35S::DML3* are pooled samples of homozygous and hemizygous individuals. Overexpression of *DEMETER* (*DME*) is weak, but sufficient to induce expression of *MEDEA* (*MEA*) in leaves (pink bars).

Supplemental Figure 2. The 08002 polymorphism in various *Arabidopsis* ecotypes.

The p4-siRNA locus 08002 contains a six nucleotide indel between *Arabidopsis* ecotypes Columbia (Col) and Landsberg *erecta* (Ler). This polymorphism is the basis of the allele-specific probes 08002.Col and 08002.Ler (hybridizing to the region in bold type). To determine if these probes would also bind siRNAs from other ecotypes, the 08002 region from Wassilewskijia (WS) and C24 was sequenced. These ecotypes are (Col)-like for the indel, but they also differ from Col at a single nucleotide (in red). However, this SNP does not appear to affect hybridization of the Col probe to C24 and WS siRNAs.

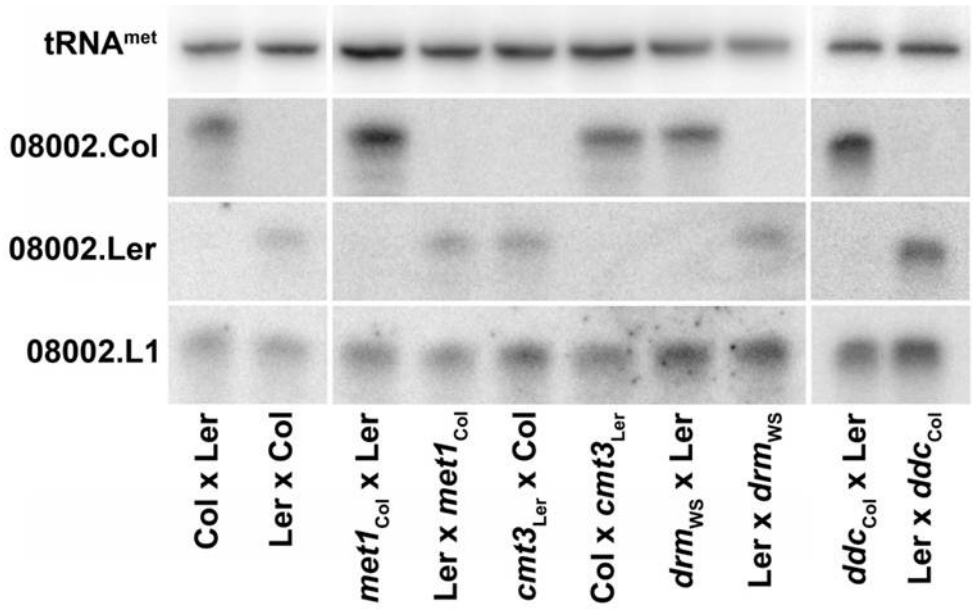

tRNA^met
08002.Col
08002.Ler
08002.L1
Col x Ler
Ler x Col
met1Col x Ler
Ler x met1Col
cmt3Ler x Col
Col x cmt3Ler
drmWS x Ler
Ler x drmWS
ddcCol x Ler
Ler x ddcCol

Figure 1

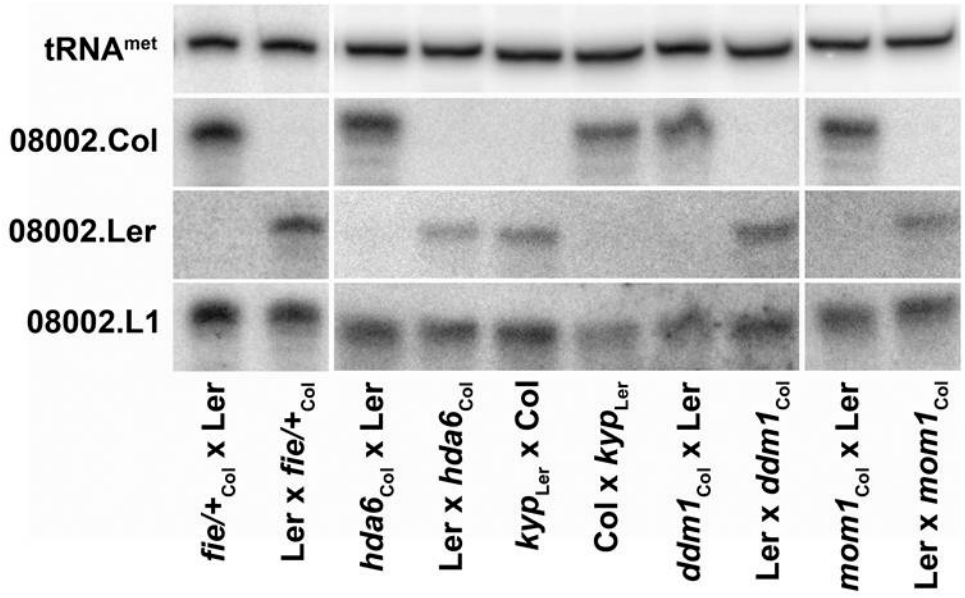

tRNA^met
08002.Col
08002.Ler
08002.L1
fie/+Col x Ler
Ler x fie/+Col
hda6Col x Ler
Ler x hda6Col
kypLer x Col
Col x kypLer
ddm1Col x Ler
Ler x ddm1Col
mom1Col x Ler
Ler x mom1Col

Figure 2

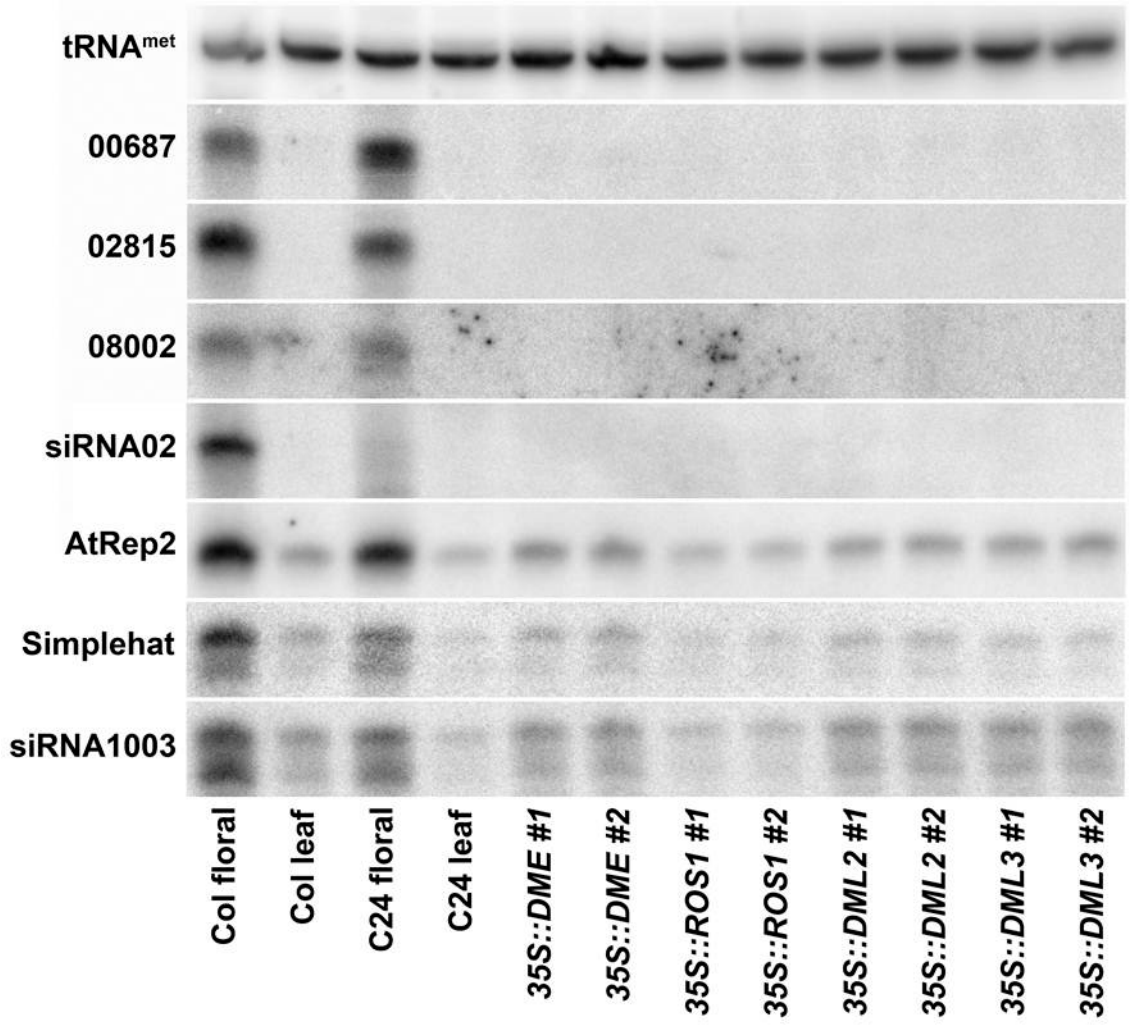

tRNA^met
00687
02815
08002
siRNA02
AtRep2
Simplehat
siRNA1003
Col floral
Col leaf
C24 floral
C24 leaf
35S::DME #1
35S::DME #2
35S::ROS1 #1
35S::ROS1 #2
35S::DML2 #1
35S::DML2 #2
35S::DML3 #1
35S::DML3 #2

Figure 3

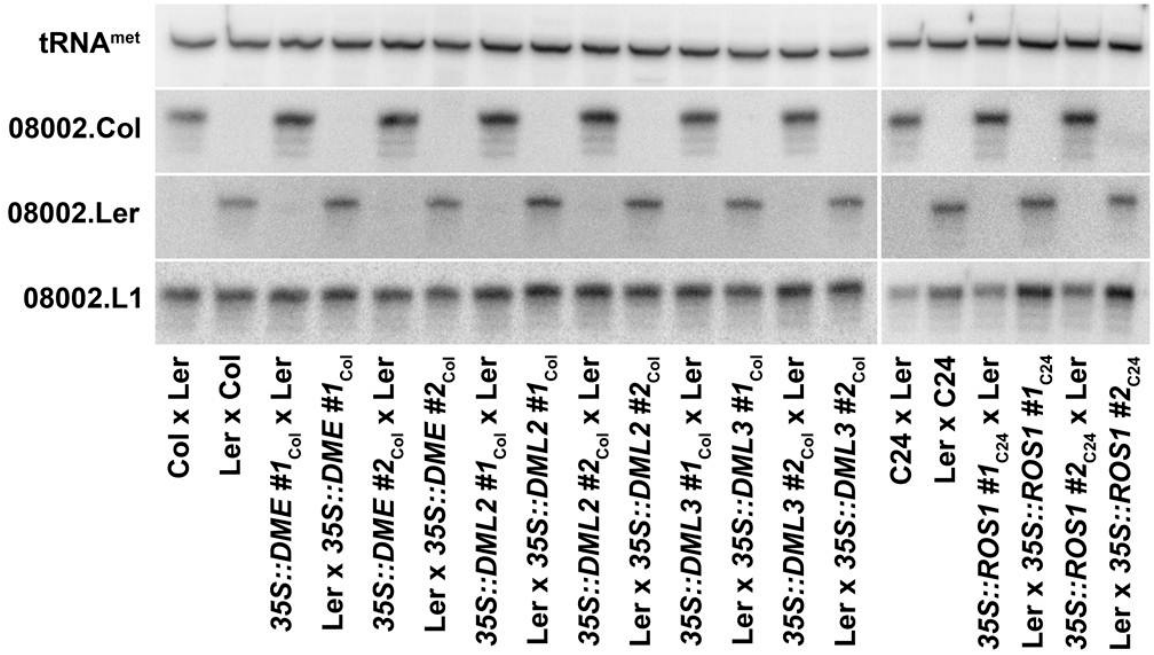

tRNA^met
08002.Col
08002.Ler
08002.L1
Col x Ler
Ler x Col
35S::DME #1Col x Ler
Ler x 35S::DME #1Col
35S::DME #2Col x Ler
Ler x 35S::DME #2Col
35S::DML2 #1Col x Ler
Ler x 35S::DML2 #1Col
35S::DML2 #2Col x Ler
Ler x 35S::DML2 #2Col
35S::DML3 #1Col x Ler
Ler x 35S::DML3 #1Col
35S::DML3 #2Col x Ler
Ler x 35S::DML3 #2Col
C24 x Ler
Ler x C24
35S::ROS1 #1C24 x Ler
Ler x 35S::ROS1 #1C24
35S::ROS1 #2C24 x Ler
Ler x 35S::ROS1 #2C24

Figure 4

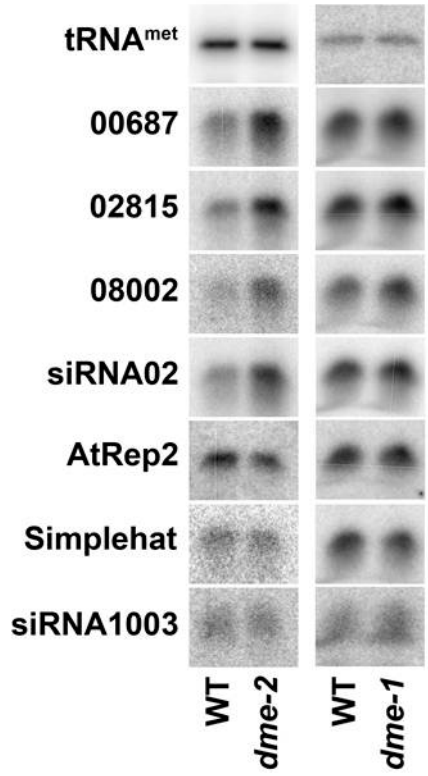

tRNA^met
00687
02815
08002
siRNA02
AtRep2
Simplehat
siRNA1003
WT
dme-2
WT
dme-1

Figure 5

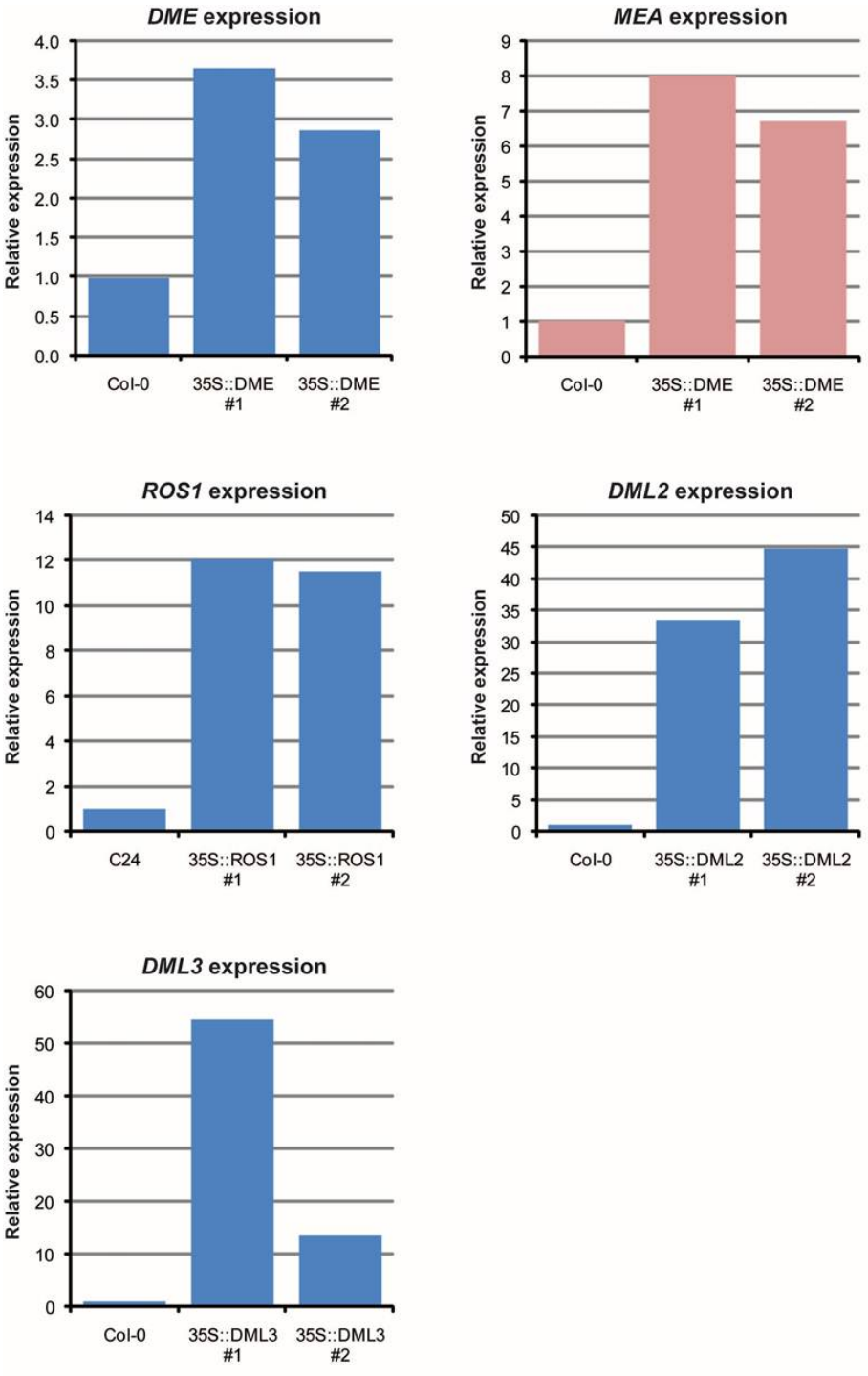

DME expression
Relative expression
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Col-0
35S::DME #1
35S::DME #2
MEA expression
Relative expression
9
8
7
6
5
4
3
2
1
0
Col-0
35S::DME #1
35S::DME #2
ROS1 expression
Relative expression
14
12
10
8
6
4
2
0
C24
35S::ROS1 #1
35S::ROS1 #2
DML2 expression
Relative expression
50
45
40
35
30
25
20
15
10
5
0
Col-0
35S::DML2 #1
35S::DML2 #2
DML3 expression
Relative expression
60
50
40
30
20
10
0
Col-0
35S::DML3 #1
35S::DML3 #2

Figure S1

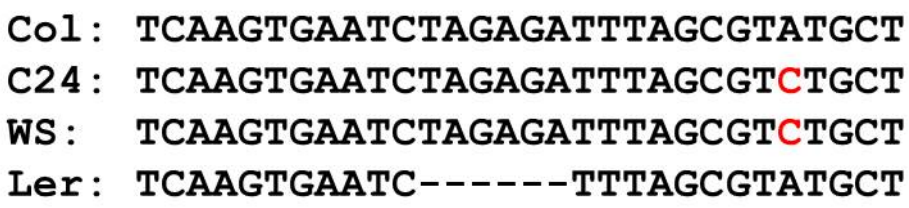

Col: TCAAGTGAATCTAGAGATTTAGCGTATGCT
C24: TCAAGTGAATCTAGAGATTTAGCGTCTGCT
WS: TCAAGTGAATCTAGAGATTTAGCGTCTGCT
Ler: TCAAGTGAATC------TTTAGCGTATGCT

Figure S2

APPENDIX C

FUNCTIONAL ANALYSIS OF NRPD1 AND NRPE1 C-TERMINAL DOMAINS REQUIRED FOR RNA DIRECTED DNA METHYLATION

My contributions to this work:

I amplified the full length NRPE1-CTD cDNA using multiple gene specific primers, and made the pENTR of clone this construct. I participated in the discussions and experimental design with regards to the NRPE1-CTD dominant negative effects with Jeremy Haag and Craig Pikaard.

Functional analysis of NRPD1 and NRPE1 C-terminal domains required for RNA-directed DNA methylation

Jeremy R. Haag, Junchen Gu, Olga Pontes, Ek Han Tan and Craig S. Pikaard[1]

Biology Department, Washington University, St. Louis, MO, 63130

[1] Corresponding author: pikaard@biology2.wustl.edu
phone 314-935-7569, fax 314-935-4432

Running title: NRPD1 and NRPE1 C-terminal domains required for RdDM

Abstract

Plant-specific RNA Polymerases IV and V are specialized forms of RNA Polymerase II and are involved in the RNA-directed DNA methylation (RdDM) pathway. The Pol IV and Pol V largest subunits, NRPD1 and NRPE1, respectively, retain the conserved DNA-dependent RNA polymerase domains A to H present in all multisubunit RNA polymerases, but lack the C-terminal heptad repeats of the Pol II largest subunit. Instead, Arabidopsis NRPD1 and NRPE1 contain unique C-terminal extensions with domains that are conserved to varying degrees among diverse plant species. Complementation assays indicate that the Defective Chloroplast and Leaves-like (DeCL-like) domain is required for full function of both NRPD1 and NRPE1. The QS-rich domain and the ten 16 aa repeats present in the NRPE1 CTD are dispensable for function, as are the majority of WG motifs implicated in AGO4 interactions. Over-expression of the NRPE1 CTD domains in wild type plants has a gain-of-function phenotype resulting in dominant suppression of RdDM.

(157 words)

Introduction

DNA-dependent RNA Polymerases (DdRPs) catalyze the production of RNA from a DNA template. Bacterial DdRP complexes have 5 core subunits, whereas eukaryotic DdRP complexes are more complex, with 12 to 17 core subunits. Pol I transcribes 45S rRNA, Pol II transcribes mRNA as well as most micro RNA precursors, and Pol III transcribes 5S rRNA and tRNAs (Grummt, 2003; Schramm and Hernandez, 2002; Woychik and Hampsey, 2002). Plants are unique in that they encode two additional DdRP complexes named Pol IV and Pol V that produce noncoding RNAs (Matzke et al., 2009).

Pol IV and Pol V are members of the RNA-directed DNA methylation (RdDM) pathway, which is important for the silencing of retrotransposons and endogenous repeats. Pol IV transcripts are precursors for small RNA biogenesis in a process that requires RNA-DEPENDENT RNA POLYMERASE2 (RDR2) and DICER-LIKE3 (DCL3) (Herr et al., 2005; Onodera et al., 2005; Pontes et al., 2006) (Chapter 5). The siRNAs associate with ARGONAUTE4 (AGO4) in a RNA-induced silencing complex (RISC) that is required for DNA methylation and the generation of secondary siRNAs at some loci (Qi et al., 2006). Pol V transcripts are hypothesized to help recruit the silencing machinery to specific chromosomal loci for DNA methylation and chromatin modifications by serving as siRNA interaction scaffolds (Wierzbicki et al., 2008; Wierzbicki et al., 2009).

The Pol II largest subunit, Rpb1, or NRPB1 in plants, contains the DdRP conserved domains A-H that are conserved in all multisubunit RNA polymerase largest subunits from bacteria to eukaryotes followed by a unique C-terminal domain (CTD)

extension (Jokerst et al., 1989). The Rpb1 CTD is composed of a heptad repeat whose consensus sequence is YSPTSPS (Allison et al., 1985). This sequence is conserved among the Pol II largest subunits of animals, plants and fungi (Stiller and Hall, 2002). The heptad repeats are a target of post-transcriptional modifications and protein-protein interactions that control Pol II initiation, elongation, termination and pre-mRNA splicing events (Cho et al., 1997; Cramer et al., 1997; Ho et al., 1998; Liao et al., 1991; McCracken et al., 1997; Nonet and Young, 1989; Otero et al., 1999; Riedl and Egly, 2000; Yamamoto et al., 2001). The total number of heptad repeats varies by species, as does the minimum number of heptad repeats required for viability (Corden, 1990). The plant-specific Pol IV and Pol V largest subunits, NRPD1 and NRPE1, respectively, are evolved from Pol II NRPB1 (Luo and Hall, 2007). They contain the core DdRP conserved domains but lack the Pol II heptad repeats at their C-termini. *Arabidopsis thaliana* NRPD1 has a CTD of 179 amino acids (aa) whereas the NRPE1 is ~370 aa, twice the length of the CTD of the Arabidopsis Pol II largest subunit, NRPB1.

The DeCL-like domain is plant-specific and has no known function. The *Arabidopsis thaliana* genome encodes five Defective Chloroplast and Leaves-like (DeCL-like) domain-containing proteins, including NRPD1 and NRPE1. AtDCL (At1g45230) is required for chloroplast rRNA processing and correct ribosome assembly (Bellaoui and Gruissem, 2004; Bellaoui et al., 2003; Keddie et al., 1996). DOMINO1 (At5g62440) is an embryo-defective mutant that is nuclear localized and proposed to be involved in a process essential for nuclear and nucleolar functions (Lahmy et al., 2004). At3g46630 remains uncharacterized but is predicted to localize to the mitochondria (Lahmy et al., 2004).

N-terminal of the NRPE1 DeCL domain is a region consisting of ten imperfect 16 amino acid repeats (aa 1451-1651) rich in WG motifs that also occur flanking the repeats (El-Shami et al., 2007; Pontier et al., 2005). WG motifs have been implicated in the binding of Argonaute proteins (El-Shami et al., 2007; Takimoto et al., 2009; Till et al., 2007) and *in vitro* and *in vivo* experiments suggest that AGO4 can interact with the NRPE1 CTD via these WG motifs (El-Shami et al., 2007; He et al., 2009; Li et al., 2006).

At its extreme C-terminus, Arabidopsis NRPE1 contains a glutamine-serine rich (QS-rich) domain (aa 1851-1976). *Spinacia oleracea* has a short proline-serine rich (PS-rich) domain at this location rather than a QS-rich domain (Pontier et al., 2005).

To address the requirements of the NRPD1 and NRPE1 C-terminal domains for Pol IV and Pol V *in vivo* function, we generated a series of deletion constructs and assayed whether or not they were capable of complementing *nrpd1* and *nrpe1* mutants defective for DNA methylation, small RNA accumulation or transcriptional silencing. My analysis reveals that the DeCL-like domains of NRPD1 and NRPE1 are required for full activity. The NRPE1 QS-rich domain is dispensable, as is the domain consisting of the ten 16 aa repeats. Contrary to a previously published report, the NRPE1 WG motifs are not fully required for Pol V activity, as deletion mutants are capable of partial complementation. Over-expression of the NRPE1 CTD leads to dominant suppression of the RdDM pathway in transformed wild type plants. Collectively, these genetic studies show that the NRPD1 and NRPE1 CTDs play an important role in Pol IV and Pol V function.

Results

NRPD1 and NRPE1 CTDs have conserved domains among diverse plant species

Predicted full-length NRPD1 and NRPE1 sequences from diverse plant species were analyzed to determine the extent of CTD conservation. The DeCL-like domain is detected by the presence of the DFSYRK consensus sequence (Bellaoui and Gruissem, 2004; Bellaoui et al., 2003) and is present in all NRPD1 and NRPE1 proteins, with the exception of the NRPD1 and one of two NRPE1 proteins in *Physcomitrella patens* (Figure S1, S2 and S3). In the context of NRPE1, the DeCL-like domain is typically C-terminal of the 16 aa repeats and WG motifs. The NRPE1 16 aa repeats are imperfect and vary in number and length in different species (Figures S1 and S2). While the WG motifs are often embedded in the repeat sequence, exceptions do occur such as the *Physcomitrella patens, Vitis vinifera, Oryza sativa* and *Zea mays* NRPE1 proteins (Figures S1 and S2). The number of WG motifs and whether they are predominantly present as WG, GW, GWG or WGW motifs varies by species (Figures S1 and S2). The QS- and PS-rich domains appear unique to Arabidopsis and spinach, respectively, as no equivalent domains were detected in NRPE1 of other plants (Figures S1 and S2).

NRPE1 C-terminal domain deletions

The Arabidopsis NRPE1 CTD can be divided into four domains: a linker region that connects the CTD to the DdRP core, the 16 aa repeat and WG motif-containing domain, the DeCL-like domain and the QS-rich domain. To test for NRPE1 CTD functions, a series of six C-terminal deletion constructs and a full-length control construct were transformed into the *nrpe1* mutant to assay for complementation (Figure 1A). Each of the HA-tagged transgenes is expressed and encodes a protein of the predicted

molecular mass (Figure 1B). NRPE2 co-immunoprecipitates with all of the NRPE1 CTD deletion constructs, even when the entire CTD is deleted, suggesting that the CTD is not required for Pol V subunit assembly (Figure 1B). NRPE1 is typically detected on immunoblots as a doublet regardless of whether the native protein or C-terminal FLAG or HA epitope tagged proteins are detected (Pontes et al., 2006; Pontier et al., 2005; Ream et al., 2009). This banding pattern is observed in each of the C-terminal deletion constructs except for the full CTD deletion construct.

The NRPE1 DeCL-like domain is required for *in vivo* complementation

It has previously been determined that Pol IV and Pol V are required for DNA methylation and silencing of the *AtSN1* retrotransposon locus (Herr et al., 2005; Kanno et al., 2005; Onodera et al., 2005; Pontier et al., 2005). DNA methylation at the *AtSN1* locus was analyzed by chop-PCR using the methylation sensitive *Hae*III restriction enzyme (Figure 1C). If the *Hae*III restriction sites in the *AtSN1* locus are methylated, DNA digestion will not occur and a PCR product will be obtained. If any of the *Hae*III restriction sites are unmethylated, the DNA will be digested and PCR amplification of the region will fail. PCR amplification of the region was successful in the NRPE1 full-length and NRPE1 Δ1851-1976 (QS-rich deletion) lines indicating these constructs successfully complement the *nrpe1* mutant and facilitate the methylation of the *Hae*III sites. The NRPE1 Δ1736-1976 protein (DeCL-like and QS-rich domain deletions) and remaining CTD deletions in the series fail to rescue *AtSN1* DNA methylation; a PCR product was not obtained, indicating that one or more *Hae*III sites was susceptible to digestion. RT-PCR analysis demonstrates *AtSN1* transcript repression in the NRPE1 full-length and

NRPE1 Δ1851-1976 lines and a failure to repress in the NRPE1 Δ1736-1976 and remaining CTD deletions (Figure 1C). DNA methylation analysis at the *5S* rDNA loci supports these results as Southern blot analysis of *Hae*III and *Hpa*II genomic DNA reveals that only the NRPE1 full-length and NRPE1 Δ1851-1976 lines complement the DNA methylation defect of the *nrpe1* mutant (Figure 1D).

While NRPE1 is not absolutely required for the biogenesis of all siRNAs, *nrpe1* mutants do affect the accumulation of some siRNAs (Mosher et al., 2008). Small RNA Northern blot analysis of AtCopia, 45S rRNA and AtSN1 sequences demonstrates the QS-rich domain is dispensable for complementation but that the DeCL-like domain is required for wild-type levels of siRNA accumulation to occur (Figure 1E).

NRPD1 DeCL-like domain deletion

The Arabidopsis NRPD1 CTD is composed of a DeCL-like domain and a small linker region that connects it to the DdRP core structure. A NRPD1 DeCL-like deletion construct, NRPD1 Δ1337-1453, as well as the previously published NRPD1 full-length control were transformed into the *nrpd1* mutant to determine if the NRPD1 DeCL-like domain is required for *in vivo* complementation (Figure 2A). The two FLAG-tagged NRPD1 constructs are both expressed at the protein level, and NRPD2 and RDR2 both co-immunoprecipitate with WT or ΔCTD proteins at equivalent levels (Figure 2B). These results suggest the NRPD1 DeCL-like domain is not required for Pol IV complex assembly or for mediation of the Pol IV-RDR2 interaction (Chapter 5).

The NRPD1 DeCL-like domain is required for siRNA biogenesis and transcript silencing but not DNA methylation

At *AtSN1*, the NRPD1 DeCL deletion mutant, NRPD1 Δ1337-1453, restores DNA methylation to the same levels as the NRPD1 full-length transgene (Figure 2C). Similar results were observed at the *5S* rDNA loci by Southern blot analysis of *Hae*III and *Hpa*II digested DNA (Figure 2D).

In contrast to the NRPD1 DeCL domain being dispensable for the restoration of DNA methylation, small RNA Northern blot analysis reveals that the NRPD1 DeCL-like domain is required for the wild-type accumulation of AtCopia, 45S and AtSN1 siRNAs (Figure 2E). Consistent with the failure to produce Pol IV-dependent siRNAs, it is found that the NRPD1 DeCL-like domain is required for suppression of *AtSN1* and *solo LTR* transcripts (Figure 2F).

NRPE1 CTD repeats are dispensable for *in vivo* complementation

Given the functional requirement for the NRPE1 DeCL-like domain, we were unable to conclude the significance of domains N-terminal to this domain using the C-terminal deletion series studied in Figure 1. To address the requirement for sequence elements between the NRPE1 DdRP core and the DeCL-like domain, three additional transgene deletion constructs were engineered and transformed into the *nrpe1* mutant for *in vivo* complementation assays (Figure 3A). NRPE1 Δ1251-1426 contains a deletion in the linker region and deletes 3 of 18 WG motifs; NRPE1 Δ1426-1651 deletes the ten 16 aa repeats and 13 of the 18 WG motifs, and NRPE1 Δ1251-1651 deletes both regions and 16 of the 18 WG motifs.

The three NRPE1 internal CTD deletion lines were analyzed for rescue of DNA methylation at the *5S* rDNA loci by Southern blot analysis of *Hae*III and *Hpa*II digested genomic DNA (Figure 3B). Deletion of the linker region (NRPE1 Δ1251-1426) or the ten 16 aa repeats (NRPE1 Δ1426-1651) resulted in full rescue of the *nrpe1* mutant. Only when these two regions were deleted together (NRPE1 Δ1251-1651) was there a failure to fully complement, although DNA methylation levels are still increased relative to the *nrpe1* mutant. DNA methylation at *AtSN1* was also assayed by chop-PCR and similar results were observed with DNA methylation fully restored with the NRPE1 Δ1251-1426 and NRPE1 Δ1426-1651 transgenes and only partially with the NRPE1 Δ1251-1651 transgene (Figure 3C).

In agreement with the *AtSN1* DNA methylation status, *AtSN1* transcription detected by RT-PCR demonstrates that only the NRPE1 Δ1251-1651 transgenic line continues to express *AtSN1* transcripts, though below *nrpe1* mutant levels (Figure 3D). Unexpectedly, there are no observable defects in siRNA accumulation in any of the three deletion lines (Figure 3E).

The NRPE1 WG motifs are important but not required for NRPE1 function

It has previously been published that the NRPE1 WG motifs are required for *in vivo* complementation of *5S* rDNA and *AtSN1* DNA methylation states in the *nrpe1-11* background (El-Shami et al., 2007). The NRPE1 transgene used in the study, NRPE1 ΔSD, had two deletions spanning aa 1411 to 1707 and aa 1875 to 1976. The transgene therefore deleted all ten 16 aa repeats, 16 of the 18 WG motifs and the QS-rich domain (Figure 4A).

Three independent NRPE1 ΔSD lines were compared side-by-side with the NRPE1 Δ1251-1426, NRPE1 Δ1426-1651, NRPE1 Δ1251-1651 and NRPE1 Δ1251-1976 deletion lines. Contrary to the published results (El-Shami et al., 2007), the NRPE1 ΔSD line does partially rescue DNA methylation at the *AtSN1* (Figure 4B) and *5S* rDNA loci (Figure 4C). NRPE1 ΔSD DNA methylation levels are roughly equivalent to the NRPE1 Δ1251-1651 transgenic line. The two do not display full complementation but they do facilitate significantly more DNA methylation than the *nrpe1* mutant. Transcription from the *AtSN1* and *solo LTR* loci in NRPE1 ΔSD and NRPE1 Δ1251-1651 lines is partially suppressed (Figure 4D) in agreement with the DNA methylation results, showing increased methylation at these loci. Thus, the WG motifs may be important, but they are not required for NRPE1 to complement an *nrpe1* mutant.

Over-expression of the NRPE1 C-terminal domains dominantly suppresses the RdDM pathway

Having analyzed loss-of-function phenotypes with CTD deletions in the NRPD1 and NRPE1 proteins, we next tested for gain-of-function phenotypes. If the CTDs are a platform for protein-protein interactions, over-expression may titrate away silencing factors required for RdDM function. A YFP over-expression vector encoding NRPE1 aa 1234-1842, referred to as YFP-CTD (Figure 5A), was transformed into wild type Arabidopsis plants. In whole mounted Arabidopsis roots, the protein signal is detected throughout the nucleoplasm, with little to no cytoplasmic localization detected (Figure 5B). *AtSN1* DNA methylation, in ten of twelve independent transgenic lines, is reduced compared to wild type plants (Figure 5C) demonstrating that the transgene is capable of

dominant suppression of RdDM. *AtSN1* transcription is correspondingly activated in the lines that have reduced DNA methylation (Figure 5D). Lack of transgene RNA expression in line 182 (Figure 5D) explains why there is no dominant suppression phenotype in this plant. Because the transgene is expressed in line 172, a post-transcriptional gene silencing mechanism or mutation that prevents the protein from being translated or functioning properly may explain the lack of a dominant negative phenotype in this plant. Similar to *nrpe1* mutants, AtCopia, 45S and AtSN1 siRNA accumulation is reduced in the YFP-CTD transgenic lines (Figure 5E) and these plants also display delayed flowering (Figure S4) similar to *nrpe1* mutants.

In an attempt to narrow down the region(s) capable of inducing dominant suppression of RdDM, three additional NRPE1 constructs were cloned, spanning aa 1426-1651, aa 1426-1851 and aa 1851-1977, in addition to the NRPD1 DeCL domain, aa 1337-1453 (Figure 5A). These cDNAs were recombined into over-expression vectors that add an N-terminal FLAG tag and transformed into wild type Arabidopsis plants. Protein blot analysis of immunoprecipitated protein samples confirmed expression of all the transgenes (Figure 5F).

Six independent lines for each transgene were analyzed for dominant suppression of the RdDM pathway. DNA methylation at the *AtSN1* locus was only marginally affected in three of the NRPD1 aa1337-1453 lines (Figure 5G). In contrast, multiple individuals for each of the three NRPE1 CTD over-expression constructs demonstrated significantly reduced *AtSN1* DNA methylation (Figure 5G). Corresponding with the DNA methylation results, transcription of *AtSN1* and *solo LTR* retroelements was activated in the NRPE1 CTD over-expression lines (Figure 5H). Weak expression of

AtSN1 is detected in several of the NRPD1 aa1337-1453 transgenic lines, although *solo LTR* expression does not appear to be activated (Figure 5H).

Discussion

Our results show that the DeCL-like domain is required *in vivo* for both Pol IV and Pol V function. NRPE1 is completely dependent upon this domain for function in the RdDM pathway, while NRPD1 requires the domain for complementation of siRNA biogenesis and suppression of retroelement transcription. Interestingly, DNA methylation is rescued despite deletion of the NRPD1 DeCL-like domain. Over-expression of the NRPD1 DeCL-like domain led to only subtle dominant negative DNA methylation defects, although release of transcriptional silencing was more pronounced, in agreement with the complementation assay results. In addition, the NRPD1 aa 1337-1453 lines displayed leaf curling and smaller plant size (Figure S5) similar to some of the reported phenotypes of plants over-expressing a plastid DeCL-like domain-containing protein, AtDCL (Bellaoui and Gruissem, 2004). The RdDM-defective phenotypes observed in the NRPD1 DeCL-like domain over-expression lines might be due to dominant-negative crosstalk with the three other DeCL-like domain containing proteins in Arabidopsis since *nrpd1* and *nrpe1* mutants lack these morphological phenotypes.

The QS-rich domain and ten 16 aa repeats in the NRPE1 CTD are not required for complementation of an *nrpe1* mutant, but each domain is sufficient to trigger dominant suppression of RdDM when over-expressed. The plants have no apparent morphological defects (data not shown). We suggest that the over-expressed domains either titrate away interacting proteins from the endogenous NRPE1 protein or in some other way interfere

with the function of the RdDM pathway. In agreement with this idea is the observation the YFP-tagged NRPE1 CTD localizes to the nucleus where other members of the RdDM pathway localize (Pontes et al., 2006). Interestingly, YFP-CTD was never observed in the nucleolus-associated Cajal body where siRNA biogenesis and processing are believed to occur (Li et al., 2006; Pontes et al., 2006), unlike the full-length NRPE1, suggesting the DdRP core is required for NRPE1 to localize here.

The NRPD1 Δ1337-1453 and NRPE1 Δ1251-1651 phenotypes are noteworthy since there is a breakdown in correlation between DNA methylation and siRNA production. In the case of NRPD1 Δ1337-1453, DNA methylation is rescued despite the failure to restore siRNA production, and in the case of NRPE1 Δ1251-1651, siRNA production is rescued despite the failure to restore DNA methylation. Neither restores retroelement transcript suppression. These results suggest siRNA production and DNA methylation are unable to establish a transcriptionally silenced state independent of one another. Building upon this idea, there may be two parallel pathways in plants that converge on the same target that are both required for the establishment of silencing. Perhaps DNA methylation provides an independent check on the siRNA-mediated silencing pathway in plants, and vice versa. At the very least, the results imply that Pol V-directed DNA methylation is important for transcriptional silencing but not Pol V-derived siRNAs and that Pol IV-derived siRNAs are important for transcriptional silencing but not Pol IV-directed DNA methylation.

In disagreement with a previously published report (El-Shami et al., 2007), the majority of NRPE1 WG motifs can be deleted and still largely complement the *nrpe1* mutant (Figure 4). This suggests that the WG motifs are important but not required for

Pol V function. Reports of *in vitro* interaction between bacterially expressed NRPE1 CTD protein and AGO4 in plant extracts (El-Shami et al., 2007; He et al., 2009; Li et al., 2006) have been confirmed (Figure S6) and demonstrate that AGO4 is capable of binding NRPE1 aa 1426-1651 but not a NRPE1 CTD construct that lacks this region. However, if NRPE1 and AGO4 do directly interact via the WG motifs *in vivo*, this interaction is not required for the RdDM pathway to function because the NRPE1 Δ1426-1651 line fully complements the *nrpe1* mutant. It must be stated that despite repeated efforts, the reported *in vivo* interaction between NRPE1 and AGO4 (Li et al., 2006) cannot be confirmed despite numerous co-IP approaches (Figure S7) and mass spec analysis of both NRPE1 and AGO4 purified samples (Haag, Ream, Pikaard, EMSL, unpublished). Thus, if NRPE1 and AGO4 do interact *in vivo*, it is possibly a weak or transient interaction mediated by AGO4 binding of Pol V transcripts (Wierzbicki et al., 2009) with the WG motifs acting to help stabilize the interaction.

While the NRPD1 and NRPE1 CTDs have little resemblance to the CTD of NRPB1, the Pol IV and Pol V complexes are evolutionarily derived from Pol II (Luo and Hall, 2007; Ream et al., 2009) and like Pol II, Pol IV and Pol V require distinct C-terminal domains for proper function. It is likely that the unique roles of these related polymerases arise from differential use of Pol II-derived small subunits (Ream et al., 2009) and their unique CTD architectures. Whether the CTDs play a role in regulating Pol IV and Pol V transcription or post-transcriptionally process Pol IV and Pol V transcripts is still an open question. The NRPD1 and NRPE1 CTDs are likely to be involved in protein-protein interactions and may be the target of post-translational modifications, like the NRPB1 CTD. Evidence for alternative splicing or post-

translational modification of the NRPE1 CTD is hinted at by the observation that the NRPE1 doublet pattern is lost when the full CTD is deleted (Figure 1B) and the over-expressed NRPE1 QS-rich domain migrates much larger than the predicted 14kD size (Figure 5F). Proteomic analyses to identify protein-protein interactions and post-translational modifications in the NRPD1 and NRPE1 CTDs are currently underway.

Materials and Methods

Plant materials. *Arabidopsis thaliana* mutant lines *nrpd1-3*, *nrpd2* (*nrpd2a-2, nrpd2b-1*) and *nrpe1-11* have been described previously (Onodera et al., 2005; Pontier et al., 2005), as have transgenic lines *NRPD1-FLAG* (*nrpd1-3*) and *NRPD1*$^{DDD\text{-}AAA}$*-FLAG* (*nrpd1-3*) (Haag et al., 2009; Pontes et al., 2006). The NRPE1 ΔSD-FLAG (*nrpe1-11*) transgenic line was kindly provided by Thierry Lagrange.

Cloning, vectors and transgenic lines. The pENTR-NRPE1 full-length genomic sequence with its endogenous promoter (Pontes et al., 2006) was recombined into pEarleyGate301 (Earley et al., 2006) using LR Clonase (Invitrogen) in order to add a C-terminal HA epitope tag in lieu of the normal stop codon. C-terminal domain deletions were obtained by using pENTR-NRPD1 and pENTR-NRPE1 full-length genomic clones with endogenous promoters (Pontes et al., 2006) as the DNA template and reverse primers that truncated the 3' end (Table S1). Pfu Ultra (Stratagene) was used to amplify the sequences. The PCR products were gel purified and cloned into pENTR-TOPO S/D (Invitrogen) before being recombined into pEarleyGate 301 (NRPE1 C-terminal truncations with HA epitope) or pEarleyGate302 (NRPD1 C-terminal truncation with

FLAG epitope). Internal C-terminal domain deletions were obtained by the SLIM method (Chiu et al., 2004) using the pENTR-NRPE1 full-length genomic clone as the DNA template and the appropriate primers (Table S1). Constructs were recombined into pEarleyGate301. CTD over-expression lines were generated by cloning *NRPD1* and *NRPE1* cDNA sequences (Table S1) and recombining into pEarleyGate104 (35S promoter with N-terminal YFP fusion) or pEarleyGate202 (35S promoter with N-terminal FLAG epitope). pEarleyGate plasmids in *Agrobacterium tumefaciens* strain GV3101 were used to transform *Arabidopsis thaliana* (Col-0) plants by the floral dip method (Bechtold and Pelletier, 1998) as modified by Clough and Bent (Clough and Bent, 1998). The NRPD1 and NRPE1 genomic clones were transformed into *nrpd1-3* and *nrpe1-11*, respectively, while the over-expressed cDNA clones were transformed into wild type plants. T1 seeds were sown on soil and transformants were selected by spraying 2-week old seedlings with BASTA herbicide. NRPE1 ΔSD-FLAG transformants were selected as described previously (El-Shami et al., 2007).

DNA methylation analysis. Southern blot analysis of *Hae*III and *Hpa*II digested DNA at the 5S rDNA locus was performed as in (Haag et al., 2009). The AtSN1 DNA methylation assay involving PCR amplification of undigested or *Hae*III-digested genomic DNA was performed as previously described (Herr et al., 2005).

RNA analysis. Small RNA was isolated and analyzed as previously described (Haag et al., 2009). RT-PCR was performed as previously described (Haag et al., 2009) using primers in Table S1.

Antibodies. Affinity purified anti-NRPD2 and anti-RDR2 have been described previously (Haag et al., 2009; Onodera et al., 2005). Anti-FLAG M2-HRP and anti-HA are commercially available (Sigma).

Immunoprecipitation and immunoblotting. Frozen leaf tissue (4.0g) was ground in mortar and pestle and protein extracted as in (Pontes et al., 2006). Supernatant was incubated with 35uL anti-FLAG-M2 or anti-HA resin (Sigma) for 3 hours at 4 °C on a rotating mixer. Resin was washed two times with extraction buffer supplemented with 0.5% NP-40. Washed immunoprecipitates were eluted from the resin with two bed volumes of 2x SDS sample buffer and boiled 5 min. Protein samples were run on Tris-glycine gels by SDS-PAGE and transferred to nitrocellulose or PVDF membrane. Antibodies were diluted in TBST + 5% (w/v) nonfat dried milk (Schnucks) as follows: 1:500 NRPD2, 1:250 anti-RDR2, 1:3,000 anti-HA and 1:2,000 anti-FLAG-HRP. 1:5,000 to 1:10,000 anti-rabbit-HRP (Amersham) was used as secondary antibody. ECL Plus (GE Healthcare) was used for chemiluminescent detection of proteins. Membranes were stripped with 1% SDS, 25 mM glycine, pH 2.0 and re-equilibrated with TBST prior to subsequent blocking and immunoblotting.

Whole mount localization. Whole roots were fixed in 4% formaldehyde in PBS, pH 7.4 for 20 min at room temperature and washed in 1X PBS, pH 7.4 at room temperature. Nuclei were stained with 2.5 ug/ml propidium iodide (Invitrogen) and observed with Leica SP2 confocal microscope using 488 nm and 561 nm laser lines.

Acknowledgements

JRH and CSP designed the study and wrote the paper. EHT cloned the pENTR-NRPE1 aa 1234-1842 cDNA. OP performed the microscopy for Figure 5B. JG helped screen the CTD over-expression lines and prepared Figures 5F, G and H. JRH performed all remaining experiments. We thank Mike Dyer and the greenhouse staff for expert plant care, members of the Pikaard lab for helpful discussion and Tom Ream for proofing the manuscript. The National Institutes of Health grant GM077590 supported this research. The Pontes laboratory is supported by a grant from the Edward Mallinckrodt Foundation. The authors have declared that no competing interests exist.

References

Allison, L.A., Moyle, M., Shales, M., and Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. Cell, 42(2), 599-610.

Bechtold, N., and Pelletier, G. (1998) In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. Methods Mol Biol, 82, 259-266.

Bellaoui, M., and Gruissem, W. (2004) Altered expression of the Arabidopsis ortholog of DCL affects normal plant development. Planta, 219(5), 819-826.

Bellaoui, M., Keddie, J.S., and Gruissem, W. (2003) DCL is a plant-specific protein required for plastid ribosomal RNA processing and embryo development. Plant Mol Biol, 53(4), 531-543.

Chiu, J., March, P.E., Lee, R., and Tillett, D. (2004) Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res, 32(21), e174.

Cho, E.J., Takagi, T., Moore, C.R., and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev, 11(24), 3319-3326.

Clough, S.J., and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J, 16(6), 735-743.

Corden, J.L. (1990) Tails of RNA polymerase II. Trends Biochem Sci, 15(10), 383-387.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997) Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A, 94(21), 11456-11460.

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.

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.

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.

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.

He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., and Zhu, J.K. (2009) An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. Cell, 137(3), 498-508.

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.

Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. J Biol Chem, 273(16), 9577-9585.

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(2), 266-275.

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.

Keddie, J.S., Carroll, B., Jones, J.D., and Gruissem, W. (1996) The DCL gene of tomato is required for chloroplast development and palisade cell morphogenesis in leaves. EMBO J, 15(16), 4208-4217.

Lahmy, S., Guilleminot, J., Cheng, C.M., Bechtold, N., Albert, S., Pelletier, G., Delseny, M., and Devic, M. (2004) DOMINO1, a member of a small plant-specific gene family, encodes a protein essential for nuclear and nucleolar functions. Plant J, 39(6), 809-820.

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.

Liao, S.M., Taylor, I.C., Kingston, R.E., and Young, R.A. (1991) RNA polymerase II carboxy-terminal domain contributes to the response to multiple acidic activators in vitro. Genes Dev, 5(12B), 2431-2440.

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.
Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A.J. (2009) RNA-mediated chromatin-based silencing in plants. Curr Opin Cell Biol, 21(3), 367-376.
McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., and Bentley, D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. Nature, 385(6614), 357-361.
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.
Nonet, M.L., and Young, R.A. (1989) Intragenic and extragenic suppressors of mutations in the heptapeptide repeat domain of Saccharomyces cerevisiae RNA polymerase II. Genetics, 123(4), 715-724.
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.
Otero, G., Fellows, J., Li, Y., de Bizemont, T., Dirac, A.M., Gustafsson, C.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (1999) Elongator, a multisubunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. Mol Cell, 3(1), 109-118.
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, 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.
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.
Riedl, T., and Egly, J.M. (2000) Phosphorylation in transcription: the CTD and more. Gene Expr, 9(1-2), 3-13.
Schramm, L., and Hernandez, N. (2002) Recruitment of RNA polymerase III to its target promoters. Genes Dev, 16(20), 2593-2620.
Stiller, J.W., and Hall, B.D. (2002) Evolution of the RNA polymerase II C-terminal domain. Proc Natl Acad Sci U S A, 99(9), 6091-6096.
Takimoto, K., Wakiyama, M., and Yokoyama, S. (2009) Mammalian GW182 contains multiple Argonaute-binding sites and functions in microRNA-mediated translational repression. RNA, 15(6), 1078-1089.

Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M.W., and Ladurner, A.G. (2007) A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. Nat Struct Mol Biol, 14(10), 897-903.

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

Yamamoto, S., Watanabe, Y., van der Spek, P.J., Watanabe, T., Fujimoto, H., Hanaoka, F., and Ohkuma, Y. (2001) Studies of nematode TFIIE function reveal a link between Ser-5 phosphorylation of RNA polymerase II and the transition from transcription initiation to elongation. Mol Cell Biol, 21(1), 1-15.

Figure Legends

Figure 1. The NRPE1 DeCL-like domain is required for *nrpe1 in vivo* complementation. (A) Genomic HA-epitope tagged *NRPE1* C-terminal domain deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a "Δ" represent deletions. (B) Western blot analysis of HA-immunoprecipitated NRPE1 proteins from whole plant extracts and co-immunoprecipitated NRPE2. (C) Agarose gel results of chop-PCR DNA methylation assay and transcript expression at the *AtSN1* retroelement. (D) 5S rDNA methylation analysis by Southern blot of *Hae*III and *Hpa*II digested genomic DNA. (E) Northern blot analysis of AtCopia, 45S rRNA, miR171 and AtSN1 small RNAs with images of ethidium bromide (EtBr) stained gels below.

Figure 2. The NRPD1 DeCL-like domain is required for *nrpd1 in vivo* complementation. (A) Genomic FLAG-epitope tagged *NRPD1* C-terminal domain deletion transformed into *nrpd1-3* mutant background. Black colored regions denoted with a "Δ" represent deletions. (B) Western blot analysis of FLAG-immunoprecipitated NRPD1 proteins from whole plant extracts with co-immunoprecipitated RDR2 and NRPD2. (C) *AtSN1* chop-PCR DNA methylation assay. (D) 5S rDNA methylation analysis by Southern blot of *Hae*III and *Hpa*II digested genomic DNA. (E) Northern blot analysis of AtCopia, 45S rRNA, miR171 and AtSN1 small RNAs with images of ethidium bromide (EtBr) stained gels below. (F) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls.

Figure 3. The NRPE1 repetitive elements and majority of WG motifs are not required for *nrpe1* complementation. (A) Genomic HA-epitope tagged *NRPD1* CTD internal deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a "Δ" represent deletions. (B) 5S rDNA methylation analysis by Southern blot of *Hae*III and *Hpa*II digested genomic DNA. (C) *AtSN1* chop-PCR DNA methylation assay. (D) RT-PCR analysis of *AtSN1* transcription with *actin* and no RT controls. (E) Northern blot analysis of 5S rRNA, AtCopia, 45S rRNA and miR163 small RNAs with image of ethidium bromide (EtBr) stained gel below.

Figure 4. The NRPE1 WG motifs are important but not required for *nrpe1 in vivo* complementation. (A) Genomic HA-epitope tagged *NRPD1* CTD internal deletion series transformed into *nrpe1-11* mutant background. Black colored regions denoted with a "Δ" represent deletions. (B) *AtSN1* chop-PCR DNA methylation assay. (C) 5S rDNA methylation analysis by Southern blot of *Hae*III and *Hpa*II digested genomic DNA. (D) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls.

Figure 5. Over-expression of the NRPE1 CTD dominantly suppresses the RdDM pathway. (A) 35S promoter driven N-terminally tagged cDNA constructs transformed into wild type *Arabidopsis thaliana*. (B) Whole mount localization of YFP-CTD in Arabidopsis root with enlargements of a single nucleus showing YFP signal, propidium iodide (PI) signal for stained DNA, and overlayed images. (C) *AtSN1* chop-PCR DNA methylation assay with YFP-CTD transformants. (D) RT-PCR analysis of *YFP-CTD*

transgene and *AtSN1* transcription with *actin* and no RT controls. (E) Northern blot analysis of AtCopia, 45S rRNA, miR171 and AtSN1 small RNAs with images of ethidium bromide (EtBr) stained gels below. (F) Western blot analysis of immunoprecipitated over-expressed FLAG epitope tagged NRPE1 and NRPD1 CTD protein domains. An arrow denotes predicted full-length proteins. (G) *AtSN1* chop-PCR DNA methylation assay of over-expressed CTD domains. (H) RT-PCR analysis of *AtSN1* and *solo LTR* transcription with *GAPA* and no RT controls in over-expressed CTD transformants.

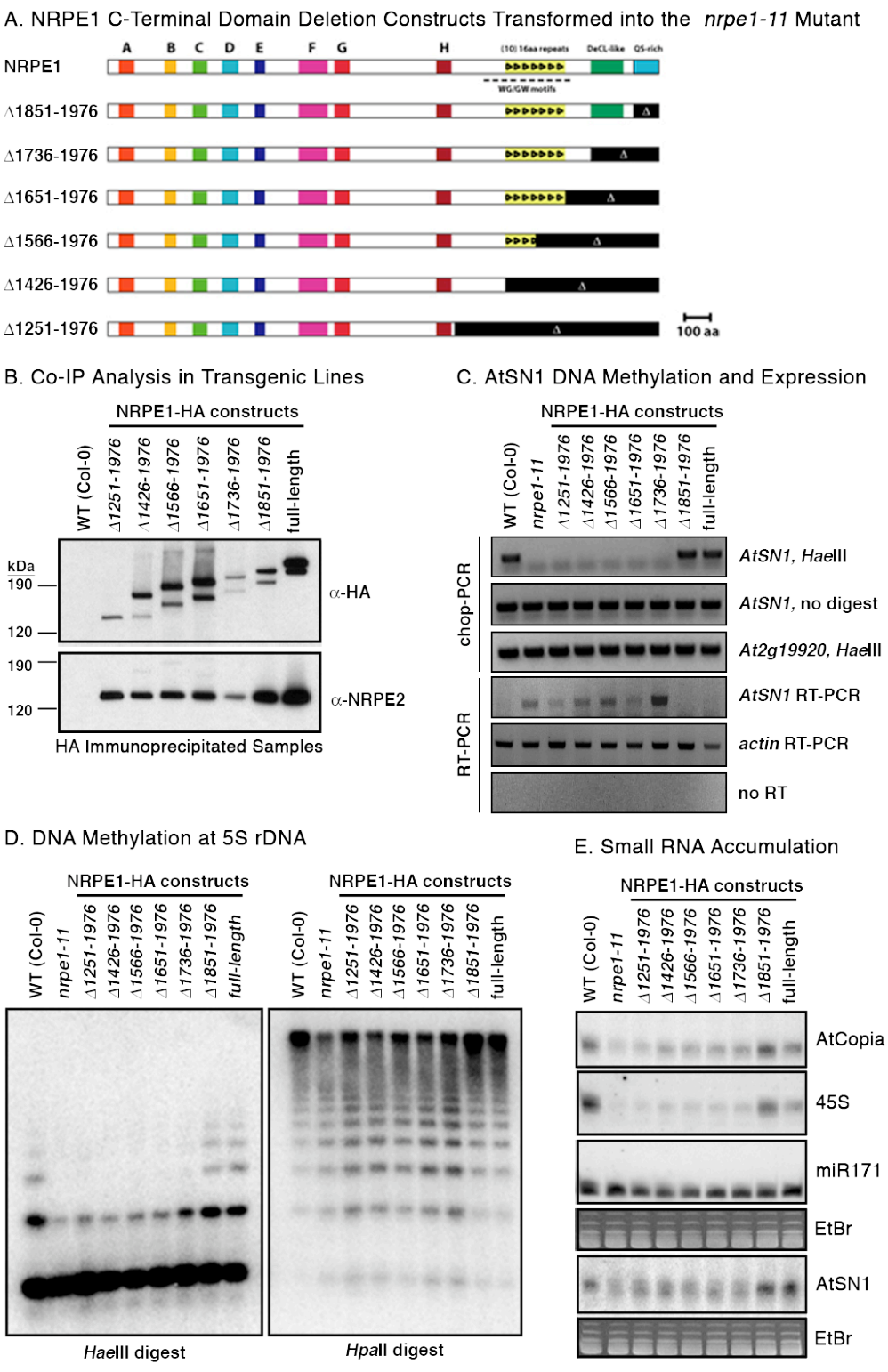

A. NRPE1 C-Terminal Domain Deletion Constructs Transformed into the nrpe1-11 Mutant
A B C D E F G H
(10) 16aa repeats
DeCL-like
QS-rich
WG/GW motifs
NRPE1
Δ1851-1976
Δ1736-1976
Δ1651-1976
Δ1566-1976
Δ1426-1976
Δ1251-1976
100 aa
B. Co-IP Analysis in Transgenic Lines
NRPE1-HA constructs
WT (Col-0)
Δ1251-1976
Δ1426-1976
Δ1566-1976
Δ1651-1976
Δ1736-1976
Δ1851-1976
full-length
kDa
190
120
α-HA
α-NRPE2
HA Immunoprecipitated Samples
C. AtSN1 DNA Methylation and Expression
nrpe1-11
chop-PCR
RT-PCR
AtSN1, HaeIII
AtSN1, no digest
At2g19920, HaeIII
AtSN1 RT-PCR
actin RT-PCR
no RT
D. DNA Methylation at 5S rDNA
HaeIII digest
HpaII digest
E. Small RNA Accumulation
AtCopia
45S
miR171
EtBr
AtSN1
EtBr

Figure 1

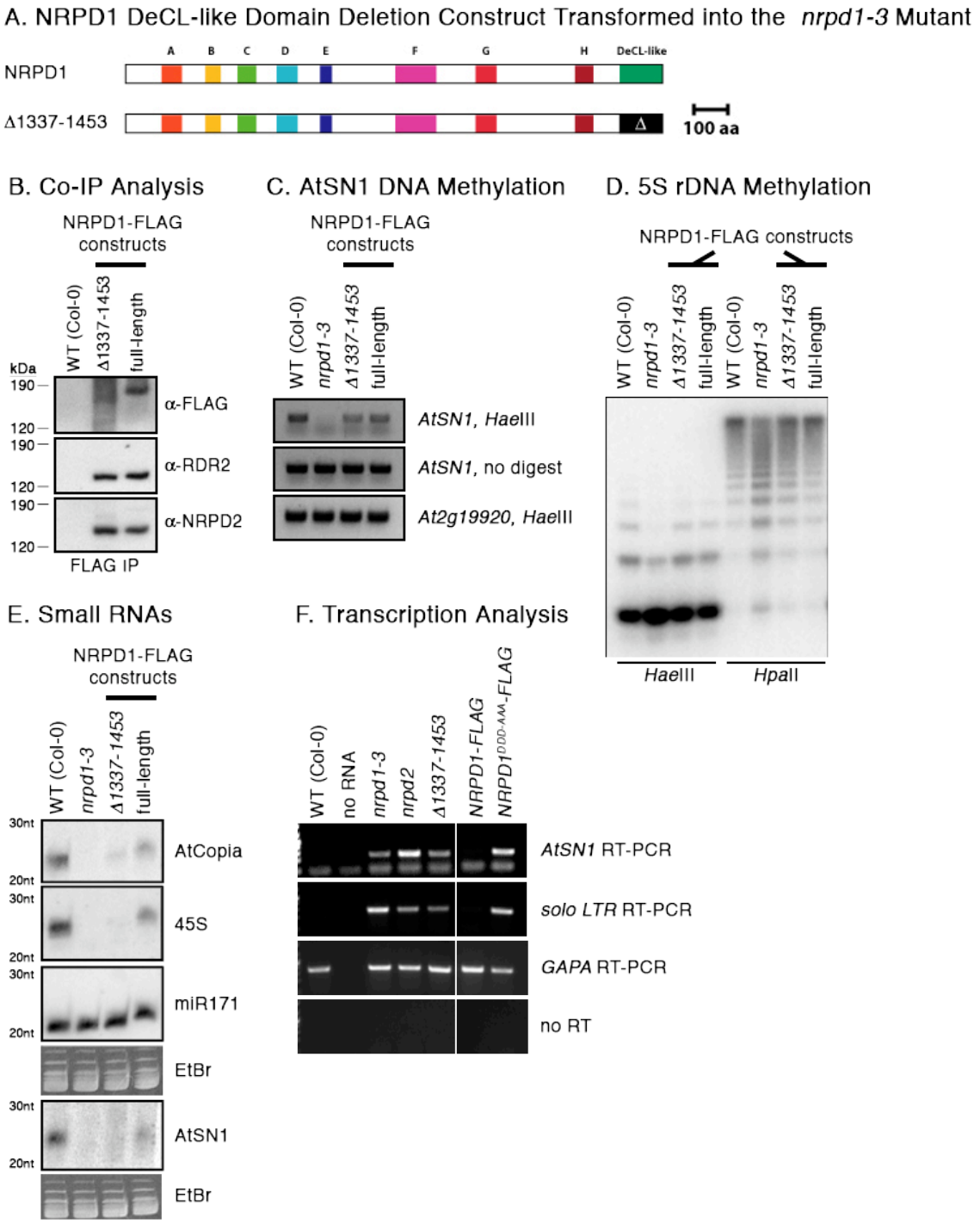

A. NRPD1 DeCL-like Domain Deletion Construct Transformed into the nrpd1-3 Mutant
NRPD1
Δ1337-1453
A
B
C
D
E
F
G
H
DeCL-like
Δ
100 aa
B. Co-IP Analysis
NRPD1-FLAG constructs
WT (Col-0)
Δ1337-1453
full-length
kDa
190
120
α-FLAG
α-RDR2
α-NRPD2
FLAG IP
C. AtSN1 DNA Methylation
NRPD1-FLAG constructs
WT (Col-0)
nrpd1-3
Δ1337-1453
full-length
AtSN1, HaeIII
AtSN1, no digest
At2g19920, HaeIII
D. 5S rDNA Methylation
NRPD1-FLAG constructs
HaeIII
HpaII
E. Small RNAs
NRPD1-FLAG constructs
30nt
20nt
AtCopia
45S
miR171
EtBr
AtSN1
F. Transcription Analysis
WT (Col-0)
no RNA
nrpd1-3
nrpd2
Δ1337-1453
NRPD1-FLAG
NRPD1DDD-AAA-FLAG
AtSN1 RT-PCR
solo LTR RT-PCR
GAPA RT-PCR
no RT

Figure 2

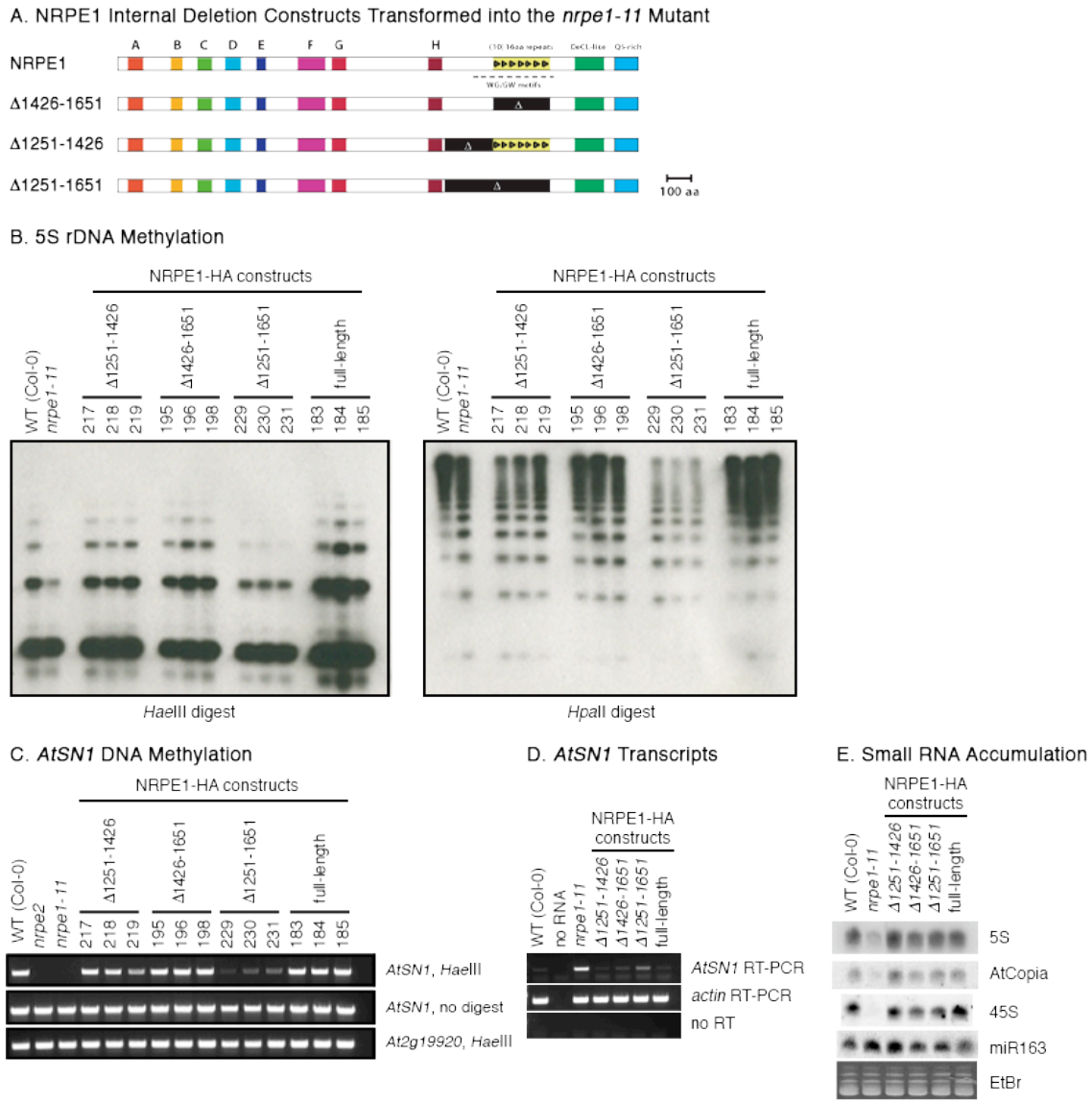

A. NRPE1 Internal Deletion Constructs Transformed into the nrpe1-11 Mutant
NRPE1
Δ1426-1651
Δ1251-1426
Δ1251-1651
100 aa
B. 5S rDNA Methylation
NRPE1-HA constructs
WT (Col-0)
nrpe1-11
Δ1251-1426
Δ1426-1651
Δ1251-1651
full-length
217 218 219 195 196 198 229 230 231 183 184 185
HaeIII digest
HpaII digest
C. AtSN1 DNA Methylation
WT (Col-0)
nrpe2
nrpe1-11
AtSN1, HaeIII
AtSN1, no digest
At2g19920, HaeIII
D. AtSN1 Transcripts
WT (Col-0)
no RNA
nrpe1-11
AtSN1 RT-PCR
actin RT-PCR
no RT
E. Small RNA Accumulation
5S
AtCopia
45S
miR163
EtBr

Figure 3

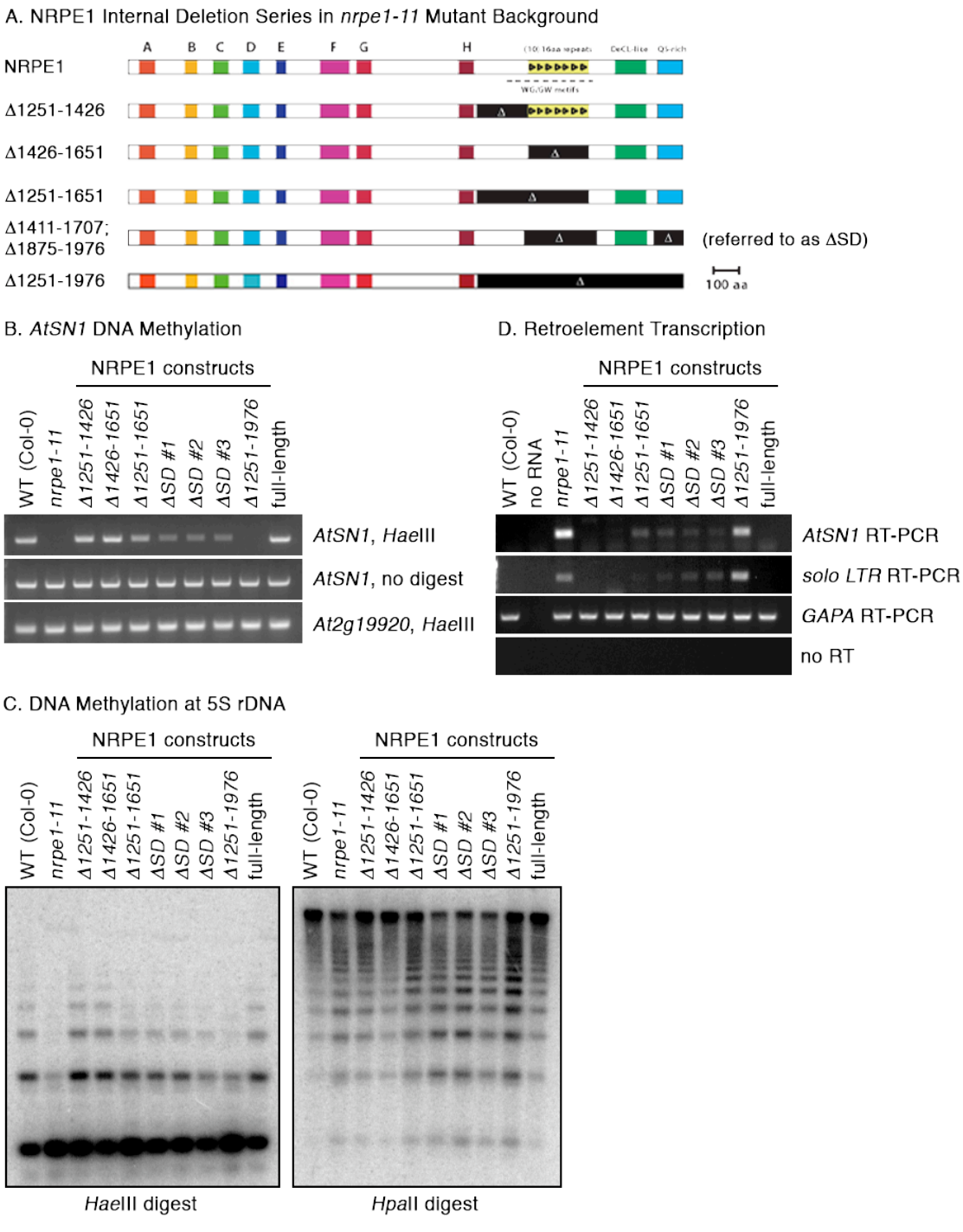

A. NRPE1 Internal Deletion Series in nrpe1-11 Mutant Background
A
B
C
D
E
F
G
H
NRPE1
Δ1251-1426
Δ1426-1651
Δ1251-1651
Δ1411-1707;
Δ1875-1976
Δ1251-1976
(referred to as ΔSD)
100 aa
B. AtSN1 DNA Methylation
NRPE1 constructs
WT (Col-0)
nrpe1-11
Δ1251-1426
Δ1426-1651
Δ1251-1651
ΔSD #1
ΔSD #2
ΔSD #3
Δ1251-1976
full-length
AtSN1, HaeIII
AtSN1, no digest
At2g19920, HaeIII
D. Retroelement Transcription
NRPE1 constructs
WT (Col-0)
no RNA
nrpe1-11
Δ1251-1426
Δ1426-1651
Δ1251-1651
ΔSD #1
ΔSD #2
ΔSD #3
Δ1251-1976
full-length
AtSN1 RT-PCR
solo LTR RT-PCR
GAPA RT-PCR
no RT
C. DNA Methylation at 5S rDNA
NRPE1 constructs
NRPE1 constructs
WT (Col-0)
nrpe1-11
Δ1251-1426
Δ1426-1651
Δ1251-1651
ΔSD #1
ΔSD #2
ΔSD #3
Δ1251-1976
full-length
HaeIII digest
HpaII digest

Figure 4

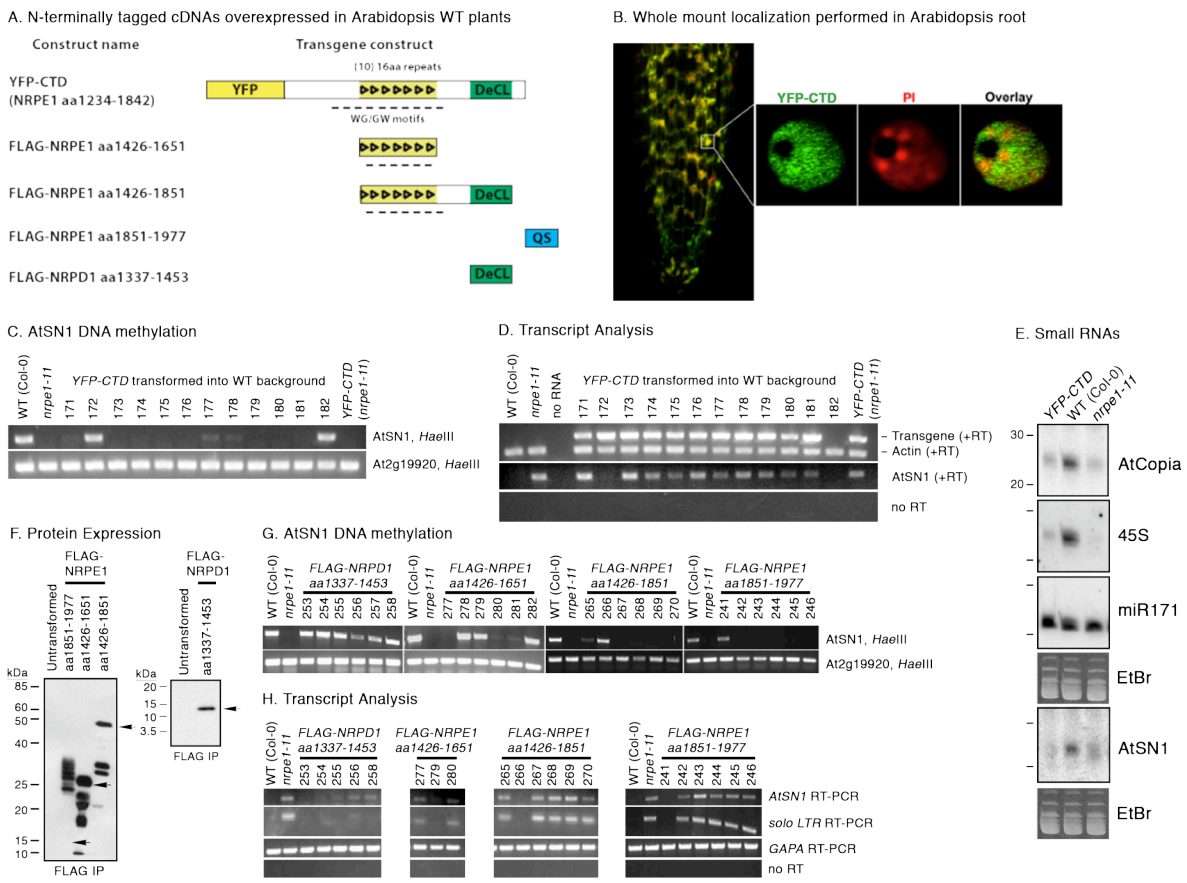

A. N-terminally tagged cDNAs overexpressed in Arabidopsis WT plants
Construct name
Transgene construct
(10) 16aa repeats
YFP-CTD
(NRPE1 aa1234-1842)
YFP
DeCL
WG/GW motifs
FLAG-NRPE1 aa1426-1651
FLAG-NRPE1 aa1426-1851
DeCL
FLAG-NRPE1 aa1851-1977
QS
FLAG-NRPD1 aa1337-1453
DeCL
B. Whole mount localization performed in Arabidopsis root
YFP-CTD
PI
Overlay
C. AtSN1 DNA methylation
WT (Col-0)
nrpe1-11
YFP-CTD transformed into WT background
YFP-CTD (nrpe1-11)
AtSN1, HaeIII
At2g19920, HaeIII
D. Transcript Analysis
WT (Col-0)
nrpe1-11
no RNA
YFP-CTD transformed into WT background
YFP-CTD (nrpe1-11)
Transgene (+RT)
Actin (+RT)
AtSN1 (+RT)
no RT
E. Small RNAs
YFP-CTD
WT (Col-0)
nrpe1-11
AtCopia
45S
miR171
EtBr
AtSN1
EtBr
F. Protein Expression
FLAG-NRPE1
FLAG-NRPD1
Untransformed
aa1851-1977
aa1426-1651
aa1426-1851
aa1337-1453
kDa
FLAG IP
G. AtSN1 DNA methylation
FLAG-NRPD1 aa1337-1453
FLAG-NRPE1 aa1426-1651
FLAG-NRPE1 aa1426-1851
FLAG-NRPE1 aa1851-1977
AtSN1, HaeIII
At2g19920, HaeIII
H. Transcript Analysis
AtSN1 RT-PCR
solo LTR RT-PCR
GAPA RT-PCR
no RT

Figure 5

Supplemental Data

Supplemental Methods

Sequence analysis. Full-length NRPD1 and NRPE1 protein sequences were obtained from NCBI GenBank and the publicly available genome sequencing efforts of JGI (http://www.jgi.doe.gov/). When necessary, cDNA predictions were made using FGENESH+ (http://www.softberry.com). Repeat elements were identified with XSTREAM (http://jimcooperlab.mcdb.ucsb.edu/xstream/) and by manual analysis.

***In vitro* co-immunoprecipitation.** NRPE1 cDNA constructs were recombined into pDEST17 (N-terminal GST fusion construct for bacterial expression) and expressed in the BL21.AI strain. A single colony of each construct was inoculated in 5 mL 1xLB (50 ug/mL Carb) and incubated overnight at 37 degrees C. Overnight culture was then used to inoculate fresh 1xLB (50 ug/mL Carb) and samples were incubated at 37 degrees C to an OD_{600} of 0.4. Expression was induced with the addition of L-Arabinose to 0.2% final concentration and incubated another 3 hours at 37 degrees C. Bacteria were pelleted and washed once with 1x Binding Buffer. The pellet was resuspended in 1x Binding Buffer and lysed by sonicating a total of 1 min at Duty Cycle 40% and Output 1.5 in a Branson Sonifier. Samples were centrifuged at 10,000 x g for 15 min at 4 degrees C. The soluble fraction was retained and GST-tagged recombinant protein purified with glutathione resin (Amersham).

MYC-AGO4 protein extract was isolated from 4.0 g of inflorescence tissue by grinding under liquid nitrogen in a mortar and pestle and resuspending in 14 mL Baumberger buffer. Extract was filtered through two layers of Miracloth and centrifuged

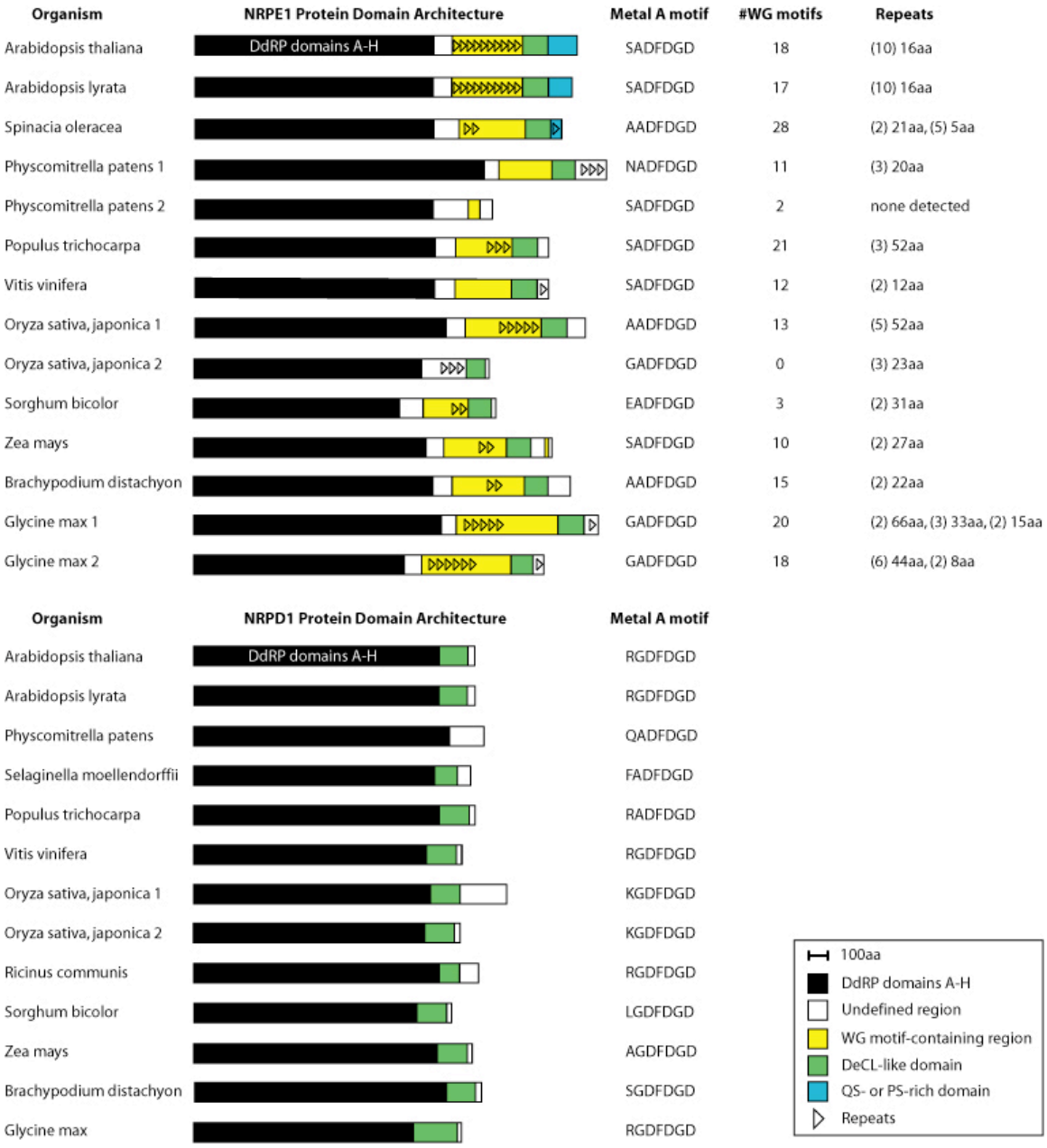
15 min at 11,500 rpm. Supernatant (300 uL) was added to the washed gluthathione resin with bound GST recombinant proteins and the volume was brought up to 1 mL with Baumberger buffer and incubated for 3 hrs at 4 degrees C. The glutathione resin was washed 5 times for 2 min each with 1 mL Baumberger Wash Buffer and pelleted by centrifugation at 200 rpm for 2 min. Protein was eluted from the resin by adding 50 uL 2x SDS loading buffer and incubating at 95 degrees C for 5 min.

Samples were split and run on 4-12% Novex gels. One sample set was Coomassie stained while the other was transferred to PVDF membrane for Western blot analysis.

***In vivo* co-immunoprecipitation using native antibodies.** All steps were performed at 4 degrees C unless otherwise stated. Frozen inflorescence tissue (0.7 g) was ground in liquid nitrogen and homogenized with 2 mL extraction buffer (50mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM MgCl2, 10% glycerol, 0.1% NP-40) containing 2 mM DTT, 1 mM PMSF, and 1/100 plant protease inhibitor cocktail (Sigma) [Li et al, 2006]. Sample was transferred to a 2.0 mL microcentrifuge tube and centrifuged twice at 13,000 rpm for 5 min. Samples were precleared with 20 uL Protein A agarose beads (Pierce) for 30 min. The samples were then incubated with 1:250 anti-NRPE1 or 1:250 anti-AGO4 for 3 hrs. Protein complexes were captured with 60 uL Protein A agarose beads (Pierce) for 2 hrs and then washed five times with extraction buffer. Samples were boiled in SDS loading buffer and run on a 7.5% Tris-glycine gel followed by transfer to PVDF membrane. Western blot was performed with 1:5000 anti-Myc monoclonal antibody (Upstate) O/N at 4C followed by anti-mouse-HRP and ECL Plus detection.

In vivo co-immunoprecipitation analysis comparing the extraction buffers from [Li et al, 2006] and [Baumberger et al, 2005] was performed as above, except one set of samples was incubated with anti-FLAG agarose beads and the other with anti-cMyc agarose beads (Sigma) for 4 hrs at 4 degrees C. The Protein A preclearing step was skipped.

Figure S1.



Comparison of NRPD1 and NRPE1 C-terminal domain architectures among diverse plant species. Domain features of illustrated full-length protein predictions are based on sequence analysis presented in Figures S2 and S3. The *Arabidopsis lyrata, Physcomitrella patens, Selaginella moellendorffii, Populus trichocarpa, Vitis vinifera, Sorghum bicolor, Brachypodium distachyon* and *Glycine max* NRPD1 and NRPE1 sequences were produced by the US Department of Energy Joint Genome Institute, http://www.jgi.doe.gov/ and are provided for use in this publication only. *Zea mays* NRPE1 was kindly provided by Lyudmila Sidorenko (Chandler lab). The *Brachypodium distachyon* sequences were identified by Tom Ream in the Pikaard lab. Remaining sequences have previously been published or are available from NCBI GenBank.

Figure S2. Predicted NRPE1 protein sequences among diverse plant species with key domain features denoted to the right-hand side. The Metal A motif is in black bold type; the conserved DdRP H domain is underlined in bold; WG/GW/WGW/GWG motifs are in bold; repeat elements are underlined with solid and dotted lines; the DeCL signature motif is in bold blue type.

>Arabidopsis_thaliana_NRPE1 (At2g40030)

MEEESTSEILDGEIVGITFALASHHEICIQSISESAINHPSQLTNAFLGLPLEFGKCESCGATEPDKCEGH
FGYIQLPVPIYHPAHVNELKQMLSLLCLKCLKIKKAKGTSGGLADRLLGVCCEEASQISIKDRASDGASYL
ELKLPSRSRLQPGCWNFLERYGYRYGSDYTRPLLAREVKEILRRIPEESRKKLTAKGHIPQEGYILEYLPV
PPNCLSVPEASDGFSTMSVDPSRIELKDVLKKVIAIKSSRSGETNFESHKAEASEMFRVVDTYLQVRGTAK
AARNIDMRYGVSKISDSSSSKAWTEKMRTLFIRKGSGFSSRSVITGDAYRHVNEVGIPIEIAQRITFEERV
SVHNRGYLQKLVDDKLCLSYTQGSTTYSLRDGSKGHTELKPGQVVHRRVMDGDVVFINRPPTTHKHSLQAL
RVYVHEDNTVKINPLMCSPL**SADFDGD**CVHLFYPQSLSAKAEVMELFSVEKQLLSSHTGQLILQMGSDSLL
SLRVMLERVFLDKATAQQLAMYGSLSLPPPALRKSSKSGPAWTVFQILQLAFPERLSCKGDRFLVDGSDLL
KFDFGVDAMGSIINEIVTSIFLEKGPKETLGFFDSLQPLLMESLFAEGFSLSLEDLSMSRADMDVIHNLII
REISPMVSRLRLSYRDELQLENSIHKVKEVAANFMLKSYSIRNLIDIKSNSAITKLVQQTGFLGLQLSDKK
KFYTKTLVEDMAIFCKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKN
LMAVLRDIVITNDGTVRNTCSNSVIQFKYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKAVLDSSPNSN
SSWELMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENAACTVRNKLNKVSLKDTAVEFLVEYRKQP
TISEIFGIDSCLHGHIHLNKTLLQDWNISMQDIHQKCEDVINSLGQKKKKKATDDFKRTSLSVSECCSFRD
PCGSKGSDMPCLTFSYNATDPDLERTLDVLCNTVYPVLLEIVIKGDSRICSANIIWNSSDMTTWIRNRHAS
RRGEWVLDVTVEKSAVKQSGDAWRVVIDSCLSVLHLIDTKRSIPYSVKQVQELLGLSCAFEQAVQRLSASV
RMVSKGVLKEHIILLANNMTCSGTMLGFNSGGYKALTRSLNIKAPFTEATLIAPRKCFEKAAEKCHTDSLS
TVVGSCS**WG**KRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVISTTNADAFVSSPGFDVTEEEMAEW
AESPERDSALGEPKFEDSADFQNLHDEGKPSGANWEKSSSWDNGCSGGSE**WG**VSKSTGGEANPESNWEKTT
NVEKEDAWSSWNTRKDAQESSKSDSGGA**WG**IKTKDADADTTPNWETSPAPKDSIVPENNEPTSDV**WG**HKSV
SDKSWDKKN**WG**TESAPAA**WG**STDAAV**WG**SSDKKNSETESDAAA**WG**SRDKNNSDVGSGAGVLGPWNKKSSET
ESNGAT**WG**SSDKTKSGAAAWNSWDKKNIETDSEPAA**WG**SQGKKNSETESGPAA**WG**AWDKKKSETEPGPA**GW**
GMGDKKNSETELGPAAMGNWDKKKSDTKSGPAA**WG**STDAAA**WG**SSDKNNSETESDAAA**WG**SRNKKTSEIES
GAGA**WG**S**WG**QPSPTAEDKDTNEDDRNPWVSLKETKSREKDDKERSQ**WG**NPAKKFPSSG**GW**SNGGGADWKGN
RNHTPRPPRSEDNLAPMFTATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPISDDDKTFVLE
KILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQ**DFSYRK**SLNNYLMKKYPDRAEEFIDKY
FTKPRPSGNRDRNNQDATPPGEEQSQPPNQSIGNGGDDFQTQTQSQSPSQTRAQSPSQAQAQSPSQTQSQS
QSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSPSQTQTQSPSQTQAQAQSPSSQSPSQTQT

Metal A

H

(10) 16 aa repeats & (18) WG motifs

DeCL

QS-rich

>Physcomitrella_patens_NRPE1-1
MQVMEAAAWRQPSQAPTADLVGLQIGLATTSEILGHSVIESRSKDTLISLVDPRLGLPAEDERCATCGGTN
YDECTGHFAHVKLTQPIFHPNYIRCVQRVLQKICLACGVPKVKKMKSFSEEAANLKQNFRDIDSEDVGGNG
EHPVLLEADAIEKDADDVVILLSSDEEEYPRDILRVVPSGPMDFLIRSTNESAIADLPQLKSYKSKSKAHA
NGFSHVDVTRKSTRKSSSKKSSSTQNPVKIYKGTPAGLDVLNADTLRTAEPLDTNTCPYCSPGYPDYRHIL
VKILPVKGRKKNDVSQIILLEVQGSDKGEKFLLPHDFWSFIKGAAYPENEEVPKSHVLSPLEALSILKKIS
DTAIGKLGMNGLVARPEGLIMKCVPIPPNCTRTTDYKYVSNTTAVRFGTDRVTRTLQNLVNEIGRIQRTRT
GKIMKRGQRDEVKVLQVLTAEYLREKGAPKAVPGKEPLKKDRNGRFTKQDDHRWTKDWISQNYLGKGGNYT
ARAVVAGDPSLAIETIGVPLEIAQKLTVPERATKWNRSKLQEYVDRTQMLQQGSGKPGATRIVRNEEAFQV
WANSTHTVQIGDVIHRNIQDGDFVYVNRPPSVHKHSLMALKVQVHYGLVLTINPLVCPPF**NADFDGD**IFHV
FIPQSLQAIAELEHLMAVPQQIISDHGGQPLLGLTQDTLLAAYLLTSSKLLVDKAGMDQLCLWALKQPPDA
AIVKSPKGGPFWTGEQIFGLTLPTDLQVGAPHEEVFIEGGEVIRWSNGAKSLRKDSEGIAAALCVQLGPVA
LVNYLNTATGLLHAWLQMHGFSTGLADFQVTSNSADRQKMLKSIFEDYYQKSIQESCDSVRILDAKVQAMG
QEVISSPDHLTRNINFLEQAAQQTFRNRESEVESIVMKYAARDNGLLMMVRSGSKGSRGKLLQQIAGMGLQ
LYKGQHLLPFSGSRRSSMSNSSELDWWEDKGLVRSSLVDGLNPSELFNHVIADRTVILRKHVEVVQPGTLF
KSLMLFLRDLHVMYDGSVRNQCGKNIVQFCYGGAIGVLKRSIPKERLSRSQFEVVNPATPIVTWEEDDLKR
WPLSILAGEPVGVLAATAISQPAYELMLDAPCLNGPFKPRPLELVQETLYPRAKSVLKPIDRTAIIRLVNC
PCTQPLCLERRVLAVQAHLKKISLKAIAESCAVEFWNMENFEVAGPSGEALRMGSPWLGHIKLSLNLMKQL
QVDVELMVERLRQRFSGIIKNPKKHPMGQIFFCVSYNCGISNGLCLHFSPKLPNKMQNQRNDEIYNTALLA
LLLKIRGTIISGLLDCTVKGDERIESVIIVSEDPSRTTWHRGLTCNQELEEELVLEVVVSPTKSKSKRGDA
WASVKQACLPLMHMVDWNRSMPYSIQEIRHALGVEASYQMISQRLGLVLDKTAPHTRSVHVKLVADMMTFS
GDANGFNFSGFQDMNKSTGISAPFTEASFQKPIKTLMDAAGRGATDSVESVLASCV**WG**KEAPLGTGSNFEL
FWQPSKDQSRLAASRKAEKDVHMIWKDLHEKCISDKVLPPSPPPSLPGLPTLPDGDVDLDDGAGFSPLHAS
NDAADDT**WG**SPHRNNGGDGVA**WG**DSPVVRDDDG**GWG**AVGKGNDSNEVDGYDQDNSTGASKELS**GW**SKPASE
RS**GWG**SMSDKEGSSRNAWDDFGKEDRHE**GWG**DGATEPINEG**GWG**SLNNEEGTTSGAKCSSD**WG**TNAVQEIG
DG**GW**DAVSIEVPEGD**GW**DSLKVPQTENAEVGSSEHADRSYGPGADGVSQEGQFRARGEESRRGGRPWTSRD
RRRWRGRGSFGKDRGSSGRMSPGNRQNSGTISRQEQTPWVQGSTKADAWAKHAWASFGSSQGEVQAGGD**GW**
DAVLPDNCGASNRAHSTYPIAGSMPPTSRQDEVEPECKDIDDLVKSMRRILFNPRNELGGRLSDEDDELVQ
TVLAYHPKLSEKAGCGTAYIKVDRSAGFVNNRCFWLVRTDGSEI**DFSFHK**CLKEKVAREFPSFLDRYDDVY
QAHKRPFPTANFEENKSAAQGNIDAGPSAAHLLEDMPIDHEDLDARPAAAHLPEGIPIDQEDLDAQPAVAH
LSEDTPIDQENLDAQPAANSISVDTHFDQQEDIDTQTGQESAPSIGVSSATKLICKKLTEPVHEHQDTSGP
H

>Physcomitrella_patens_NRPE1-2 (phya_79970)
MQIKSEDWTWTPGNVPIPPPPSAEIVGLQFGLTTANEINRARDTLSSLIDPRLGLPAENERCATCSGTNIN
ECTGHFGHLKLTQPIFHPHHVRLLQQVLSKICLACGSLKGKKKALAILKKIPEGAIGKLGMNRLVARPEGL
IMKCVLIPPNCTRTTDYKHVNNTTAVRFGTDNVTRTLQKLVAEIVHIRKTRAGKATNRTQRDESTKLQILT
AEYLREKGAPKAVPGKEPLKRDRNGRVTKQDYHRWTKEWLSQNVLGKSGNFTAKAVLAGDPFLGIEQIGIP
WLIAQKLTLPERASQWNHTKLQEYVNVSQKLQQESENTAHATRVERNEVVYQVLSKTSLKVQIGDIVHRHI
QDGDYVYVNRPPSVHRHSLVALKVHIHHQPTITVNPLICPPF**SADFDGD**IFHIFAPQSLQAIAELDQLMAV
KQQVISEHGGQPLLELTQSQSLIAFNVLNQNDTLLAAHLLTSKKLFLDKATMDQLCLWASKKPPEAAILKS
PKGGPFWTGEQVFALTLPEDFELGAPQEEVFIQGGEIIRWRNGTKLLRKGNDSVAAALCVQLGPVALVDYL
NTATGVLHTWLQVQGFSTGLTDFQVTPNRTKRQEMLKSILEESFLKSIQESCDFVRILDAKVQALDSDENP
SPESLTKNIRFLEQVAREIFQKRRSEAGRIVAKYAEQRNSLLMMVESGSKGSMEKLLQQIAGMGLQLYKGQ
HLLSYSSSRRPAMTYSSQLDWWEDMGLVRSSLVDGLKANELFRHVIADRTGILRKHVEVVQPGTLFKALMF
FLRDLHIMYDGSVRSQCSKNLIQFCYGGARGSLIPRKPTEETLAWEEDDHRRWPLSVLAGEPVGVLAAAAI
SQPAYELMLDAPSLNGPFKPRPLNLIQRLSTTWRFAHETLYPREKSSLKPTDRCVVLRLVHCECTESLCLE
RRVLEVQAHLKRINLRMMAESVAVEYWNMEDSRAAGPSGDLVRLGSPWLGHINLSQDAMKQCEVNVEDIVK
RLCQKFSQTAGYVLKKNKMGQIFFCHRIQETIIPGLLDCTMKGDERIETVRVVCEGPASTTWHRRFAHCTG
NLDEELVLEVYVSPSSSKSRGMAWASVKQACVSLKDLVDWNRSMPYSIQEIRCSLGIEVAYQIVVQTAPHT
HFVHVKLVAEMMTFSGDAIGFTFSGFKDMNRSISVSAPFSEASFQASAQPIRTLLGAAGRGATDSVEGVMT
NCI**WG**KEAPLGTGGNFGLFWQKPKAIKSFLCCVVKQRFTNICLLIGSHLQKFIVFYALMVLVLFDLKQVPL
IFQGIQRFGASKEAVKDVHTILKDLEDECIPDRFISSMPTLLPPHLHILPEGNLEFDDGAGFSPQRVSDCN
EGLDDRNHGNSSVDDQRGVSDTAVDGNVPIDWIKEEIYQNSDIKPDEELGAWQPTSYQGG**GW**DDIDTVPGL
RSLDNVSSDATGFKCYDTSKNSKNEEVVMVETTGMFGSINWGTNCIQDIGSDG**GW**DVPSSEVATGGSWDFL
DKKCQNDSSGCCGSKHLDHKHGSSGKSILLQERQFTAHEALDQDPAK

Metal A

H

(11) WG motifs

DeCL

(3) 20aa repeats

Metal A

H

(2) WG motifs

>Spinacia_oleracea_NRPE1
RYVPVPPNCLSVPDISDGVSVMSSDLCSAMLKKVLRQIEVIRSSRSGEPNFESHEVEANDLQVAVSQYLQV
RGTGKAARAADNRYGVSKEGNNSSKAWVEKMRTLFISKGSGFSSRSVITGDAYRAVNEVGVPCEIAQKMTF
EERVNVHNIQYLQGLVDKNLCLTFRDGLSTYSLREGSKGHTFLRLGQMVHRRIMDGDIVFINRPPTTHKHS
LQALRVYIHDDHVVKINPLMCGPL**AADFDGD**CVHLFYPQSLSARAEVLELFSVEKQLLSSHSGNLNLQLST
DSLLSLKTMFEVYFLDRASANQLAMYASSLLPSPALWKACSSNAKKKKAHSSGPRWTAQQVLQTALPSHFE
CHGDRLLIHDSEILKLDFNRDIVASVISDVLTSLFFNKSPKDALDFFDSLQPLLMENLFSEGFSVSLHDFF
FPKSELQNIQRNIQDLSPLLLQLRSSFNELVQVQFENHIREFKSPVGNFILISSALGSMIDSRSDSAIDKI
VQQIGFLGLQLSDRRKFYSRGLVEDVASLFHQKYPFADVYPSEEFGFVSRCFFHGLDPYEEIVHSIATREV
IVRSSKGLAEPGTLFKNLMAVLRDVVICYDGTVRNISSNSVIQFEYGVGGMQSQNLFPAGDPVGVLAATAM
SNPAYKAVLDSSPNSNSSWDMMKEILFCRANFRNDINDRRVILYLNDCCCGRKYCQENASCLVKNHLKKVS
LRDAAIELAIEYKRPKLEPESCEIDAGLVGHIHLNSGLLKASGIGMHDILQKCEEQVNLLRKKKKYGYHFK
RILLSVSDCCFFNHSDSKWTDMPCLKFFWQDMTDTDLERTKHIMADMICPVLLDTIIKGDPRISTVNIIWI
NPGTTTWVQSPCSSTKGELAVEVALEKEAVRLTGDAWRIVLDCCLPVFHLIDTRRSIPYAIKQIQDLFGIS
CAFDQAVQRLSTSVTMVTKGVLKEHLLLLASSMTCAGNLVGFNTSGIKALCRALNVQVPFTEATLYTPRKC
FERASEKCHVDTLASIVGSCS**WG**KRVSIGTGAKFDLLWETKEIEMADKPTDVYNFLHLVSSANEEEVDSGG
LGEDIESFEKDVYMEPALSPEQENKAVFEETLEIGVDSDITGADESSWDAFPSSGT**GW**NANKIDTGSGSAE
G**GW**SS**WG**SKKDQANPEDSSKTG**GW**SSGGSKQKPQPEDSSKSG**GW**DASKS**WG**GSNQGDPSPV**WG**QPVKATND
ISIENDHGSGSAEGG**GW**ANSGMKKDLSKQENSSTAG**GW**DASKSWSGSKPKDPSSA**WG**AGKKTDDNN**GW**KKS
DSKKDLASGSVEDGGCS**GW**GPKKDLLQPEDSAGEN**GWG**ASKSKSKEPSSA**WG**KPAQETDNI**GW**KKNNPQRD
SENLEGTS**GW**NDKLQKENKSFSKQSQPASSKDWDSTGNITAGSTGFGVEKGNEKPWDVASNVSVKKST**WG**Q
TGGNSWKKNEQDEKDGDPQGLP**WG**KSHKSSDSWTSGQGNQHPVSQGVSEKQGTLSS**WG**QPRDSSQKNNNEN
GVSSNFNRQGAGKSWDSKKKESNVQSSWAQQGDSTWKDSKEARSSVKANNSTNSG**GW**STGKALVDGVSSS**W**
GSQKEDRPQPKSNDRSVGDGNFDKDAKEEGLSSWDAKKVERKTQSS**WG**QPSESKNSAQSSADH**WG**SDKSNQ
PGKSS**GWG**SEDTNAGKDSEKQDS**WG**KSNVSTWKKESGEKLHGSDDSQSP**WG**QPGGS**GW**NKKQPEGGR**GWG**S
SNTGEWKSRKNQNQNQNQNQNRPPRGPNDDSPRVALTATRKRMDEFPTEEKDVLSEVESLMQSIRRIMHQS
GCVDGEPLLPDDQTYLIDNILNYHPDKAAKIGAGVDFITVKKHSNFQESRCFYVVSTDGKDT**DFSYIK**CIE
TFVKGKYPSVAESFTSKYFRRSQRPQPASPSPASPSPTSPSPASPSPAPPNPTPPT

Metal A

H

(28) WG motifs & (2) 21aa repeats

DeCL

PS-rich repeats

>Populus_trichocarpa_NRPE1
CTASISDCPISHSSQLTNPFLGLPLEFGKCESCGTSEPGKCEGHFGFIHLPIPIYHPSHISELKRMLSLIC
LKCLKLKRNKIQIKSNGVAERLLSCCEECAQISIREVKNTDGACFLELKLPSRSRLRDGCWNFLERYGFRY
GDDFTRPLLPCEVMQILKRIPAETRKKLSGKGYFPQDGYILQQLPVPPNCLSVPVVSDGITVMSSDLSISM
LKKVLKQAEVIRSSRSGAPNFDAHKDEATSLQSMVDQYLQVRGTTKTSRDVDTRYGVKKESSESTTKAWLE
KMRTLFIRKGSGFSSRSVITGDAYTLVNQVGIPYEIAQRITFEERVSVHNMRYLQELVDNKLCLTYKDGSS
TYSLREGSKGHTFLRPGQVVHRRIMDGDIVFINRPPTTHKHSLQALSVYVHDDHAVKINPLICGPL**SADFD**
GDCVHLFYPQSLAAKAEVLELFSVEKQLLSSHSGNLNLQLTTDSLLSLKMMFKACFLGKSAAQQLAMFISP
YLPQPALLKVNCFFPHWTAHQILQMALPACFNCSGERFLIINSNFLKVDFNRDVVASVINEILISMFFEKG
SGAVLKFFNSLQPMLMENLFSEGFSVSLEDFSISRAVKQRIPESFKAISPLLCNLRSTFNELVELQVENHI
RDVKQPVREFILTSSALGYLIDSKSDAAVTKVVQQIGFLGLQVSDRGKLYSKTLVEDLASHFLSKYPANLF
DYPSAQYGLIQNSFFHGLDAYEEMAHSISTREVIVRSSRGLSEPGTLFKNLMAILRDVVICYDGTVRNVSS
NSIIQFEYGVKVGTESQSLFPAGEPVGVLAATAMSNPAYKAVLDSTPSSNCSWDMMKEILLCKVGFKNDLA
DRRVILYLNDCGCGRNYCQERAAYLVKNHLEKVSLKDIAKCFMIEYKSQQIPESFGSDAGLVGHVHLDKRK
LQDLNITAQVILEKCQETVNTFRKKKKVGNLFKKTILLVSESCSFQQCIDESPCLMFFWQGADDVHLERTS
NILADMICPVLLETIIKGDHRISCANIIWATPETNTWIRNPSRTQKGELALDIVLEKSVVKKSGDAWRIVL
DSCLPVLHLINTTRSIPYAIKQVQELLGVSCAFDTAVQRLSKSVTMVAKGVLKEHLILLGNSMTCAGSLIG
FYTGGYKTLSRSLDIQVPFTEATLFTPRKCFEKAAEKCHTDSLSSIVASCA**WG**KHVTVGTGSHFDVLWDTK
EACLNPEGSMDVYSFLNMVRSTAGGEESVTACLGAEVDDLMLEDEDWNLSPEHNSSSDKPTFEDSAEFQDF
LGNQPAESNWEKISSLKDRSRSSGNWDVDKNDGAVKEKPWSLGMNTAEANDVASS**GW**DTAAARTTNNSWNS
ENNVAQSNSFS**GW**ATKKPEPHNGFATKVQEEPTTSNDWDAGAA**WG**RKDRDNKFAETNASKSW**WG**KVTDGDE
SGQNKSKNKRPEDQDVGTH**GW**DDKMSQDQSIS**GW**ASKTTQEATTESL**GW**DSKGNSNPGDAAC**GW**KAAST**WG**
AENTDGDKL**WG**KEVSSNQADTAS**GWG**KPKSPEISL**GWG**STKESVKSDR**GWG**VSSSGGGRDKKTENQSLAGQ
GKESG**GWG**NKVTSNQADTAS**GWG**KPKSSENSQ**GWG**LSKESGKEVHE**WG**VPNSAGGNGSETNNNNENQSLVE
QGKES**GW**DNKASSNQEGTAS**GWG**KPKSPALSE**GWG**SPREPVKAVH**GWG**VPNSGGGNDWKNKRNRPSKPHED
LNASGIFTTTRQRLDVFTSQEQDILSDIEPLMLSIRRIMHQTGYNDGDPLSADDQSYVLDNVFHYHPDKAV
KMGAGIDHVTVSRHSNFQESRCFYIVSTDGCKQ**DFSYRK**CLENFIKGKYPDLADEFIA

Metal A

H

(21) WG motifs (3) 52aa repeats

DeCL

>Vitis_vinifera_NRPE1
MEEDSSTILDGEISGIRFGLATRQEICIASVSDCPISHASQLTNPFLGLPLEFGKCESCGTAEPGQCEGHF
GYIELPIPIYHPGHVSELKRMLSLLCLKCLKIRKSKVTNNGITEQLLAPCCQDSPQVSVREFRPTEGACFL
ELKIPSRSRPKDGFWDFLARYGYRYGHNLSRILLPSEVMEILRRIPEDTRKKLVRKGYFPQDGYILQYLPV
PPNCLSVPDISDGVSIMSSDLSVSMLKKVLKQIEVIKGSRSGEPNFESHKIEANNLQSSIEQYLEVRGTAK
TSRSLDTRFGSSKEPNESSTKAWLEKMRTLFIRKGSGFSSRSVITGDAYKRVNEIGLPFEIAQRITFEERV
NVHNMKHLQNLVDEKLCLTYRDGLSTYSLREGSKGHTFLRPGQVVHRRIMDGDIVFINRPPTTHKHSLQAL
SVYVHDDHTVKINPLICGPL**SADFDGD**CVHLFYPQSLGAKAEVLELFSVEKQLLSSHSGNLNLQLATDSLL
SLKVLFERYFLNKAAAQQLVMFVSMSLPRPALLKSPCSGPCWTALQILQTALPSYFDCIGERHWISKSAIL
KVDYNRDVLQSLVNEIVTSIFSEKGPNEVLKFFDSLQPLLMENLFSEGFSVSLEDFSIPSEVTQNIQKNVE
DISSLLYNLRSMYNELLQLQAENHLRLTKVPVANFILNSSALGNLIDSKSDSAINKVVQQIGFLGQQLSEK
GKFYSRTLVEGMAYLFKSKYPFHGADYPSGEFGLIRSCFFHGLDPYEEMVHSISTREIIVRSSRGLSEPGT
LFKNLMAILRDVVICYDGTVRNVCSNSIIQFEYGVKARTKPQHFFPAGEPVGVLAATAMSNPAYKAVLDSS
PSSNSSWELMKEILLCQVNFKNDLIDRRVILYLNDCDCGRKYCRENAAYLVKNQLKKASLKDTAVEFMIEY
VKQHAVSGSSEPGTGLVGHIHLNKLLLQDLNVSMQEVCQKCEETINSFRKKKNVGPFFKKIILSFRECCTF
QHSCQSKGSDMPCLLFFWQGNRDDNLEQILHILAHKICPVLLQTIIKGDSRVCTVNIIWISPDTTTWIRNP
CKSRKGELALDIVLEKAAVKQRGDAWRIVLDACLPVLHLIDTRRSIPYAIKQVQELLGISCAFDQAVQRLS
KSVTMVAKGVLKEHLILLANSMTCAGNLIGFNSGGYKALSRALNLQVPFTEATLFTPRKCFEKASEKCHTD
SLSSIVASCS**WG**KHVTVGTGSRFDVLWDTKEIGPAQDGGIDIYSFLHLVRSGSYGKEPDTACLGAEVEDLI
LEDENLELGMSPEHSSNFEKPVFEDSAEFQNTWENHVPGSGGDWAVNQNKETTASTLKPSAWSS**WG**TDKVT
MKDTFSTREPDESSRSA**GW**DDKGT**WG**TDKAQNTAFRRTHEDSPRSSGRDETFRDGRPQFASSA**WG**KKIDEA
DKT**GW**NKNDGKPQMDKLRESYDWDCKVAQEKTTQSTYGGISSTTGDWKKNELQMEVVQHDESPVNEHSWDA
NLPEDPLAQATTSV**GW**DSSTGKDWTKRKLQSPSEQQRDPAIKSWSSSHNVMKEQSNQPASTH**GW**DSPGAK**G**
WNDVEEQSQWNQRGSAVKNDQSESSH**GWG**PSNEQNQLPSSQ**GWG**SPNAGAGHESETQSQ**WG**QPSGKKSRPE
GSR**GWG**SNNTEWKNKKNRPNKPQGPLNDDYSAGGIFTATRQRVDIFTSEEQDILLDVEPIMQSIRRIMHQA
GYNDGDPLSADDQSYILDKVFNNHPDKAVKMGTGIDYVMVSRHSSFLESRCFYVVSTDGHKE**DFSYRK**CLE
NFIKEKYPDNAETFIGKYFRRPRAGGNRERSVIPEDGGNREQSVVPEETGSENRQ

>Oryza_sativa_J_NRPE1-1 (OsJ_05410)
MEEDQSAIPVAEGAIKSIKLSLSTEDEIRTYSINDCPVTHPSQLGNPFLGLPLETGKCESCGASENGKCEG
HFGYIELPVPIYHPCHVTELRQILNVVCLKCLRVKKGKVKQTEGKDNTSALSCYYCRDLPALSLKEIKTAD
GAFRLELKMPPRKFMTEGSWNFLDKYGFHHGGTSHCRTLLPEEALNILKKIPEETKRKLAARGYIAQSGYV
MKYLPVPPNCLYIPEFTDGQSIMSYDISISLLKKVLQKIEQIKKSRAGSPNFESHEVESCDLQLSIAQYIH
LRGTTRGPQDNTKRFAISTDPSALSTKQWLEKMRTLFISKGSGFSSRSVLTGDPYIGVDVIGLPSEVAKRI
TFEEQVTDINLNRLQEIVDKGLCLTYRDGQATYAITVGSKGHTTLKVGQTISRRIVDGDVVFLNRPPSTHK
HSLQAFRVYVHEDHTVKINPLICAPF**AADFDGD**CVHIYYPQSLAAKAEALELFSVEKQLTSSHSGKVNLQL
VSDSLLALKHMSSRTMLSKEAANQLAMLVTCSLPDPAVIKSKPYWTISQIVQGALPKALTSQGDKHVVRDS
TIIKLDLDKESVQTSFSDLVYSTLSVKGPGEALQFLNVLQPLLMELILLDGFSVSLQDFNVPKVLLEEAQK
NIEKQSLILEQSRFAENQVVEMRVDNNLKDIKQQISDFVVKRSHLGLLIDPKSDSSVSKVVQQLGFVGLQL
YREGKFYSRRLVEDCYYTFVNKHPAVREEHSPEAYGLVRSSYFHGLNPYEELVHAISTREAIVRSSRGLTE
PGTLFKNLMALLRDVVICYDGTVRNVCSKSIIQLNYTEDDALDFPSAIGPGEPVGVLAATAISNPAYKAVL
DASQSNNTSWERMKEILQTTSRYKNDMKDRKVILFLNDCSCAKKFCKEKAAIAVQGCLRRITLEDCATDIC
IEDGNWAAPAGFQHPVPPPQCKILPVPIPIPAHGSVKFPPVPIPAPEHLKYNIHVVRYQKQIGLDGTSEAA
PALVGHIHLDRAHLERINISTEDILQKCQEVSGKYGKKKGHLSNLFKNITFSTCDCLFTQKLVDGKLPKLP
CLQFFVSDNMIVSESVERAVSVLADSLCGVLLNTIIKGDPRIQEAKIVWVGSDATSWVKNTQKASKGEPAV
EIIVEEEEALHIGDAWRTTMDACIPVLNLIDIRRSIPYGIQQVRELLGISCAFDQVVQRLSTTVRMVAKDV
LKDHLVLVANSMTFTGNLNGFNNAGYKATFRSLKVQVPFTESTLITPMKCFEKAAEKCHSDSLGCVVSSCS
WGKHAASGTGSSFQILWNESQLKSNKEYGDGLYDYLALVRTDEEKARYTFFDDVDYLAEENEADVCLSPEL
DGTIGQPIFDDNLEEQDVQNNSSWDNGTTTNASWEQNGSAGNDSDK**WGGW**NDAAAGADTGVTKPANQGNSC
WDVPATVEKSSSD**WGGWG**TEKAKEKEKISEEPAQHDAWSVQGPKRATDGGASWKKQSSTQNDGNSWKENKG
RGSNGGSWEKDNAQKGS**WG**RGNDEAENNNDVQNKSWETVAADAHASTEKS**WG**NVTASPSDNAWSAAPVSQG
NGSSDTKQSDSWD**GW**KSAGVDKAINKDKESLGNVPASPSFSAWNASPVSQGNERSDAKQSDSWD**GW**KSAGV
DKAINKDKESLGNVPASPSFSAWNAAPVSQGNERLDAKQSDSWD**GW**KSAGVDDSVKDKES**WG**NVPASPSDS
AWNAAPVSQGNESSDAKQSDSWD**GW**KSAGVDASTNKDKES**WG**NVPASPSDSAWNAAPVSQGDDVWNSAEAN
ESRNKDWKSD**GWG**ARGGNWRGQRNNPGRPPRKPDGRGLPRRPDERGPPRRHFDLTAEEEKILGEIEPTVLS
IRKIFRESIDSIKLSPEDEKFIKENVLEHHPEKQSKVSGEIDHIMVDKHQVFQDSRCLFVVSSDGTRS**DFS**
YLKCMENFVRKTYPEHGDSFCKKYFKRRRDQPPAADGGTAPGTPAGATQSTAVDTQEGTSQQTQPDIATAP

Metal A

H

(12) WG motifs

DeCL

(2) 12aa repeats

Metal A

H

(4.3) 52 aa repeats &
(13) WG motifs

DeCL

AATQQETLQDTPAPPADDGLLGKGPSPSD

>Oryza_sativa_J_NRPE1-2 (OsJ_04874)
MEGHPDPTSAATAMIPEASIRRINLSITSNEEILKAQPVNELEKPIPITHQSQLLNNPYLGLPLQVGSCQS
CGSNAIEECEGHFRFIELPMPIFHPSHVTELSQILNLICLRCLKIKNRKELPPLCVAEVKKSNGARGLELR
APIKKELEEGFWSFLDQFGSCTRGTSHCRPLLPEEVQNIIKKIPEETRRWLSVRGYIPQDGFILSYLCVPP
NCLRVSNVLDGNTFSCSGTSTNLLRKALRKIQQIRGSRIGSSNIQVDQVADDLQVDVANYINLGGTTKGHG
DDTFTSQPTAMQWKQKMKTLFISKSSSFSSRGVITGDPYIGLNVVGVPEEVAKRMSVEEKVTDHNIAQLQD
MMNKGLCLTYTDANSITYSLDAGKDNPNKKHTILKVGEIVNRRVFDGDIVFLNRPPSTDKHSVEAFYVQVH
NDHTIKINPLICDPL**GADFDGD**CVQIFYPRSLSARAEAKELYTVDKQLVSSHNGKLNFQFKNDFSLALKIM
CGREYSEREANQITNAMFSSGMYPQKPLIGGPYWTFPQILETTKSNAITLADHLDRESVGALATGTTISSI
LSTKGPREATEFLNLLQPLLMESLLIDCFSINLGDFTVPSPILEAIQNNPLELNKYREPIMDFITHSSAIG
LLVDPKSDSNMNKVVEQLGFLGPQLQHNGRLYSSRLVEDCLSKSLHRCCGSTNCCNPLEEYGTVRSSIYHG
LNPYEALLHSICEREKIMRASKGLVEPGSLFKNMMSRLRDVTACYDGSIRTSSGNLVLQFGSRDASNCVTP
GDPVGILAATAVANAAYKAVLAPNQNNIISWDSMKEVLLTRASTKADANHRKVILYLNQCSCENECMERAL
TIRACLRRIKLEDCTTEISIKYQQQATQAAHHLVGHIHLDKKQLNQIETIMDSVLHKCQETFRNNIKKKGS
MREILKTVTFISSTSLCDQHTDDDKKFQVSCLQFFLPGSITKNISESTERVIDFMTNAIFPIILDTVIKGD
PRVEEANLVRIEPESTFWVQSSGAEQKGEAALEITVEEAAAAESGNAWGVAMNACIPVMDLIDTTRSMPYD
IQQVRQYLSKSVGMITKSVLQEHLTTVASSMTCTGDLHGFNNSGYKATCQSLKV**QAPFMEATLSRSIQCFE**
KAAAKAYSDQLGNVVSACS**WG**NNAEIGTGSAFEILWNDENMSSSKSILGGYGLYDFLEAVETTGATKDKAI
VPHNYCLYDVDCIPEDKVCLEENNQITWTDKPKAEFLMESEGRRAGMHSTGQKHPRKPNWHEGNTKSSPNS
TAVEFTGQVFQRRQLKTKSNWNSDATQQDDKPSWYSSNSAGTQNFTIAGSSRPGEWNRKNNNRGQGGGREV
WKSEGPHRGGSSSNRNQGGGRAVWKSEASHRGSGNNRNRGGGRAVWKSEASRRGGSMRQVASCAFTPVEQQ
IFEQIEPITKNVKRIIRESRDGIKLPPDDEKFIVTNVLMYHPERKKKIAGNGNYITVDRHQVFHGSRCLYV
MSSDGSRK**DFSYKK**CLENYIRAQYPDAADSFCRKYFK

>Oryza_sativa_I_NRPE1-1 (OsI_05888)
MEEDQSAIPVAEGAIKSIKLSLSTEDEIRTYSINDCPVTHPSQLGNPFLGLPLETGKCESCGASENGKCEG
HFGYIELPVPIYHPCHVTELRQILNVVCLKCLRVKKGKVKQTEGKDNTSALSCYYCRDLPALSLKEIKTAD
GAFRLELKMPPRKFMTEGSWNFLDKYGFHHGGTSHCRTLLPEEALNILKKIPEETKRKLAARGYIAQSGYV
MKYLPVPPNCLYIPEFTDGQSIMSYDISISLLKKVLQKIEQIKKSRAGSPNFESHEVESCDLQLSIAQYIH
LRGTTRGPQDNTKRFAISTDPSALSTKQWLEKMRTLFISKGSGFSSRSVLTGDPYIGVDVIGLPSEVAKRI
TFEEQVTDINLNRLQEIVDKGLCLTYRDGQATYAITVGSKGHTTLKVGQTISRRIVDGDVVFLNRPPSTHK
HSLQAFRVYVHEDHTVKINPLICAPF**AADFDGD**CVHIYYPQSLAAKAEALELFSVEKQLTSSHSGKVNLQL
VSDSLLALKHMSSRTMLSKEAANQLAMLVTCSLPDPAVIKSKPYWTISQIVQGALPKALTSQGDKHVVRDS
TIIKLDLDKESVQTSFSDLVYSTLSVKGPGEALQFLNVLQPLLMELILLDGFSVSLQDFNVPKVLLEEAQK
NIEKQSLILEQSRFAENQVVEMRVDNNLKDIKQQISDFVVKRSHLGLLIDPKSDSSVSKVVQQLGFVGLQL
YREGKFYSRRLVEDCYYTFVNKHPAVREEHSPEAYGLVRSSYFHGLNPYEELVHAISTREAIVRSSRGLTE
PGTLFKNLMALLRDVVICYDGTVRNVCSKSIIQLNYTEDDALDFPSAIGPGEPVGVLAATAISNPAYKAVL
DASQSNNTSWERMKEILQTTSRYKNDMKDRKVILFLNDCSCAKKFCKEKAAIAVQGCLRRITLEDCATDIC
IEYQKQIGLDGTSEAAPALVGHIHLDRAHLERINISTEDILQKCQEVSGKYGKKKGHLSDPRIQEAKIVWV
GSDATSWVKNTQKASKGEPAVEIIVEEEEALHIGDAWRTTMDACIPVLNLIDIRRSIPYGIQQVRELLGIS
CAFDQVVQRLSTTVRMVAKDVLKDHLVLVANSMTFTGNLNGFNNAGYKATFRSLKV**QVPFTESTLITPMKC**
FEKAAEKCHSDSLGCVVSSCS**WG**KHAASGTGSSFQILWNESQLKSNKEYGDGLYDYLALVRTDEEKARYTF
FDDVDYLAEENEADVCLSPELDGTIGQPIFDDNLEEQDVQNNSSWDNGTTTNASWEQNGSAGNDSDK**WGGW**
NDAAAGADTGVTKPANQGNSCWDVPATVEKSSSD**WGGWG**TEKAKEKEKISEEPAQHDAWSVQGPKRATDGG
ASWKKQSSTQNDGNSWKENKGRGSNGGSWEKDNAQKGS**WG**RGNDEAENNNDVQNKSWETVAADAHAST**EKS**
WGNVTASPSDNAWSAAPVSQGNGSSDTKQSDSWD**GW**KSAGVDKAINKDKESLGNVPASPSFSAWNASPVSQ
GNERSDAKQSDSWD**GW**KSAGVDKAINKDKESLGNVPASPSFSAWNAAPVSQGNERLDAKQSDSWD**GW**KSAG
VDDSVKDKES**WG**NVPASPSDSAWNAAPVSQGNESSDAKQSDSWD**GW**KSAGVDASTNKDKES**WG**NVPASPSD
SAWNAAPVSQGDDVWNSAEANESRNKDWKSD**GWG**ARGGNWRGQRNNPAEEEKILGEIETTVLSIRKIFRES
IDSIKLSPEDEKFIKENVLEHHPEKQSKVSGEIDHIMVDKHQVFQDSRCLFVVSSDGTRS**DFSYLK**CMENF
VRKTYPEHGDSFCKKYFKRRRDQPPAADGGTAPGTPAGATQSTAVDTQEGTSQQTQPDIATAPAATQQETL
QDTPAPPADDGLLGKGPSPSD

Metal A

H

(3) 22aa repeats

DeCL

Metal A

H

(4.3) 52aa repeats & (13) WG motifs

DeCL

>Oryza_sativa_I_NRPE1-2 (OsI_05331)
MGFSPAISLRNLSMVALRIWESGTTVYIAAAAVPGAKLVLVLILTAGRPFFLTHLYMRYSRTEMEGHPDPT
SAATAMIPEASIRRINLSITSNEEILKAQPVNELEKPIPITHQSQLLNNPYLGLPLQVGDAIEECEGHFGF
IELPMPIFHPSHVTELSQILNLICLRCLKIKNRKVQNIIKKIPEETRRWLSVRGYIPQDGFILSYLCVPPN
CLRVSNVLDGNTFSCSGTSTNLLRKALRKIQQIRGSRIGSSNIQVDQVADDLQVDVANYINLGGTTKGHGD
DTFTSQPTAMQWKQKMKTLFISKSSSFSSRGVITGDPYIGLNVVGVPEEVAKRMSVEEKVTDHNIAQLQDM
MNKGLCLTYTDANSITYSLDAGKDNPNKKHTILKVGEIVNRRVFDGDIVFLNRPPSTDKHSVEAFYVQVHN
DHTIKINPLICDPL**GADFDDD**CVQIFYPRSLSARAEAKELYTVDKQLVSSHNGKLNFQFKNDFSLALKIMC
GREYSEREANQITNAMFSSGMYPQKPLIGGPYWTFPQILETTKSNAITLADHLDRESVGALATGTTISSIL
STKGPREATEFLNLLQPLLMESLLIDGFSINLGDFTVPSPILEAIQNNPLELNKYREPIMDFITHSSAIGL
LVDPKSDSNMNKVVEQLGFLGPQLQHNGRLYSSRLVEDCLSKSLHRCCGSTNCCNPLEEHGTVRSSIYHGL
NPYEALLHSICEREKIMRASKGLVEPGSLFKNMMSRLRDVTACYDGSIRTSSGNLVLQFGSRDASNCVTPG
DPVGILAATAVANAAYKAVLAPNQNNIISWDSMKEVLLTRASTKADANHRKVILYLNQCSCENECMERALT
IRACLRRIKLEDCTTEISINTSLCDQHTDDDQEFRVSCLQFFLPASITKNISESTERVIDFMTNAIFPIIL
DTVIKGDPRVEEANLVRIEPESTFWVQSSGAEQKGEVALEITVEKAAAAESGNAWGVAMDACIPVMDLIDT
TRSMPYDIQQVRQYLSKSVGMITKSVLQEHLTTVASSMTCTGDLHGFNNSGYKATCQSLKVQAPFMEATLS
RSIQCFEKAAAKAYSDQLGNVVSACS**WG**NNTEIGTGSAFEILWNDENMSSSKSILGGYGLYDFLEAVETTG
ATKDKAIVPHNYCLYDVDCIPEDKVCLEENNQITWTDKPKAEFLMESEGRRAGMHSTGQKHPRKPNWHEGN
TKSSPNSTAVEFTGQVFQRRQLKTKSNWNSDATQQDDKKPSWYSSNSAGTQNFTIAGSSRPGEWNRKNNNR
GQGGGRAVWKSEGPHRGGSSSNRNQGGGRAVWKSEASHRGSSNNRNRGGGRAVWKSEASRRGGSMRQVASC
AFTPVEQQIFEQIEPITKNVKRIIRESRDGIKLPPDGEKFIVTNVLMYHPERKKKIAGNGNYITVDRHQVF
HGSRCLYVMSSDGSRK**DFSYKK**CLENYIRAQYPDAADSFCRKYFK

>Zea_mays_NRPE1
MEEDHSVILISEGAIKSIKLSLSTGEEICTYSINECPVTHPSQLGNPFLGLPLEAGKCESCGASENDKCEG
HFGYIELPVPIYHPCHVTELRQLLSLICLKCLRIKKGKDIPALSLKEIKTTDGAIRLELRAPHNKHMTERS
WNFLDKYGFHHGGCSHHRTLLPEEALNILKKVPDDTRRKLAARGYIVQTGYVMKYLPVPPNCLYIPEFTDG
QSIMSYDISIALLKKVLQKIEQIKRSRSGSPNFESHDAESCDLQLAIGQYIRLRGTTRGPQDNTKRFTVGS
ADSAALSTKQWLEKMRTLFISKGSGFSSRSVLTGDPYIGLGVVGLPSEVAKRMTFEEQVTDININRLQDVV
DKGLCLTYRDGQATYAITVGSKGYTTLKVGQTISRRIVDGDVVFLNRPPSTHKHSLQAFYAYVHDDHTVKI
NPLMCGPF**SADFDGD**CVHIYYPQSLAAKAEALELFSVERQLISSHSGKVNLQLGNDSLVAMKAMSHTTMLH
KELANQLAMFVPFSLLAPAVIKPVPSWTISQIVQGAFPANLTCQGDTHLVRDSTIIRLDLGKESVQDSFPD
LVSSILREKGPKEALQFLNVLEPLLMEFLLLDGLSISLRDFNVPKALLEEAQKDIRNQSLILEQSRCSTSQ
FVEFRVENNLKNVKQQISDSVGKFSDLGLLIDPKKEASMSKVVQQVGFVGLQLYREGKLYSRRLVEDCFTN
FVNKHLAIGDEYPPEAYGLVQSSYFHGLNPYEELIHAISTREAMIRSSRGLSEPGTLFKNLMAILRDVVIC
YDGTVRNICSNSIIQLKYGEDDETDSSSVVPPGEPVGVLAATAISNPAYKAVLDSSQSNNASWESMKEILQ
TRTSYKNDVKDRKVVLFLNDCSCAKKFCKERAALAVQSCLKRVTLGDCATDICIEHQKQINLDGTSEAAPT
LVGHIHLDKGHLERINISTQDILQKCQEMPIDGKLHKVPCVQFAFSDDIVLSESIERAVNVIADSVCSVLL
DTIIKGDPRIQAAKVIWVESDAASWVKHTRKVSKGESALEIIVEKDDAVSNGDAWRTAIDACLPVLNLIDT
RRSIPYGIQQVRELIGISCAFDQVVQRLSTTVKMVNKGVLKDHLILVANSMTCTGNLIGFNIAGYKATFRS
LKVQVPFTESTLFTPMKCFEKAAEKCDSDSLGCVVSSSA**WG**KHAAVGTGSSFQILWNENQVCLSYQPELIA
YISLYQTDYMFLDDVDYLVEENAADDMCLSPEPDGTLGKPTFEDNFEEQNIQKGSSWEIGITTNSSWEQNA
SVANDSGD**WGGW**SSGGGAAAKPADQDNSWEVHAKVQDNSTTD**WGGW**SVEKPTGEATVSGEPAETDTWADKG
AKMESDAGDGNWEKSSTPEASKKNDSSENTWDKRKGDGGDGA**WG**NRSDDGHGNWEHPSNWNGQSLDVDQDT
WGNARGKKKADGNYCQWEEQPSNYKQKKTNADHDSSYNNVMPSSEIAWNAGDGTGRPNAKSNAESS**WG**EED
KMESDDHPKVPKESDTWNTGRSNESPWDNTDALQDSWVKSAARNNNTQDGSWDKVVSMKDLDSLQDSWSKA
TIQTNDAQNDSWDNVAKNAPDSAAEDS**WG**AATPAETTDSGNKEWKSD**GWG**AKSGNWSSQRNNPGRPPRRPD
ERGPPPPRQRFELTVAEKNILLEVEPIKLRVRSIFREACDGVRLNPEDEKFILEKVLEHHPEKQSKVSGEI
DYLTVNKHQTFQDTRCFFVVSTDGSQA**DFSYLK**CLENFVRKSYTEDADTFCMKYLRPPETEQGTPPAPQAE
VPQETWGSPAVPLEGGTHIAGPDSTGDAVILGEQHDLTPASPAVAPQVASEPDTTDGTGLLGKAPQAD**WG**P
RFDAD

Metal A

H

(3) 22a repeats

DeCL

Metal A

H

(2) 27aa repeats & (10) WG motifs

DeCL

>Solanum_lycopersicum_NRPE1 (DQ020653) - incomplete N- and C-termini

DFDGDCVHLFYPQSLSAKAEVLELFAVGKQLLSSHTGNFNSQLATDSLLSLKLMFSHYFFDKAAAQQLAMF — Metal A
LPMALPDSAVVDVRKSGAMWTTLQILGAALPDGFDSCGETHTIGKSQFLGIDYHRDLISSILNDVITSIYF
MKGPNDVLKFFNSLQPLLMENLCTEGFSISLRDFYMTKAVRDGIQERIQCMSKLLHHLRSSYNESVEVQLE
HHLRNEKLPVIDFVLKSSGMGVLIDSKSESAFNKVVQQIGFLGLQISDRGKFYXXTLVHDMAQLFQKKYPS
VGTNPSEEFGLVRSCLFYGLDPYQGMIHSISSREVIVRSTRGLTEPGTLFXNLMAILRDVVICYDGTVRNV
SSNSIIQFEYGSSGGSNLPSEFCAGDPVGVLAATAMSNPAYKAXLDSSPSSNSSWEMMKEILLCGVSFKND
VSDRRVILYLNDCGCRRGYCREKAAYVVKNHLSKVCLKDAADEFLIEYAGRQAGYENSETGTGLIGHIRLN
QGQLENLGISVLEVHERCQENISSFRXKKKIGNLFKRIVLSVSEFCSFCHNSGSKCLNAPCLRFSWPDASD
DHLERVSHILADMIXPILLDTVIKGDPRVSSANIAWISPDTMSWIRSPSKSQRGELALDIVLEKEAVKXRG
DAWRXLMDSCLPVIHLIDTTRSIPYAIKQVQELIGISCAFEQAVXRLSTSVTMVTKGVLKDHLVLLANSMT
CAGNLVGFNAGGIKALSRSLNVQIPFTEATLFTPRKCFERAAEKCHVDSLSSIVASCS**WG**KHVAVGTGSRF — H
EVLLNTRNVEWNIPDTRDVYSFLHLVRNTSAQEVEGTSCLGAEIDELEEDEDMGLYLSPNRDSGSEMPTFE
DRAEFDYNENLDEGKPSGSAWEEASSGSVKSGGSWDMAGKTQNGAEEGVNQSDSWSS**WG**KKVDEPENNRQQ — (3) WG
SGSGEQSGSWSPWGRRWKKMVVLGDEPKQLNSESS**WG**KAPNGGGLGSATAEGNRRLDQSVNDWSSSVSRDG — motifs
QYKKWWLEFFKRWWLELSG**GW**QWKNNRPARSADDSNRGGHFTATRQKIDLFTAEEQEIISDVDPIMLKVKS
DPLSADDQSYIIDTVLNYHPDKAVKMGAGLDYITVSKHTNFQDTRCFYVVSTDGAKQELAAV

>Glycine_max_NRPE1-1 (Glyma15g37710)

MEDNPPSSVLDGTVVGIKFGMATRQEICTASISDSSISHASQLSNPFLGLPLEFGRCESCGTSEVGKCEGH
FGYIELPIPIYHPSHISDLKRMLSMVCLNCLKLRKTKLPASSSGLAQRLISPCCQEDKAALVSIREVKTSD
GACYLALKVSKSKMQNGFWSFLEKYGYRYGGDHTRALLPCEAMEIIKRIPIETKKKLAGKGYFPQDGYVLK
YLPVPPNCLSVPEVSDGVSVMSSDPSITILRKLLRKVEIIKSSRSGEPNFESHHVEANDLQSVVDQYFQIR
GTSKPARDIETHFGVNKELTASSTKAWLEKMRTLFIRKGSGFSSRNVITGDCYKRINEVGIPVEVAQRITF
EERVNIHNIRYLQKLVDEHLCLTYKEGGSTYSLREGSKGHIYLKPGQIVHRRIMDGDIVFINRPPTTHKHS
LQALYVYIHEDHTVKINPLICGPL**GADFDGD**CVHLFYPQSLAAKAEVVELFSVENQLLSSHSGNLNLQLST — Metal A
DSLLSLKMLVKRCFFDRAAANQLAMFILLPLPRPALLKASSGDACWTSIQILQCALPLGFDCTGGRYLIRQ
SEILEFEFSRDVLPATVNEIAASVFFGKGPKEALNFFDVLQPFLMESLFAEGFSVSLEEFSISRAIKRIIR
KSIGKVSSLLYQLRSLYNELVAQQLEKHIRDVELPIINFALKSTKLGDLIDSKSKSAIDKVVQQIGFLGQQ
LFDRGRFYSKGLVDDVASHFHAKCCYDGDGYPSAEYGLLKGCFFNGLDPYEEMVHSISTREIMVRSSRGLS
EPGTLFKNLMAILRDVVICYDGTVRNICSNSIIQFEYGIQAGDKSEHLFPAGEPVGVLAATAMSNPAYKAV
LDASPSSNSSWELMKEILLCKVNFRNELVDRRVILYLNDCDCGGSYCRENAAYSVKDQLRKVSLKDAAVEF
IIEYQQQRTQKENSETDVGLVGHIYLDEMMLEELKISMAYVFDKCHERLKSFSQKKKKKMTLFLSYLIVRG
TVKCSIFVVSRIQDLYFIDHEYCTWKTMVFLSVSETIKNEIFPGLFMTISYLLFFTIPTESCSSSHPAAPC
LTFWLKNYDSDLDNAVKVLAEKICPVLFKTIIQGDPRISSASIIWVSPDTNTWVRNPYKSSNGELALDIIL
EKEAVKQSGDAWRVVLDACLPVLHLIDTRRSIPYAIKQIQELLGISCTFDQAIQRVAASVKMVAKGVLREH
LILLASSMTCGGNLVGFNIGGYKALSRQLNIQVPFTDATLF**TPKKCFERAAEKCHTDSLSSIVASCS**WGKH — H
VAVGTGSKFDVVWDANEIKSNEIEGMDVYSFLHMVKSFTNGEEETDACLGEDIDDLLEEEYMDLGMSPQHN
SGFEAVFEENPEVLNGSTSN**GW**DVSSNQGESKTNEWS**GW**ASSNKAEIKDGRSEIAPKNS**WG**KTVNQEDSSK
SNPWSTSTIADQTKTKSNEWSA**WG**SNKSEIPV**GW**ASSNKTEIKDGRSETAQENS**WG**KTVNQEDSSKSNAWN
TSTTVDHANTKSNEWSA**WG**SNQSEIPAGGSKAVQEDS**WG**SSKWKADVAQEDNSRLGAWDANAADQTKSSEW — (2) 66aa
S**GWG**KKKDVTQEDNSRLGAWDANAADQTKSRDWS**GWG**KKKDITQEDNSRLGAWDANAADQTKSSEWS**GWG**K — repeats, (3)
KKDQIRQNLMNGQVGERRKKLPKKTIPGLVLGMQIQQIRQNLMNEDQTKSNEWS**GWG**KKKDVTQEDNSRLG — 33aa repea
AWDANAADQTKSNEWSD**WG**KKKEVTQEDNVQDS**WG**SGKRKDKVTQEDNSGSG**GWG**ANRTDLAKSKSSEWSS — & (19) WC
WGKNKSEIPAGGSENVQNDS**WG**SGKLEDDTQKENSGSAWVRNKAETIDGGSEKPQEDAWNSGNWKAESKVG — motifs
NAS**WG**KPKSSESQAWDSHNQSNQNSSSQ**GW**ESHIASANSESEKGFQ**WG**KQGRDSFKKNRFEGSQGRGSNAG
DWKNRNRPPRAPGQRLDIYSSGEQDVLKDIEPIMQSIRRIMQQQGYNDGDPLAAEDQLFVLENVFEHHPDK
ETKMGTGIDYVMVNKHSSFQESRCFYVVCKDGESK**DFSYRK**CLANYISKKYPDLAESFLGKYFRKPRARGD — DeCL
QTATPGRDEAATPGEQTATPGRDEAATPAEQISTPTPMETNE* — (2) 15aa repeats

>Glycine_max_NRPE1-2 (Glyma13g26690)
MEIIKRIPIETKKKLAGKGFFPQDGYVLKYLPVPPNCLSVPEVSDGASVMSSDPSMTILRKLLRKVEIIKS
SRSGEPNFESHHVEANDLQSVVDQYFQIRGTSKPARDIETHFGVNKELTASSTKAWLEKMRTLFIRKGSGF
SSRNVITGDCYKRINEVGIPVEVAQRITFEERVNIHNIRYLQKLVDEHLCLTYKEGVSTYSLREGSKGHIY
LKPGQIVHRRIMDGDIVFINRPPTTHKHSLQALYVYIHEDHTVKINPLICGPL**GADFDGD**CVHLFYPQSLA
AKAEVVELFAVENQLLSSHSGNLNLQLSTDSLLALKMLVKRCFLGRAAANQLAMFLLLPLPRPALLKASSD
DACWTSIQILQGALPMGFDCTGGRYLIRQSEILEFDFSRDALPATINEIAASIFFGKGPMEALKFFDVLQP
FLMESLFAEGFSVSLEEFSISRAIKRIIRRSIGKASSLLYQLRSLYNELVAQQLEKHIQDVELPIINFALK
STKLGDLIDSKSKSTIDKVVQQVGFLGQQLFDRGRFYSKGLVDDVASHFHAKCCYDGDGYPSAEYGLLKGC
FFNGLDPYEEMVHSISTREIMVRSSRGLSEPGTLFKNLMAILRDVVICYDGTVRNICSNSIIQFEYGIQAG
DKTEHLFPAGEPVGVLAATAMSNPAYKAVLDASPNSNSSWELMKEILLCKVNFRNEPVDRRVILYLNDCDC
GGSCCRENAAYSVKNQLRKVSLKNAAVEFIIEYQQQRTQKENSETDAGLVGHIYLDEMMLEELKISMANVF
EKCLERLKSFSRKKKARQSFLIIRGTVNESCSSSHPAAPCLTFWLKNHDSDLDNAVKVLSENICPVLFETI
IKGDPRISSASIIWVSPDTNTWVRNPYKSSNGELALDIVLEEEAVKQSGDAWRIVLDSCLPVLHLIDTRRS
IPYAIKQIQELLGISCTFDQAIQRVAASVKMVAKGVLREHLILLASSMTCGGNLVGFNTGGYKALSRQLNI
QVPFTDATLFTPKKCFERAAEKCHTDSLSSIVASCS**WG**KHVAVGTGSKFDIVWDSSEVFDNTDLILDLIRI
GIKSNEIEGMDVYSFLHMVKSVTNGEEETDACLGEDIDDLLEEEYMDLGMSPQHNSGFEAVFEENPEVLNG
STSN**GW**DVSSNQTQSKTNEWS**GW**ASSNKDGRSETAQENS**WG**KTVNQEDSSKSNAWNTSTTADQTKTKSNEW
SD**WG**SNKSEIPAGGSKAVQEDSSKSNAWNTSTTSNQTKTKSKEWSA**WG**SNKSEIPACGSKAVQEDSSKSNT
WNTSTTADQTKTKSNEWSA**WG**SNKSEIPAGGSKAVQEDSSKSNAWNRSTTADQTKTKSNEWSA**WG**SNKSEI
PAGGSKAVQEDSSKSNAWNTSTTADQTKTKSNEWSA**WG**SNKSEIPAGGSKAVQEDSSKAWNTSTTADQTKT
KSNEWSARVSNKSEIPAGGSKAVQEDS**WG**SSKWKADVAQEDNSRLGAWDANAADQTKSNEWS**GWG**KKKDVT
QEDNVQHS**WG**SGKRKDKVTQEDNSGSGD**WG**ANRTDLAITKSSEWSS**WG**KNKTEIPAGGSANVQNDS**WG**LGK
LNDTQKDNSGCGA**WG**ENSGSAWPQEDAWNSGNWKAESKVGNTT**WG**KPKSSESHAWDSHNQSNQNSSSQ**GW**E
SHIASANSENEKGFQ**WG**KGRDSNRPPRAPGQRLDIYSSEEQDVLKDIEPIMQSIRRIMQQQGYSDGDPLAA
EDQLFVLENVFEHHPDKETKMGAGIDYVMVNKHSSFQESRCFYVVCKDGQSK**DFSYRK**CLANYISKKYPDL
AESFLGKYFRKPRARGDQTATLGGDQTATPAQDEAATSGPGQRQE*

>Brachypodium_distachatyon_NRPE1 (Bradi4g45070 and Bradi4g45060)
MEEDQSAVLVAEGAIKSIKLSLSTEDEILTYSINDCPVTHPSQLGNPFLGLPLETGKCESCGASENGKCEG
HFGYIELPVPIYHPCHVSELRQLLSLVCLKCLRIKKGKAKQSNGKENVSVTACSYCRDVPALSLKEVKTAD
GAFRLELRAPPRRLMKDSSWNFLDKYGFHHGGASHFRTLLPEEALNILKKIPDDTRKKLAARGYIAQSGYV
MKYLPVPPNCLYIPEFTDGQSIMSYDISISLLKKILHRIEQIKKSRAGTPNFESHEAESSDLQISIAQYIH
LRGTTKGPQDTKRFTISTDSSHLSTKQWLEKMRTLFISKGSGFSSRSVLTGDPYIGVDVVGLPSEVAKRIT
FEEQVTDINIKRLQEVVDKGLCLTYRDGQTTYAITVGSKGYTTLKVGQTISRRIVDGDVVFLNRPPSTHKH
SLQAFYVYIHDDHTVKINPLICSPL**AADFDGD**CVHIYYPQSLAAKAEALELFSVEKQLTNSHNGKVNLQLS
NDSLLALKHMSSRTVLSKESANQLAMLLSFSLPDPAVVKLKPCWTITQIIQGALPAALTCEGGRFLVKDST
VIKLDLAKESVQASFSDLVSSILCVKGPGGALQFLNALQPLLMEYLLLDGFSVSLQDFNVPKVLLEEVHKS
IQEQSLVLEQSRCSKSQFVEMRVDNNLKDVKQQISDFVVESSHLGLLIDPKSEPSMSKVVQQLGFVGLQLY
REGKFYSSRLVEDCFSSFVDKHPPIVGNQHPPEAYGLVQNSYFHGLNPYEELVHSISTREAIVRSSRGLTE
PGTLFKNLMAILRDVVICYDGTVRNICSNSIMQLKYNEDDATDIPSALTPGEPVGVLAATAISNPAYKAVL
DASQSNNTSWASMKEILQTKVSYKNDTNDRKVILFLNDCSCPKKFCKEKAAIAVQNRLKRVTLEDCATDIC
IEYHKQILDGSSEATPALVGHIHLEKARLDMINVSTEDILQKCQEVSLKHGKKKGHLGHLFKKITFSTCDC
SFTQKPMIDGKLPKVPCLQFSFSEDIPMLSESVERAVSVLANSLCDVLLDTIIKGDPRIQEAKIMWVGSDA
QSWVKNTRKVSKGEPTVEIVVEKNEASKQGDAWRIAMDACIPVIDLIDTRRSIPYGIQQVRELLGISCSFD
QIVQRLSTTMKTVAKGILKDHLILVANSMTCTGNLYGFNTGGYRATFRALKVQVPFTESTLFTPMKCFEKA
AEKCHSDALGCVVSSCS**WG**KHAALGTGSSFQILWNENQLKSNKEYGDGLYDFLAMVRTDQEKARYTFLDDV
DYLVEDNAMDDICLSPELNGTHGVPTFEDNFEHQDTQNGNSWENGTKANASWEQNASAGNDSDN**WGGW**SNA
AAAADTGAAKPADQGNSSWDVPATAENDSTD**WGGW**GNEKAKDNRTVSTEPAELDTWSDRGAKKGTDGGGGS
WGKQTNTCEDSGTNLERNSWAKRPSSPSLSTWAKKNSDGGDGTWDKQANSCKKNVEQDSWKNMPVSPARNA
WNKKESSRGDATWEMRASTLEEKKTSESNEGSWEKSNAQKDS**WG**NTQHGSSDKMAVKDNDMQQDP**WG**HIAT
QNINAQDDL**WG**SVAAKAQTSTAENTDAQDDS**WG**AVAAKAQTSTAQES**WG**NVAASPSDNAWKAPPISQTSAA
EHTDAHNDS**WG**IVAAKAQTSTAQQES**WG**NATASPSDNAWNAAPMDLDAKQPGSWD**GW**SSALAEDSNKADDS
SNKNK**GW**KSD**GWG**AKGNRRDQRDNPSMPPMRPDERPPRPRFEVPAEAKKILREIEPIVSMVRKIFRESCDG
VRLPLEDEKFIKESILEHHPEKERKVPGEIDHIMVNKHHIFQESRCFYVVLADGTHT**DFSYNK**CMDNYVRK
TYTDAAEHADLVSQMYFKKRDRDRAAAVDGGSTPANASQSTQVMETSQDEAPQEAQPETCVATQEETRVSP
QETPAATTQQEETENNPDSASEADYHSASEAGLPEGV

Metal A

H

(6) 44aa repeats & (18) WG motifs

DeCL

(2) 8aa repeats

Metal A

H

(2) 22 aa repeats & (15) WG motifs

DeCL

>Sorghum_bicolor_NRPE1 (Sb03g046922)
MEDDDPAAAGLTVPEAFIRRVKLSVTSNQEIVSTSPLFPSQDPIPITHCSQLQDNPSLGLPLQDGSTCESC
GATQLDKCDGHFGFIKLPEPIYHPSHIAELGKILNLVCLRCLRLKKPKKVTGKESRFTSCSYCQELSPLCV
SQVKKSNGARSLELKLPLKQEVADGFWSFLDQFGFHTSGTSHRRPLHPKEVQDIMKKITEKTRARLAARGY
NLQDGFVMDNMSIPPNCLQISNMLDENTEMCPPTSKGLLHKVLRTIEQIESLNISHPNIEARELGADDLQV
AVADYMNMGGAAKVSQHVTFTRQPAPKQWHKKMKTLFLSKSSSYTCRAVITGDPYIGLDVVGVPDEIARRM
SVQECVTNYNIARLQDMMNKGLCLTYTDLNTNTYDLDGKKGNKKCIMLRVGETVDRRVLDGDLVFLNKPPS
TDMHSIQALYVHVHDDHTIKINPLICGPL**EADFDGD**CVHIFFPRSVLARVEAAELFAVEKQLLNSHNAKLN
FQIKNDYLLALRIMCDRSYSKEKANQIAMFSSGMIPPCNPWTICDRWTIPQILQTTDALRIVPSHPNTVGA
SVTAIITSTLSEKGPREAIKLINLLQPLLMESLLMDGFSISLKDLDGQSAMQKANQSISLEIDKFSKSIVD
FIANSSALGLLVDPKNDSALMNLVEQVGFLGYQLQSTDRLYSNNLVEDCYNFLEKRSGSTKCYDPPKGHDF
VTSSFYNGLNPYEELLHSISVREKIERSSSKGLAEAGNLFKNMMAMLRDVTVCYDGTMRTSYNNSIVQFDS
TNVSSSLTPGDSIGILAATVFANAAYKAVLVPNQKNMTSWDSMKEVLLTNACSKTGTIDQKAILYLNKCFC
GLKFCSELAAHRVQSCLKRIKLEYCAIEVSIKYQQEATQAAQCLVGHIHLDKEQLNWMEITMGNILQTCQK
NVNKHVMKNRQLMQILKTTEIISSEYCLCGQDIGDERALQVSCLQCFIHASTTTVQPESNVIQMMTNTIFP
ILLDTVIKGDPQVQEAKLIWVEPKLTRWVKNSSAEQKGELAVEITVEKIAAAENGGT**WG**VVMDACVPVMDL
IDTTRSAPCNIQEVQKVFGISSVFDRVVQFLMFCPPLGSFFQHLSKAVGMVTKSVLMEHLITVASSMTCTG
SLHGFNRSGSKATFQSLKVQAPFTEATLSRPMQCFRKSAEKVDSDQLDSVVSTCS**WG**NHAAIGTGSAFKIH
WNDENQSASNEILREYNLYDFLEAVGRIGATEQKTDAPHSLCLYDVGQLPEDEVQEDEVVCFGGTSPISWT
DKPKGDSLLHDFMGRAGMWSTVQKHQEMQNKTKWNSASTRGQNKRQFTGQVYARKQPKHSWSQAATHQNNK
LSWCGENVAGAQDFANAESSKG**GW**NRKNSGFGRGGHRGGGRGMAFANAESSSSG**GW**NRKNSGFGRGGRRGG
GRGMWKSEGSHRGGSNSTNWRAQNNNSARQCGISYSFTPVEQQIYTQVEPIIKNVKRIIRESRDGMKLSQD
DEMFIMNKILMYHPEKEKKMAGQGNYIMVNKHQTFPSSRCLYVASSDGSSS**DFSYKK**CLENFIRIHYPHAA
ESFCRKYFK

>Arabidopsis_lyrata_NRPE1 (483042)
MEEESSSEILEGEIVGIKFALATHHEICIASISGSAINHPSQLTNSFLGLPLEFGKCESCGATEPDKCEGH
FGYIQLPVPIYHPAHVNELKQMLSLLCLKCLKIKKAKSTSGGLADRLLGVCCEEASQISIRDRASDGASYL
ELKLPSRSRLQAGCWNFLERYGYRYGSDYTRPLLAREVKEILRRIPEETRKKLTAKGHIPQEGYILEYLPV
PPNCLSVPDVSDGYSSMSVDPSRIELKDVLKKVIAIKSSRSGETNFESHKAEANDMFRVVDTYLQVRGTAK
AARNIDMRYGVSKISDSSSSKAWTQKMRTLFIRKGSGFSSRSVITGDAYRHVNEVGIPIEIAQRITFEERV
SVHNIGYLQKLVDDKLCLSYTQGSTTYSLRDGSKGHTVLKPGQVVHRRVIDGDVVFINRPPTTHKHSLQAL
RVYVHEDNTVKINPLMCSPL**SADFDGD**CVHLFYPQSLSAKAEVMELFSVEKQLLSSHTGQLILQMGCDSLL
SLRVMLEGVFLDKATAQQLAMYGSLTLPPPALRKSSKSGPAWTVFQILQLAFPERLSCKGDRFMVDGSDLL
KFDFGVDAMASIINEIVTSIFLEKGPKETLGFFDSLQPLLMESLFAEGFSVSLEDLSMSRADMDVIHNLII
REISPMVSRLRLSYRDELQLENSLHKVKEVAANFMLKSYSMRNLIDIKSNSAITKLVQQTGFLGLQLSDKK
KFYTKTLVEDMALFCKRKYGRISSSGDFGIVKGCFFHGLDPYEEMAHSIAAREVIVRSSRGLAEPGTLFKN
LMAVLRDIVITNDGTVRNTCSNSVVQFTYGVDSERGHQGLFEAGEPVGVLAATAMSNPAYKAVLDSTANSN
SSWEQMKEVLLCKVNFQNTTNDRRVILYLNECHCGKRFCQENAAYTVRNKLKKVSLKDTAVEFLVEYRKQQ
TISEIFGIDSCLHGHIHLDKTLLQDWNISMQDILQKCEDVINSLGQKKKKATDDFKRTSLSVSECCSFQD
PCGRKDSDMPCLMFSYSATDPDLERTLDVLCNTIYPVLLETVIKGDPRICSANIIWNSSDMTTWIRNCHAS
RRGEWVLDVTVEKSAVKQSGDAWRVVIDACLSVLHLIDTKRSIPYSIKQVQELLGLSCAFEQAVQRLSASV
RMVSKGVLKEHIILLANNMTCSGNMLGFNSGGYKALTRSLNIKAPFTEATLITPRRCFEKAAEKCHTDSLS
TVVGSCS**WG**KRVDVGTGSQFELLWNQKETGLDDKEETDVYSFLQMVRSTTNADAYVSSPGFDVTEEEMAEW
AESPERDSALGEPKFEDSAEFQNLHDEGKPSESNWEKSSSWDNGCSGGSE**WG**VSKNTGGEANPESNWEKTT
NVEKEDAWSSWNTKKDAQESSKSDSGVA**WG**LKTKDDDADTTPNWETRPAQTDSIVPENNEPTSDV**WG**HKSG
SDKSWDKKNGGTESAPAA**WG**STDAAV**WG**SSDKKNSETESDAAA**WG**SRDKKNSEVGSGAGVLGPWNKKSSKT
ESDGAT**WG**SSDKTKSGAAAWSSWDKKNMETDSEPAA**WG**SQSKNKPETESGPST**WG**AWDTKKSETESGPA**GW**
GIVDKKNSETESGPAAMGNWDKKKSNTESGPAA**WG**STDAAV**WG**FSDKNNSETESDAAA**WG**SRDKKTSETES
GAAA**WG**S**WG**QPTPTAANEDANEDDENPWVSLKETKSRDKDDKERIQ**WG**NPAKKFPSSG**GW**SNGGGADWKGK
RNHTPRPPRSEDNLAPMFTATRQRLDSFTSEEQELLSDVEPVMRTLRKIMHPSAYPDGDPISDDDKTFVLE
KILNFHPQKETKLGSGVDFITVDKHTIFSDSRCFFVVSTDGAKQ**DFSYRK**SLNNYLMMKYPDRAEEFIDKY
FTKPRPSGNRDRNNQDATPPGEEQSQPPTQSIGNGGDDFNTQTQSPSQTQAQAQAQAQAQSPSQTQTQSPS
PSQTQTQSPSQTQAQAQSPSQSPSQTQTYS

Metal A

H

(2) 31aa repeats & (3) WG

DeCL

Metal A

H

(10) 16 aa repeats & (17) WG motifs

DeCL

QS-rich

Figure S3. Predicted NRPD1 protein sequences among diverse plant species with key domain features denoted to the right-hand side. The Metal A motif is in black bold type; the NRPD1 signature motif (Erhard et al, 2009) in the DdRP G domain is underlined; the conserved DdRP H domain is underlined in bold; the DeCL signature motif is in blue bold type.

>Arabidopsis_thaliana_NRPD1 (At1g63020)
MEDDCEELQVPVGTLTSIGFSISNNNDRDKMSVLEVEAPNQVTDSRLGLPNPDSVCRTCGSKDRKVCEGHF
GVINFAYSIINPYFLKEVAALLNKICPGCKYIRKKQFQITEDQPERCRYCTLNTGYPLMKFRVTTKEVFRR
SGIVVEVNEESLMKLKKRGVLTLPPDYWSFLPQDSNIDESCLKPTRRIITHAQVYALLLGIDQRLIKKDIP
MFNSLGLTSFPVTPNGYRVTEIVHQFNGARLIFDERTRIYKKLVGFEGNTLELSSRVMECMQYSRLFSETV
SSSKDSANPYQKKSDTPKLCGLRFMKDVLLGKRSDHTFRTVVVGDPSLKLNEIGIPESIAKRLQVSEHLNQ
CNKERLVTSFVPTLLDNKEMHVRRGDRLVAIQVNDLQTGDKIFRSLMDGDTVLMNRPPSIHQHSLIAMTVR
ILPTTSVVSLNPICCLPF**RGDFDGD**CLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTAA
YLVNVEKNCYLNRAQMQQLQMYCPFQLPPPAIIKASPSSTEPQWTGMQLFGMLFPPGFDYTYPLNNVVVSN
GELLSFSEGSAWLRDGEGNFIERLLKHDKGKVLDIIYSAQEMLSQWLLMRGLSVSLADLYLSSDLQSRKNL
TEEISYGLREAEQVCNKQQLMVESWRDFLAVNGEDKEEDSVSDLARFCYERQKSATLSELAVSAFKDAYRD
VQALAYRYGDQSNSFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSFGFPRELTCAAWNDPNSPLRGAKG
KDSTTTESYVPYGVIENSFLTGLNPLESFVHSVTSRDSSFSGNADLPGTLSRRLMFFMRDIYAAYDGTVRN
SFGNQLVQFTYETDGPVEDITGEALGSLSACALSEAAYSALDQPISLLETSPLLNLKNVLECGSKKGQREQ
TMSLYLSEYLSKKKHGFEYGSLEIKNHLEKLSFSEIVSTSMIIFSPSSNTKVPLSPWVCHFHISEKVLKRK
QLSAESVVSSLNEQYKSRNRELKLDIVDLDIQNTNHCSSDDQAMKDDNVCITVTVVEASKHSVLELDAIRL
VLIPFLLDSPVKGDQGIKKVNILWTDRPKAPKRNGNHLAGELYLKVTMYGDRGKRNCWTALLETCLPIMDM
IDWGRSHPDNIRQCCSVYGIDAGRSIFVANLESAVSDTGKEILREHLLLVADSLSVTGEFVALNAKGWSKQ
RQVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALAWGKVPGFGTGDQFEIIISPKVHGFTTPVD
VYDLLSSTKTMRRTNSAPKSDKATVQPFGLLHSAFLKDIKVLDGKGIPMSLLRTIFTWKNIELLSQSLKRI
LHSYEINELLNERDEGLVKMVLQLHPNSVEKIGPGVKGIRVAKSKHGDSCCFEVVRIDGTFE**DFSYHK**CVL
GATKIIAPKKMNFYKSKYLKNGTLESGGFSENP

Metal A
NRPD1 sig
H
DeCL

>Physcomitrella_patens_NRPD1 (phya_90112)(complete?)
MELQDPEAGEAPLAEVMGIQFGILSAKDIVTLSVFEREHSIITAKDLWDSRLGIYNLPGNNNHCQTCGARK
ASDCDGHFGHITLPMPIYHPLHIYFLKKLLNQICLVCKRFKEKVFTLTSYFNSPLQYSSESSDDGKACKWC
GVNNSYETIEMKASVKEGKLPLDYWNFVCGNPERAYNILQSLSKKVIQKLGMDEYVARPEALILHFVPVPP
SGSRITEVDFGSSLPRTHMVGGRRFRFDKQHKLLQRLSFEVKRLQSLRTGMPDWATTKNEVMELQLLASSY
LTGSKWEHGLNPKAYDAVVKSDVQKSDRYMKGHILAKTNNSSARMIVVGDPSIKIEEILLPVFLVEQLTIP
EKVTAFNIERLQRYVDNGPYADLPGRDRVRLHSRLKRMVVEIGDTVHRHIKDGDLVIVNRPPSLTKHAIMA
MEVRLHHSCSLAINPLICAPF**QADFDGD**CMHLFVPQTSEAHAEAHELLKVSNQLINPQGGQSNSALTEDSR
LGAYLMTSSCIFLNKMEVSQLSTSSLVSLPIPAILKSPNKREPLWTGQQLYSTILPEGICYKVTDKKFSTD
VERGILISNGELLVCNGNSNWLGDAFDALTAVIHTSQGPAAALVYLNRAQELANLFLRDRGFSVGLQDFQL
SRDRSQLLRRRLEEVSIGNREALFRTLLMDEHVQREELNKNPASKRGLTAETECIKSKGLYLGATGIVKQV
EALDKVAVDRFQTKFRESTKRLAKDYCKRMNPLLVMINAGSKGSMSKLVQQTISVGLQLFKGEHLLPLNVP
DFCQKQLTDVSTLRATDFLQFERRVPSANLSGYWESRGIITSSYLDGLSPLQFFIHTLSSRYGIMRSKVEE
PNLLLKRLLLFLRNLYVEYDGSVRSLEGQQIVQFKYGRYIEGQRGAITTLEGPKIWCEAGEPVGILAATAI
TEPAYQLKLDSPHNVGAKAIGPLDLINETLSPSNPLKLIDRRVLLRFPLALKSRRHGQENGAMRILQHLKP
VSLSMVATTTMIEYRKAQTVVGEHGRSSPWVGHIRLGVVKLKIYQLLVADLVGSLETQYTNCKFASSHSCQ
FGSSGVTQEQPNPCIHFFVDDSTLVATLDDKEYDEVLSNSLEVMKNVILPILLRTPIKGDARIESVNLLWE
DMEWNPRCTKYLSSKKPCKNGTGELVLEVTVKKECCKSRGKAWKIVTESCLPIMQLLDWQRCTPYSIQELN
HVFGLEAAKGVLLQRLELAIAGMGKPVNLEHLELIADTMVTSGKVSGASLSGYKDLCKTISRSAPFSTAAF
LNPKNSFVVAGRHGISETMEGALSSSV**WG**KAPSLGTGSNFEFFWQAKARELREVCNIREGFDIHEYLAKLNS
SALKPCEGVPVPQHHNESQCVSTTMIQGHCDMVMSPDDFKLKQTNDELEIHLRSKEDFPQVGNHNGVLKQQ
ASSPTHISHPPVTDPIRTEGAVTSRSEACEDSSSFHTPNETLELTRQDSSNSSPCSSFRKDLFPTPVLHDD
SEGDETSGIV

H

>Populus_trichocarpa_NRPD1
CSTCGSRDLKSCEGHFGVINFPYTIVHPYFLSEVVQILNKICPGCKSIRLAKATELITKENPQRKGCKYCA
GNSLGWYPPMKFKVSSKEIFRKTAIIAEIRETLSKKPQKGFKKILAADYWDIFPKDEQEEEEETNAKPNRR
VLSHSQVRHMLKDVDPNFIKLSILKTDTIFLNCFPVTPNSHRVTEVTHAFSNGQRLIFDERTRAYKKMVDF
RGVANTLSFHVMDCLKTSKLNPDKSGNIDPWTAQPKKSNDYVNNASGLRWIKDVVLGKRNDHSFRMVIVGD
PHLQLHEIGIPCHIAERLQISESLTAWNWEKLNACFEKSRFEKGDMHVRREGNLVRVRHMKELRLGDIIYR
PLNDGDTVLINRPPSIHQHSLIALSVKVLPVPSVLAINPLCCPPF**RADFDGD**CLHGYVPQSVDTRVELTEL Metal A
VSLDKQLTNWQSGRNLLSLSQDSLTAAHLVLEDDVFLSSFELQQLQMFRPERFLLPAVKAPSANALVWTGK
QLISMLLPVGFDHDFPSCNVCIRDGDLVSSEGSFWLWDTDGNLFQSLVKHCHGQVLDFLYAAQRVLCEWLS
MRGLSVSLSDLYLCPDSNSRKNMMDEIWYGLQDADYACNLKHLMVDSCRDFLTGNNEEDQCNVERLRFLSG
CSEEDYCVMAFDGERLCYEKQRSAALSQSSVDAFRLVFRDIQSLVYKYASQDNSFLAMFKAGSKGNLLKLV
QHSMCLGLQHALASLSFRIPHQLSCAGWNKQKADDATESAKRYIPHAVVEGSFLSGLNPIECFVHSVTSRD
SSFSDNADLPGTLFRRMMFFMRDLHGAYDGTVRNAYGNQLVQFSYNIDDMDPSGSVDEINNSDGIAGRPVG
PLAACAISE<u>AAYSALDQPISLLEKSPLLN</u>LKNVLECGLKRNSAHQTMSLFLSEKLGRQRHGFEYAALEVQN NRPD1 sig
HLERLLFSDIVSFVRIIFSPQSDGRMHFSPWVCHFHVYKWYILHKVFFSFQEIVKKRSLKVHYIIDALEKQ
CKSKTRFPKVQITSRYALWFLLNTHQIRDWRTIYADTWKEKKETFCITVTIVETSKNEFIELETIQDLMIP
FLLETVIKGFMEIQKVDILWNDKPKIPKSHNRLRGELFLRVHMSRGSDKTRLWNQLMDDCLSIMDLIDWAR
SHPDNIHECCLAYGIDAGWKFFLNNLQSAMSDVGKTVLPEHLLLVANCLSVTGEFVGLNAKGLKRQREHAS
V<u>STPFVQACFSNPGDCFIRAAKAGVVDDLQGSIDALAWGKVPAIGTGQFDIVYSGK</u>GLEFSKPVDVYNLLG H
SQMISTEQNTEFGVLDAQIYKSDKCGAQFLHKFGGCGPKGFKVKEGIPRSFLRRLLTYDDIQRMSYTVRKI
LNKYSVDQQLNESDKSVLMMTLYFHPRRDEKIGIGAKDIKVINHPEYQDTRCFSLVRTDGTIE**DFSYRK**CL DeCL
HNALEIIAPQRAKRYCEKYLTSKVSATDNSG

>Vitis_vinifera_NRPD1
MDNDFLEEQQVPSGLLIGIKFDVSTEEDMGADSGSRRLRSKGCKYCAANSNDWYPTMKFKVSSKDLFRKTA
IIVEMNEKLPKKLQKKSFRPVLPLDYWDFIPKDPQQEENCLNPNRRVLSHAQVHYLLKDIDPGFIKEFVSR
MDSFFLNCLPVTPNNHRVTEITHALSNGQTLIFDQHSRAYKKLVDFRGTANELSCRVLDCLKTSKLRSEKS
TSKDSASKMSGLKWIKEVLLGKRTNHSFRMIVVGDPKLRLSEIGIPCHIAEELLISEHLNSWNWEKVTNGC
NLRLLEKGQTYVRRKGTLAPVRRMNDFQAGDIIYRPLTDGDIVLINRPPSIHQHSVIALSVKVLPLNSVVS
INPLCCSPF**RGDFDGD**CLHGYIPQSVDSRVELSELVALNRQLINRQSGRNLLSLSQDSLSAAHLVMEDGVL Metal A
LNLFQMQQLEMFCPYQLQSPAIIKAPLLDTQVWTGKQLFSMLLPPGFNYVFPLNGVRISDGELISSSDGSA
WLRDIDGNLFSSLVKDCQGKALDFLYAAQEVLCEWLSMRGLSVSLSDIYLSSDSISRKNMIDEVFCGLLVA
EQTCHFKQLLVDSSQNFLIGSGENNQNGVVPDVQSLWYERQGSAALCQSSVCAFKQKFRDIQNLVYQYANK
DNSLLAMLKAGSKGNLLKLVQQGLCLGLQHSLVPLSFKIPHQLSCAAWNKQKVPGLIQNDTSEYAESYIPY
AVVENSFLMGLNPLECFVHSVTSRDSSFSDNADLPGTLTRRLMFFMRDLYIAYDGTVRNAYGNQLVQFSYN
IEHTSTPSDGINEDTCAYDMGGQPVGSISACAISE<u>AAYSALDQPISLLEPSPLLN</u>LKRVLECGLRKSTADR NRPD1 sig
TVSLFLSKKLEKRKHGFEYGALEVKNHLEKLLFSDIVSTVMIVFSPQNGSKTHFSPWVCHFHVCEEIAKKR
SLKPHSIIDALYMKCNSARAESKINLPDLQITSNGRDCFVDMEKEDSDCFCITVSIVNSKKSCIQLDTVRD
LVIPFLLGAVWVIPSSIKDAILSWHGLLDVKKVDILWNDNPDSDVLKSSSGRLYLRVYVSGDCGKKNFWGV
LMDACLQIMDMIDWERSHPDNIHDIFVVYGIDAGWKYFLNSLKSAISDIGKTVLPEHLLLVASCLSATGEF
VGLNAKGMARQKELTSI<u>SSPFMQGCFSSPGSCFIKAGKRAVADNLHGSLDALA**WG**KIPSVGSGGHFDILYS</u> H
<u>A</u>KGHELARPEDIYKLLGSQTSCHEQNLKVKVPITCYQTTTKCGAQLVYANGDSASKGCKSLEKISKSVLRS
FLSLNDIQKLSRRLKFILQKYPINHQLSEIDKTTLMMALYFHPRRDEKIGPGAQNIKVRYHSKYHNTRCFS
LVRTDGTEE**DFSYHK**CVHGALEIIDPRRARSYQSRWLPYSEV DeCL

>Oryza_sativa_J_NRPD1-1 (OsJ_15844)
MLLEPELSPGSLGTRTRGEGWMEEPSLEVNNPVAELNAIKFSLMTSSDMEKLSSATIIEMCDVTNAKLGLP
NGAPQCATCGSRSIRDCDGKKKLTGKLLGHFGVIKLAATVHNSYFIEEVVQLLNQICPGCLTLKQNGDTKK
ADGTTIQGTCKYCSKDGSKLYPSIIFKMLTSPRVTLSRSKLHRNTSVMDKMSIIAEVAGGVAHKSKNKAPH
ETLPQDFWDFIPDDNQPPIFNVTKKILSPYQVFHMLKKLDPELINQDDRTKAYKRMVDLYSKKSDDESSAS
TDTYGTKWLKDIILSKRSDNAFRSIMVGDPKINLNEIGIPMGLALNLVVSEQVSSYNFETINLKCNLHLLT
KEVLLVRRNGNLIFVRKANQLEIGDIAYRLLQDGDLVLVNRPPSVHQHSLIALSAKLLSTQSAVSINPLCC
DPF**KGDFDGD**CLHGYIPQCLQSRIELEELVGLSGQLLNQQDGRSLVSLTHDSLAAAHQLTNADVFLEKAEF
QQLQMLSSSISLTPMPSVFKSTNSQGPLWTGKQLFGMLLPYGMNISFDQKLHIKDSEVLTCSSGSFWLQNN
TSSLFSVMFKEYGCKALEFLSSTQDVLCEFLTMWGLSVSLSDLYLFSDHYSRRKLSEEVHLALDEAEEAFQ
IKQILLNSVSIPNLKYYDGGDDRSNTDEQSGFTQVSLPIIRSSMTSFKSVFNDLLKMVQQYVSKDNSMMTM
INSGSKGSVLKFVQQTACVGLQLPASKFPFRIPSQLSCVSWNRHKSLNCEITDGTSECVGGQDMYAVVRNS
FLDGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLRDTYVAYDGTVRSSYGQQIVRFSYDTADGMY
SDHDLEGEPGAPVGSWAACSISEAAYGALDHPVNSLEDSPLMNLQEVLKCHKGTNSLDHTGLLFLSKHLRK
YRYGFEYASLEVKDHLERVDFSDMVDTVIILYGGSDMQKTKGNPWITHFHLNQETMKIKRLGLEFIVREII
DQYNTLRKQLNNAIPSVSISNSETLHLKMENKSGKLGKNLGTGNECVKNQTCCVTMVVQVEINSMSQLDVI
KERVIPSILATLLKGFLEFKNVKVQCQEDNELVLKVGMSEHCKSGKFWATLQNACIPIMELIDWERSRPER
VYDNFCSYGIDSAWKFFVESLRSTTDAIGRNIHRQHLLVVADCLSRPAHSFINAAKRDSVDNLSGTLDAIA
WGKEPCAGSSGPFKILYSGKSHETKQNEHIYDFLHNPEVQALEKNVMDTYRKRTEKTSKRRSALNSEGNAT
INGGAISFNQKFLNAKVGIWENIIDMRTSLQNMLREYTLNEVVTEQDKSCLMEALKFHPRGYDKIGVGIRE
IKIGVNPGHPSSRCFIVLRNDDTTA**DFSYNN**RFPCRYLHSELPEAPPERLRPSHRPSAAACGGGGGNCVV
SSTREKPCKFFLSGDCRYGDECRCYLHAGSINDGFSLLTPLRGHQKEPLLFVGIPDAVKIWDTGAEMSLSE
PTGEYMHWRLAMGCSSLQCNYTSLGCYGKLETGSLAVTYTHNEDHGALALAGMQDAQLNPILLWSTNYNIV
HLYELPSMEEQVRKAVFLNRETFGSQFALAISRIPYSVVEEYTSTGLEELFADVGTWKKQN

>Oryza_sativa_J_NRPD1-2 (OsJ_30285)
MAGGVREGREIEMAPRRATILLGRIGMEEPSLEVKMPEADLKAVKFSLMTSSDMEKLSSASIIEMCDVTNA
KLGLPNGAPQCATCGSQSVRDCDGHFGVIKLAATVHNPCIEEVVQLLNQICPGCLTLKQNGDTKKTDGTTI
QTTCKYCSKDGAKLYPSVIFKMLTSPRVTLSRSKLHRNTSVMDKISIIAEVAGGVTHNSKNKAPHETLPQD
FWDFVPDDNQPPQSNVAKKILSPYQVFHMLKNLDPELINQLYSRKSDGEDPTSPDTYGTKWLKDIILSKRS
DNAFRSIMVGDPKINLNEIGIPTDLALNLVVSEQVSFYNFETINLKCNLHLLTKEVLLVRRNGKLIFVRKA
NKLEIGDIAYRLLQDGDLVLVNRPPSVHQHSLIALSAKLLPIQSAVAINPLCCDPF**KGDFDGD**CLHGYVPQ
TLQSRVELDGLVSLSGQMLNAQDGRSLVSLTHDSLAAAHQLTSADVFLQKAEFQQLQLLCSSISPTPEPSV
VKSANFQGSLWTGKQLFGMLLPSGMNISFDQKLHIKDSEVLTCSSGSFWLQNNTSSVFSVMFKEYGSKALE
FLSSTQDVLCEFLTMKGLSVSLSDFYLFSDHYSRKKLSEEIHLALDEAEEAFQIKQILLNTVSIPNLKHYD
GPDNLSNSHGQSDFTQVSLPIIKSSITGFKSVFNDLLKMVLQHVSKDNSMMAMINSGSKGSVLKFVQQTAC
VGLQLPASTFPFRIPSELSCVSWNRQKSLNCEITNNTSECMAGQNMYAVIRNSFLDGLNPLECLLHAISGR
ANFFSENADVPGTLTRKLMYHLRDTYVAYDGTVRSSYGRQIVQFSYDTADGMNNDHDLEGEPGAPVGSWAA
CSISEAAYGALDHPVNALEDSPLMNLQEVLKCHKGTKSAVHTGLLFLSKYLKKYRYGFEYASLEVKDHLER
VDFSDLVDTETMKIKRLRLGFIVRELIDQYNALRKKLNNMIPSVCISYSKCSVGNECVKNRSCCVTMVAQV
ESNSTSQLDIIKERVIPSILATLLKGFLEFENVKVECQQDSELVVKVGMSEHCKTGKFWATLQNACIPIME
LIDWERSRPERVYDIFCSYGIDSAWKYFVESLRSTTDAIGRNIHRQHLLVVADCLSISGQFHGLSSQGLKQ
QRAWLSISSPFSEACFSRPAYSFINAAKRDSVDNLSGALDAIA**WG**KEPCAGTSGPFKVLYSGKSQKTKQNK
NIYDFLHNPEVQALEKNFMDTYKQRTEKPSKQRSAFSSKGNATINGGTISVNQKFLDSKVGIWENIIDMRT
CLQNMLREYTLNEVVTEQDKSCLIEALKFHPRGYDKIGVGIREIKIGVNPGHPNSRCFIVQRSDDTSA**DFS**
YNKCVLGAANSISPELGSYIEKILSNRAIRPHQL

Metal A

NRPD1 sig

H

DeCL

Metal A

NRPD1 sig

H

DeCL

>Oryza_sativa_I_NRPD1 (OSIGBa0147H17.3)
MEEPSLEVNNPVAELNAIKFSLMTSSDMEKLSSATIIEMCDVTNAKLGLPNGAPQCATCGSRSIRDCDGHF
GVIKLAATVHNSYFIEEVVQLLNQICPGCLTLKQNGDTKKADGTTIQGTCKYCSKDGSKLYPSIIFKMLTS
PRVTLSRSKLHRNTSVMDKMSIIAEVAGGVAHKSKNKAPHETLPQDFWDFIPDDNQPPIFNVTKKILSPYQ
VFHMLKKLDPELINQVTRRRELLFLSCLPVTPNCHRVAEMPYGHSDGPRLAFDDRTKAYKRMVDLYSKKSD
DESSASTDTYGIKWLKDIILSKRSDNAFRSIMVGDPKINLNEIGIPMGLALNLVVSEQVSSYNFETINLKC
NLHLLTKEVLLVRRNGNLIFVRKANQLEIGDIAYRLLQDGDLVLVNRPPSVHQHSLIALSAKLLSTQSAVS
INPLCCDPF**KGDFDGD**CLHGYIPQCLQSRIELEELVSLSGQLLNQQDGRSLVSLTHDSLAAAHQLTNADVF
LEKAEFQQLQMLSSSISLTPMPSVFKSTNSQGPLWTGKQLFGMLLPYGMNISFDQKLHIKDSEVLTCSSGS
FWLQNNTSSLFSVMFKEYGCKALEFLSSTQDVLCEFLTMWGLSVSLSDLYLFSDHYSRRKLSEEVHLALDE
AEEAFQIKQILLNSVSIPNLKYYDGGDDRSNTDEQSGFTQVSLPIIRSSMTSFKSVFNDLLKMVQQYVSKD
NSMMTMINSGSKGSVLKFVQQTACVGLQLPASKFPFRIPSQLSCVSWNRHKSLNCEITDGTSECVGGQDMY
AVIRNSFLDGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLRDTYVAYDGTVRSSYGQQIVRFSYD
TADGMYSDHDLEGEPVAPVGSWAACSISE**AAYGALDHPVNSLEDSPLMN**LQEVLKCHKGTNSLDHTGLLFL
SKHLRKYRYGFEYASLEVKDHLERVDFSDMVDTETMKIKRLGLEFIVREIIDQYNTLRKQLNNAIPSVSIS
NSKCSVGNECVKNQTCCVSMVVQVEINSMSQLDVIKERVIPSILATLLKGFLEFKNVKVQCQEDNELVLKV
GMSEHCKSGKFWATLQNACIPIMELIDWERSRPERVYDNFCSYGIDSAWKFFVESLRSTTDAIGRNIHRQH
LLVVADCLSVSGQFHGLSSQGLKQQRTWLSI**SSPFSEACFSRPAHSFINAAKRDSVDNLSGTLDAIASDMV**
DKEPCTGSSGPFKILYSGKSHETKQNEHIYDFLHNPEVQALEKNVMDTYRKRTEKTSKRRSALNSEGNATI
NGGAISFNQKFLNSKVGIWENIIDMRTSLQNMLREYTLNEVVTEQDKSCLIEALKFHPRGYDKIGVGIREI
KIGVNPGHPSSRCFIVLRNDDTTA**DFSYNK**CVLGAANSISPELGSYIENRRSNRAVRPHQL

Metal A

NRPD1 sig

H

DeCL

>Solanum_lycopersicum_NRPD1 (DQ020654) - incomplete N-terminus
FRTVVVGDPNIELGEIGIPCXXAENLHMAETLSLRNWERMTDLCDLMILQRGGILVRRNGVLVRISVMDGL
QKGDIIHRPLVDGDVVMINRPPSIHQHSLIALSVRILPINSVLSINPLVCSPF**RGDFDGD**CLHGYIPQSID
STIELSELVALKQQLLDGQNGQNLLSLSHDSLTAAHLILEPGVFLDRFQMQQLQMFCPRQLGMTAIVKAPP
GNICYWTGKQLFSLLLPSDLEYVFPSNGVCISEGEIVTSSGGSSWLRDASDNLFYSLVKHNGGDTLDLLYA
AQTVLCEWLSMRGLSVSLSDLYISADSYSRENMIDEVCSGLQEAERLSYIQLLMIKYNKDFLSGNLEESKN
SMGFDFEFMSIMQQKSASLSQASASAFKKVFRDIQNLVYNYASNDNSLLAMLKAGSKGNLLKLVQHNMCLG
LQQSLVPVSFRMPRQLSCDAWNNHKSHLVIEKPHKVPECPGSYIPSAVVKSSFLAGLNPLECFVHSLTTRD
SSFSGHADVSGTLNRKLMFFMRDLYVGYDGTVRNAYGNQIVQFSYYEAEQIASTKVTGEALESHNHAIGGH
PVGSLAACAISE**AAYCALDQPVSALESSPLLN**LKKILESGAGSRTGEKTASMFLSKRLGRWAHGFEYGALE
VKGHLERLLLSEVVSTVMICFSPETRKSTHNCPWVCHFHIDKENVKTRRLKLRSVLDALNMRYRAATTKAG
NDLPNLHITCKDCSVAEVQKEKSEICITVSVVETSKDPSSLLDTLRDVVIPFLLETVIKGFSAFKKVDILW
KELPSPSKSSRGPTGELYLQVFMSESCDRIKFWNALVDSCLQIRDLIDWERSYPDDVHDLTVAYGIDVAWE
YFLCKLHSAVSETGKKILPEHLVLAADSLTTTGEFVPLSAKGLTLQRKAAGV**VSPFMQACFTNPGDSFVRA**
AKMGLSDDLQGSLESLA**WG**KTPSIGTGSSFDIMYSGKGYELAEQINVYTLLRNLVTVDTPNVKVTLGKDGG
MDGMSLVRRLDRLDDLDKKSCKSELSFTKLRSYFSFNDIKKLSQSLKQMLSKYDIGRELNEADKCLAMMAL
QFHPRRNEKIGKGAPKEIKIGYHQEFEGSRCFMVVRSDDTVE**DFSYRK**CMQHALELIAPQKAKTSRWLNGA
SA

Metal A

NRPD1 sig

H

DeCL

>Ricinus_communis_NRPD1 (RCOM_1683300)
MEADLFEERQQLPSALLTAITFGVSTEAEKEKLSVLTIDTVSEVTDSKLGLPNPTNQCSTCGSKDLKSCEG
HFGVIKFPFTILHPYYLSEVVRILNQVCPKCKSIRKESKVRCLNHLNPKLPVLLILLCWYPAMKFSVSSEE
IFRKNVIIAKFSERPTNKSQKRGFKKKLAADYWDIIPKDEQQEENITRPNQRVLSHAQVIHLLENIDPNFI
RKFVLKRDSIFLNCFSVTPNCHRVTEVTHAFSNGQRLVFDDRTRAYKKMVDFRGIAKELSFRVLDCLKTSK
INPDKSVNNDDYMALQRKMNDSSSSSSGLRWIKDVVLGKRNDNSFRMVVVGDPNIKFSEIGIPCPIAERLQ
ISEHLTTWNWDKLNTCCEVRLLEKGDMHVRREGKLVRVRRTKELRIGDIIYRPLNDGDTVLINRPPSIHQH
SLIALSVKVLPATSVLAINPLICAPF**RGDFDGD**CLHGYVPQSVDTRVELRELVALDKQLINVQNGRNLLSF
SQDSLVAAHLVMEDGVLLSLQQMQQLQMFCPHQLFSPAVRKAPSLNGCAWTGKQLISMLLPRGFDHECPSS
DVYIRDGELISSEGSFWLRDTDGNLFQSLIKQCQDQVLDFLYIAQEVLCEWLSMRGLSVSLSDLYLCPDSD
SRENMMDEVLFGLQDAKGTCNMKQFMVDSCRDFLASIDEDEQYSVNFDVEHLCHEKQRSAALSQASVDAFK
HVFRDIQTLGYKYASKDNALMAMFKSGSKGNLLKVVQHSMCLGLQHSLVPLSFRMPLQLSCDAWNKQKAEN
AVECARSYIPSAVVEGCFLTGLNPLECFVHSVTSRESSFSDNADLPGTLTRRLMFFMRDVHAAYDGSVRSA
YGNQLIQFSYNIDEGRSAETYGTAKIVDNYDGMAGKPVGSLAACSISEAAYSALDQPISLLEKSPLLNLKN
VLECGLKKSNAHKSMSLFLSEKLGRRRHGFEYGALKVQDHLERLLFSDIVSVSRIIFSSQSESKTCFSPWV
CHFHVYKEIMKKRNLNVDSIINILNGRCKSNTNLPNVQISCKSCSIADNHREKEETLCITVTIVERSKNSS
TRLATIQDLMIPFLLETVLKGLMEINKVDILWKDWPRISKTHNQPYGELYLRVSMSADSEKTRLWNLLMDY
CLPIMDMIDWTCSRPDNVRDFSLAYGIDAGWKFFLQRLESAISDVGKSVLPEHMLLVANCLSVTGEFVGLN
AKGWKRQREDASVSSPFVQACFSSPGNCFIKAAKAGVKDDLQGSLDALA**WG**KVPSVGTGQFDIVYSGKVKL
LLFLLVKRVKLKTPPSFVVLTVFLETPLINLLVWYSVDQQLNEADKCTLTMALYFHPRKEEKIGSGFKDIK
VVKHPEYQDSRCFSLVRSDGTIE**DFSYRK**CVYGALEIIAPHKARSQIEFFQNSDVVAIIGRITYKLFVGQS
EVKELPWEVVHACGLGKHSNRVISMLCYVQGSCKVDLALCNGLGRRLALVTANRA

Metal A
NRPD1 sig
H
DeCL

>Zea_mays_NRPD1
MELHREPPEAILNAIKFDLMTSTDMEKLSSMSIIEVSDVTSPKLGLPNGSLQCETCGSQRGRDCDGHFGVT
KLAATVHNPYFIDDVVHFLNRICPGCLSPREGIDTKRLEREKVQATCKYCSKDGSKLYPSIVFKTLSSPRV
LLFKSKLHRNASVMERISIVAEAADRMPNRSKGKGSLEGLPLDFWDFVPSENKQVQSNMTKIILSPYQVFY
MLKKSDPELIKQFVSRRELLFLSCLPVTPNCHRVVEIGYGLPDGRLTFDDRTKAYKRMVDVSRRIDDYRQH
PHFSVLASSLVSSRVSECLKSSKLYSKKADGETSTDTYGMKWLKDVVLSKRSDNVFRSIMVGDPKIKLWEI
GIPEDLSSSLVVSEHVSSYNFQSTNLKCNLHLLAKQELFIRRNGKLMFLRKADQLEIGDIAYRPLQDGDII
LINRPPSVHQHSLIALSAKILPIHSVVSINPLCCTPF**AGDFDGD**CLHGYIPQSIRSRVELEELVSLHNQLL
NMQDGRNLVSLTHDSLAAAHLLTSTDVFLKKSELQQLQMLCLSVSTPAPAVIKSMNFQGSLWTGKQLFSML
LPSGMNFSCDTELHIMDSEVLTCSLGSSWLQNNTSGLFSVMFKQYGCKALDFLSSAQEVLCEFLTMRGLSV
SLSDLYMFSDHYSRRKLAEGVKLALYEAEEAFRVKKILLDPINIPVLKCHDETEDVTYRQSDCIQSNPSVI
RSSIMAFKDVFRDLLKMVQQHVSNDNSMMVMINAGSKGSMLKYAQQTACIGLQLPASKFPFRIPSQLSCIS
WNGQKSLNYEAESTSERVGGQNLYAVIKNSFIEGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLR
DIHVAYDGTVRSSYGQQIVQFSYDSVDDLVDKLGAPVGCRAACSISEAAYGALEHPVNGLEDSPLMNLQEV
FKCHKATNSGDHIGLLFLSRHLKKYRYGLEYASLEVKNHLERVNFSDLVETIMIIYDGHDKIRNEGMWTTH
FHINKAMMKKKRLGLRFVVDELAKEYDTTRDQLNNAIPSIRISRRKCLVGDEGVKSSSCCIAVVAHAERNS
ISQLDTIKTRVIPSILDTLLKGFLEFKDVEIQCPHDGELLVKVCMSEHCKGGRFWPTLQNACIPVMELIDW
ELSQPSNVSDIFCSYGIDSAWKYFVESLKSATTDTGRNIRREHLLVIADSLSVTGQFHALSSQGLKQQRTR
LSISSPFSEACFSRPAQSFINAAKQCSVDNLCGSLDAVA**WG**KEPFNGTSGPFEIMHSGKPHEPEQNESIYD
FLCSSKVRNFEKNHLDTRRQSTENASICRLACKSSKGSTTVNGVAITIDQDFLHAKVSIWDNIIDMRTSLQ
NMLREYPLNGYVAEPDKSQLIEALKFHSRGAEKIGVGVREIKIGLNPSHPGTRCFILLRNDDTTE**DFSYHK**
CVQGAADSISPQLGSYLKKLYYRA

Metal A
NRPD1 sig
H
DeCL

>Glycine_max_NRPD1 (Glyma11g02920)

MENIAVLEINAAGQVTGSSLGFPNASDECATCGSKDKRFCEGHFGVIKFPTPILHPYFMSEIAHILNKICP
VCKSIRHKSKVIYLLLVPNTGILSFYELASMDFIITCFLPPIYSSIVFLQGVRLIYGTKRSNDCNYCSAYP
SMKFRVSSNDLFRRTAIIVEVKASKKTLGTEIPADYWNFIPCDAQQEENYVNRRVLSPAQVLNLLNGVDPD
FIEKYIPRKNLLYLNCFPVTPNCHRVTEVPYAISIFNIIIFINCHMGTPNELSSRVLDCLRISKARCSAVL
AFRLCFSFDEMQLNPDKTPNSIFADIQQRKIGENACNSSGLRWIKDVVLGKRNDSSLRTVVVGDPDLELSE
VGIPCHIAESLQVSEYVNRQNREKLLYCCELRLLEKGKIDVCRNGSKVHLYKKEDLQIGDKIYRPLADGDK
VLINRPPSIHQHSMIALTVRVLPISSVVCINPLCCSPL**RGDFDGD**CLHGYIPQSVTARIELNELVALDRQL — Metal A
INGQSGRNLLSLSQDSLTAAYLLMEDGVLLNVYQMQQLQMLSISDKRLIPPAVVKAPSSNSSLWSGKQIFS
MLLPYDFDYSFPSDGVVVSDGELVSSSEASGWLRDSDYNVFQSLVEHYQGKTLNFLYTAQKVLCEWLSMTG
FSVSLSDLYLSSDSYARKNMIEEIFYGLQDAEQAYKYLLLSVKRQLMLLGKFFAIFKAGSKGNLLKLVQHS
MCLGMQNSLVRLSYRLPRHLSYVFCSFLTGLNPLECFVHSVTNRDSSFSDHADLPGTLTRRLMFFMRDLHD
AYDGTVRNLYGNQLIQFSYDIEEDSSCDKGFQEYAIGGEPVGAISACAISE<u>AAYSALGQPVSLLETSPLLN</u> — NRPD1 sig
LKNVLECGSRKRNGDQTVSLFLSEKLGKQRHGFEYAALEVKNYLERLLFSNIVSTVMIIFTPHDGSSQEKY
SPWVCHFHLDKEIVTRRKLKVHSIIDSLYQRYYSQRKDSKVCFTNLKISSNILRFSHHHEFLYCSLGFLDV
KKVDVLWNNQSKVKNSCNGFSGELYLRVTLSSEGSRGRFWGVLLNLCHKIMHIIDWTRSHPDNINHFSSAY
GIDAGWQYFFN<u>VCMIKNFPSFNPGSCFIKAAKSGVTDNLQGSLDALA**WG**NCLSMGTSGMFDIIYSE</u>KYFSP — H
CNAHDKCYTGLFLTIDTTSFPYLLIYRKEVDKNSISCYSKNHETTFCPRYKVAKSGNVYELLEASFDKPNN
KAGTHLHKYSSDKCGSEFRHKNGYALKEGKQWKTILRNFVTYCWKVVFVIMPCNEFMLLCLLGKYYSQLGS
RVVNFVLRMDFSRKYSIDELLSESDRSTMLRVLNFHPRKSEKFGIGPQDIKVGWHPKYKDSRCFHIVRIDG
TVE**DFSYRK**CILGALDIVDPKKSKIQEKKWSGHGNT* — DeCL

>Selaginella_moellendorffii_NRPD1 (Smo:441655)

MASSKRRSSHRDRALEEATGTLIALDFRPLTSEEIIRASVYEVKTVRALQNNRFGLPNLSDCCTSCGAKRT
DASNSACPGHSGHIELPVLVYHWDRISALEAILNRVCLHCYSFKHKGRKKELRTLSSLEQVASGVDAHQAD
IGAVPNGARAPEAEENPGKCTGPAAAVKKIFKKVGTANVPALLLEIDGKVRREDIPPGFQSLILKDEMTPQ
WRSKMLDPNQVLRILKCLPQETIDKLRDEKLPSIPAEDYFIKSLPVPPNWMRYSTNEFYFQDKTTKNLKHL
LTKIKSIVYTRDEDKISLLTEQKVMEIQAAATQCIRANPLYGNVSDEDPRYGNVSDESKPLSGLHFLRSLT
GKYCGSSARAVVIGDPALKLEEIGISARIAAGLVVLETVTSSNIIFLQSYAYNNPGLKVVRGGEVCTARSC
KKLQVGDVIHRSLKDGDQVFVNRPPTFHKHALIGLKSKVIRNNVFAVNPLICPPL**FADFDGD**TLALYLPQS — Metal A
LQVRAEVAELVALPKQLVSSQGGQSIIGLTQDALLGAHLMTRKNVFLDKLDMDQLRMWCPSAEVPVPAIVK
SPRKSPLWTGQQLFQMTLPTTFDWESDDGGLIIRQGEILRTSDKSSAWLGKDGLMTTICRRYGPDRALEHL
DIAQGIAVDWISERGFSVGLCDFYMAADAVSRRKLEEETLCAVEEAKISSLAHQIVSDPRFQVNSVSRPRC
NSWNERVQPVTSVNEATQQAAISAFQSTMKAFERTIEEHVRENSRENSLLRMVEANSKGSFSKMMQQGGCL
GLQLRQGEFVYHRVKSLFPRAVENESRGYLTSSELWKSMGLVESSFLDGLDPREFFIHSLSSRKGNDGSQQ — no NRPD1
RCASFFRFLMSYMKDIRVEYDNTIRSTHGGHIFQFSYGATAEPGEPVGLLAGTAVIEPVYDQVMSSSPQAS — signature
TMLKTLQNILFSNSFKDIDRCVTLKLQKLPVQPEWIALQVQDFLKPVTIGMLASKIWIEYSPCSEVGGQKK — motif
RVPWIGCFQLRAEAMERCSLNIDTIVCHLRKLLPTSLDDPDAFIQGLHFFSRDVEVLCFFPITSSVSNYDS
KQIHKHMIGTMFGNLLQVVVKGCPRGIEFVNVKWEDELCIEVAFLSRTRGVPWTHALEACGSISHLVDWQK
STPLSIQEVHVAFGIEAAYQYLLEKLKEFTKGSGVLRKPWKNIDANESGYEAFVKNLSG<u>CSPLAFAMGKSP</u> — H
<u>GGVFEAAAMNREVDYLAGANELAFCGKSPSLGTGANIELFFK</u>EDKGPVSRFPDFESLVFSRRVVDDTVSAT
LSAKDREIVWARIDQRSQKLHDILRKSLTGTPVSAANEAVILDTLKYHPMMDSKVGCGVRHIRVDNHHSFG
GRCFHIVRLDGSVE**DFSYHK**CLLERIKGNTVLVQRYKKKFMGGKNGRKEEVPVEIFSQKNDTGRMYDKKTH — DeCL
GFLLVENHFVPVKTLKKT*

>Sorghum_bicolor_NRPD1 (Sb06g025933)

MELHRELPEATLNAIKFDLMTSTDMEKLSSMSVIEVSDVTSPKLGLPNASPQCETCGSKSGRDCDGHFGVT
KLAATVHNPYFIDDVVHFLNQICPGCLSPREGINMKKDGSKLYPSVIFKTLSSPRVLLSKSKLHRSPSVME
RISIVAEAAERVSNRSKGKGLLEGLPQDYWDFVPSENKQVQSNMTKIILSPYQVFHMLKKSDPELIKQFVS
RRELLFLSCLPVTPNCHRVVEIGYGLSDGRVTFLYSKKTYGETSTDPSGMKWLKDAVLSKRSDNAFRSTMV
GDPKIKLWEIGIPEDLASNLVVSDHVNSYNFENINLKCNLHLLTKEELFIRRNGKLMFLRKADQLEIGDIA
YRPLQDGDLILINRPPSVHQHSLIAFSAKILPIHSVVSINPLCCTPF**LGDFDGD**YGRSLVSLTHDSLAAAH Metal A
LLTSTDVFLKKSEFQQLQMLCLSVLTPVPAVIKSMNFQGSRWTGKQLFSMLLPSGMKFSCDRMLHILNGEV
LTCSLGSSWLQNNTSGLFSVMFKQYGCKALDFLSSAQEVLCEFLTMRGLSVSLSDMFSDHYSRRKLTEGVK
LALDEAEEAFRIKQILLDPINIPVLKCQDETEDVTYRQSDCIQNNPSVIRSSIMAFKDVFSDLLKMVQQHV
SNDNSMMVMINAGSKGSMLKYAQQTACVGLQLPASKFPFRVPSQLSCIRWNRQKSLNYEAEGTNERVGGQN
LYAVIRNSFIEGLNPLECLLHAISGRANFFSENADVPGTLTRKLMYHLRDIHVAYDGTVRSSYGQQIVQFS
YDSADDPVDKLGAPVGCWAACSISEAAYGALEHPVNGLEDSPLMNLQEVFKCHKATNSGDHIGLLFLSRHL NRPD1 sig
KKYRYGLEYASLEVKNHLEQVNFSDLVETIMIMLEMMKKKRLGLRFVIEELTKEYNATRDQLKNAIPSICI
SRRKCVVGDEGVKISACCIAVVALAEPNSMSQLDTIKKRVIPIILDTLLKGFLEFKDVEIQCQHDGELLVK
VCMSHHCKGGRFWATLQNACIPVMELIDWELSRPSNVADIFCSYGIDSAWKYFVESLKSATTDIGRNIRRE
HLLVIADSMSVTGQFHAISSHGLKQQRTRLSISSPFSEACFSRPAQSFIDAAKQCSVDNLCGSLDAIA**WG**K H
EPFNGTSGPFEIMHSGKPHEPEQDESIYDFLRSPKVQNVEKNHLDTRRQSTENASICRLACKSKGSATVNG
VAITSDQDFLHAKVSIWDNIIDMRASLQNMLREYPLNGYVMEPDKSKLIEALKFHPRGAEKIGVGVREIKV
GLNPNHPGTRCFILLRNDDTTE**DFSYHK**CVHGAANSISPQLGSYLKKLYHRA DeCL

>Brachypodium_distachyon_NRPD1 (Bradi2g34870 and Bradi2g34880)

MVRSLLSVIREVTQGSEHSPTKEVQNTGELEKGGVSLPRPAVHLPLLVQGVRAPPRRSSDMSEWTDGPNNE
MDVPMAELKALKFDLLSSADIETLSSANIIEASDVTSAKLGLPNAAPQCVTCGSQNVRDCDGHSGVIKLPA
TVYSPYFLEQLVQFLNQICPGCWTPKQNRDTKRSDAATIQEPCKYCSKDGLYPSVIFKVLTSPRITLSKSK
LQRNTSVMDKVSVTAEVINMSKNKSSLEVLPHDYWNFVPHNQPPQPNTTKILLSPYQVFHILKQVDLELIT
KFAPRRELLFLSCLPVTPNRHRVAEMPYRFSDGPSLAYICMLYSKKTDKESSTDSYGTSVKKNDSYGTKWL
KDAILSKRSDYAFRSIMVGDPKIRLHEIGIPMDLADLFVPEHVSIYNFKSINLKCNLHLLAKELLIARRNG
KLIYVRKENQLEIGDIVYRPLQDGDLILVNRPPSVHQHSLIALSAKLLPVQSVVAINPLNCAPL**SGDFDGD** Metal A
CLHGYVPQSIGSRVELGELVSLSHQLLNMQDGRSLVSLTHDSLAAAHLLTSSGVLLNKTEFQQLQMLCVSL
SPTPVPSVIKSINPQGPLWTGKQLFGMLLPSGMNFSPDPKLHIKDSEVLACSGGSFWLQNNTSGLFSVLFK
QYGGEALEFLSSAQDMLCEFLTMRGLSVSLSDIYLFSDHYSRRKFAEEVNLALDEAEEAFRVTQILLSPNF
IPHLKCYDDCDDLSDSYEQSDFVQSNLPIIKSSIMAFKSVFSDLLKMVQQHTPKDNSMMAMINAGSKGSML
KFVQQAACVGLQLPAGKFPFRIPSELTCASWNRHKSLDCDISEGARKRLGGQNSHAVIRNSFIEGLNPLEC
LLHSISGRANFFSENADVPGTLTKNLMYHLRDIYVAYDGTVRSSYGQQIVQFTYDTAEDIYTDCGQEGEFG
APVGSWAACSISEAAYGALDHPVNVIEDSPLMNLQEVLKCQKGTNSLDHFGLLFLSKNLKKYRYGFEYASL NRPD1 sig
YVQNYLEPMDFSELVNTVMIQYDGGGVQKTKGSPWITHFHISKEMMKRKRLGLRLLVEDLTEHYNAKRDQL
NNVIPKVYISKCKCSDDDDCINNQTCCITVVAQDESNSTSTSQLDDLKKRAIPVLLATPVKGFLEFKDVEI
QCQRDNELVVKVNMSKHCKSGIFWTTLKKACIGIMGLIDWERSRPGSVYDIFCPCGIDSAWKYFVESLRSK
TDDIGRNIHREHLLVVADTLSVSGQFHGLSSQGLKQQRTQLSTSSPFSEACFSRPADTFIKAAKQCSVDNL H
CGNIDALA**WG**KEPPAGTSGPFKIMYAGKPHEPVQNENIYGFLHNPEVWGPEKNHMETDSTRTKNASERWSS
GNATFNGGTISVEQNYLGAKVGVWDSIIDMRTCLQNMLREYQLDEYVVELDKSRVIEALRFHPRGREKIGV DeCL
GIRDIKIGQHPSHPGTRCFILVRNDDTTE**DVSYKK**CVQGAADSISPQLGSHMEKILQTRSFCRDSWR

>Arabidopsis_lyrata_NRPD1 (924683)
MEDDCEELQVPVGTLTSIGFSISNNTDRDTMSVIKVEAPNQVTDSRLGLPNPDSICKTCGSKDRKVCEGHF
GVINFQYSIINPYFLKEIAALLNKICPGCKYIRKKQFQITEDQPERCRYCTSNTGYPLMKFRVTTKEVFRR
SGIVVEVNEESLMKLKKRGVLALPPDYWSFVPQDSNIDESCLKPTRRILTHAQVYALLSGIDQRLIKKDIP
MFDSLALTSFPVTPNGYRVTEIVHQFNGARLVFDERTRIYRKLVGFEGNTLELSSRVIECMQYSRLFSENV
SSSQDSANPYQKKSDTPKLCGLRFMKDVLLGKRSDHTFRTVVVGDPSLKLHEIGIPERIAKRLQVSEHLNN
WNNERLVTFCSPNLFDNKEVHVRRGDRLVAIRVSDLQTGDKIFRNLMDGDTVLMNRPPSIHQHSLIAMTVR
VLPTTSVVSLNPICCLPF**RGDFDGD**CLHGYVPQSIQAKVELDELVALDKQLINRQNGRNLLSLGQDSLTAA
YLVNVEKNCYLNRAQMQQLQMYCPFQLPPPAIIKASPSSTEPQWTGMQLFGMLFPPGFDYTYPLNDVVVSN
GELLSFSEGSAWLRDGEGNFIQGLIKHDKRKVLDIIYSAQEMLSQWLLMRGLSVSLADLYLSSDPQSRKNL
TEEISYGLREAEQVCNKQQLMVESWRDFLAVNGEDEGEDSVARDLARFCYERQKSATLSKIAVSAFKDAYR
DVQALAYRYGEQSNSFLIMSKAGSKGNIGKLVQHSMCIGLQNSAVSLSYGFPRELTCASWNDPNSPLRGAK
GEDSTATESYVPYGVIENSFLTGLNPLESFVHSVTSRDSSFSGNADLPGTLSRRLMFFMRDIYAAYDGTVR
NSFGNQLVQFTYETDGPVEDITGEALGSLSACALSE<u>AAYSALDQPISLLETSPLLN</u>LKNVLECGSKKGQRE
QTMTLYLSETLSKKKHGFEYGSLEIKNHLEKLSFSEIVSTSMIIFSPSTNTKVPLSPWVCHFHISEKVLKR
KQLNVESVVSSLNEQYKSRNRELKLDIVDLDIQSTNHCSSDDKAMKDDSFCITVTVIEASKHSVLELDAIR
LVLIPFLLDSPVKGSQEIKKVDILWTDRPKAPKRNGDHLAGELYLRVTMYGDRGKRNCWTALLETCLPIMD
MIDWSRSHPDNIRQCCSVYGIDAGRSIFVANLESAVSDTGKTILKEHLLLVADSLSVTGEFVALNAKGWSK
QRQVEST<u>PAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALA**WG**KVPGFGTGDQFEIIISPK</u>VHGFTTPV
NVYDLLSSTPPKTNSAPKSDKVTVQPFDLLGTAFLKGIKVLDGKGISMSRLRTIFTWENIEKLSQSLKRIL
TSYEINDPLNGRDEELVMMVLHLHPNSADKIGPGLKGIRVAKSKHGDSRCFEVVRIDGTFE**DFSYHK**CVLG
ATKIIAPKKVNLYKSKYLKNGTHQPGRLSENPQTVK

Metal A

NRPD1 sig

H

DeCL

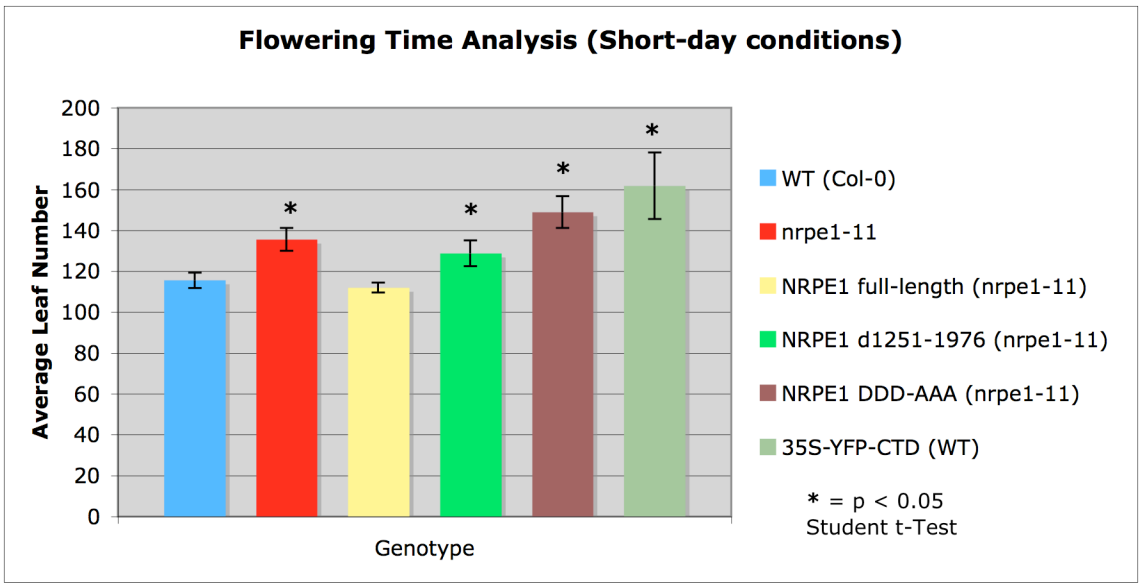


Figure S4. Flowering time experiment with Arabidopsis plants grown under short-day conditions (8 hrs light/16 hrs dark) and randomly rotated every 4 to 6 days. Rosette leaf number was counted when the bolt reached 5 cm in height.

Figure S5. Visible phenotypes observed among wild type Arabidopsis plants transformed with pEarleyGate202-NRPD1 aa1337-1453 (Line #258, T2 generation). Plants display a range of smaller statures and curled rosette leaves. The survival rate was lower than that of other CTD over-expressed domains transformed and planted side-by-side. This rate was not quantified but it took three flats of planted seed to obtain (9) T1 individuals after BASTA selection (~0.5 to 1.0 mL seed planted per flat) compared to the typical single flat that results in at least (30) BASTA survivors.

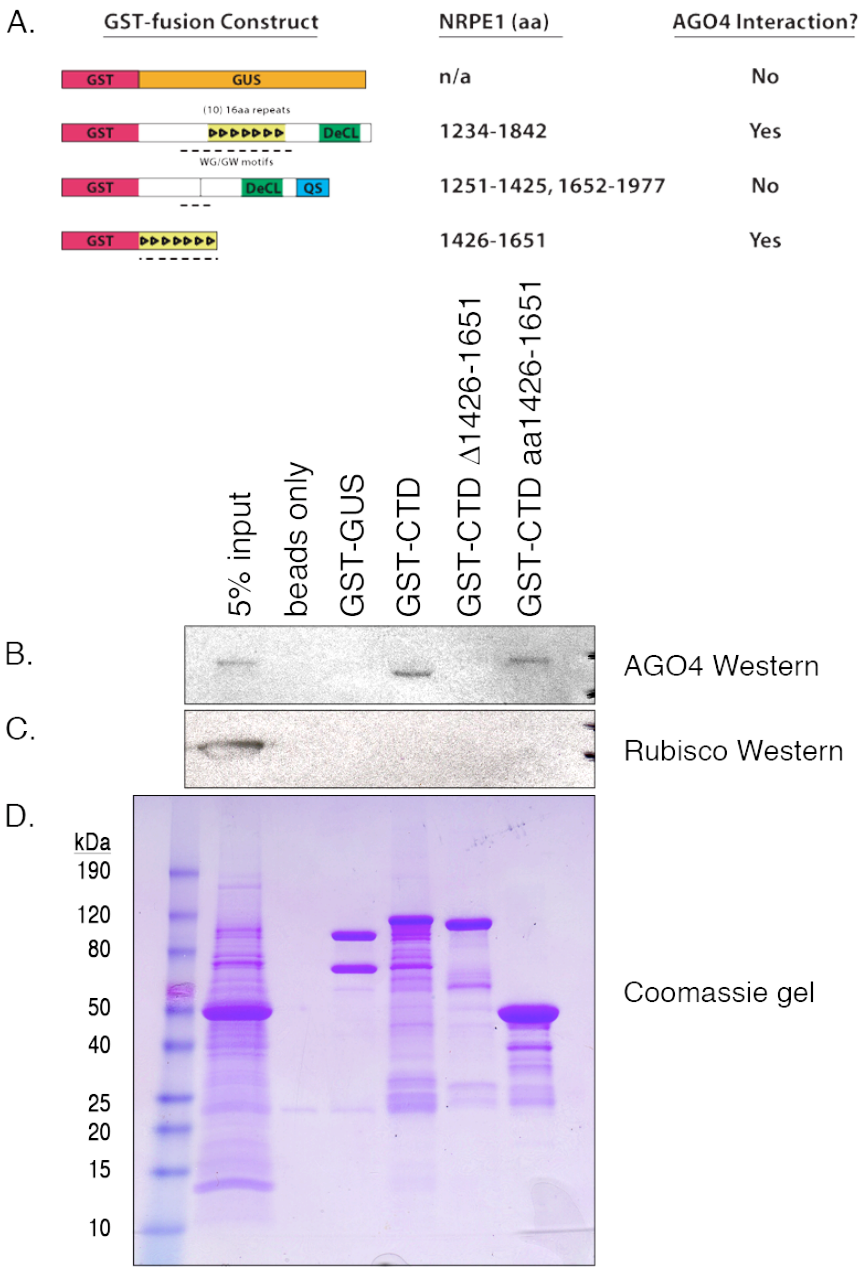


Figure S6. AGO4 *in vitro* interaction with the NRPE1 CTD. (A) Bacterially expressed N-terminal GST tagged constructs used for the *in vitro* protein-protein interaction experiment. Total protein extract from MYC-AGO4 expressing plants was incubated with GST-tagged proteins bound to glutathione resin. The resin was washed and bound proteins analyzed by Western blot. (B) AGO4 Western was performed using the anti-cMyc, clone 9E10. (C) Rubsico Western to demonstrate adequate resin washing. (D) Coomassie stained gel of the eluted bound protein fractions demonstrating roughly equal protein inputs.

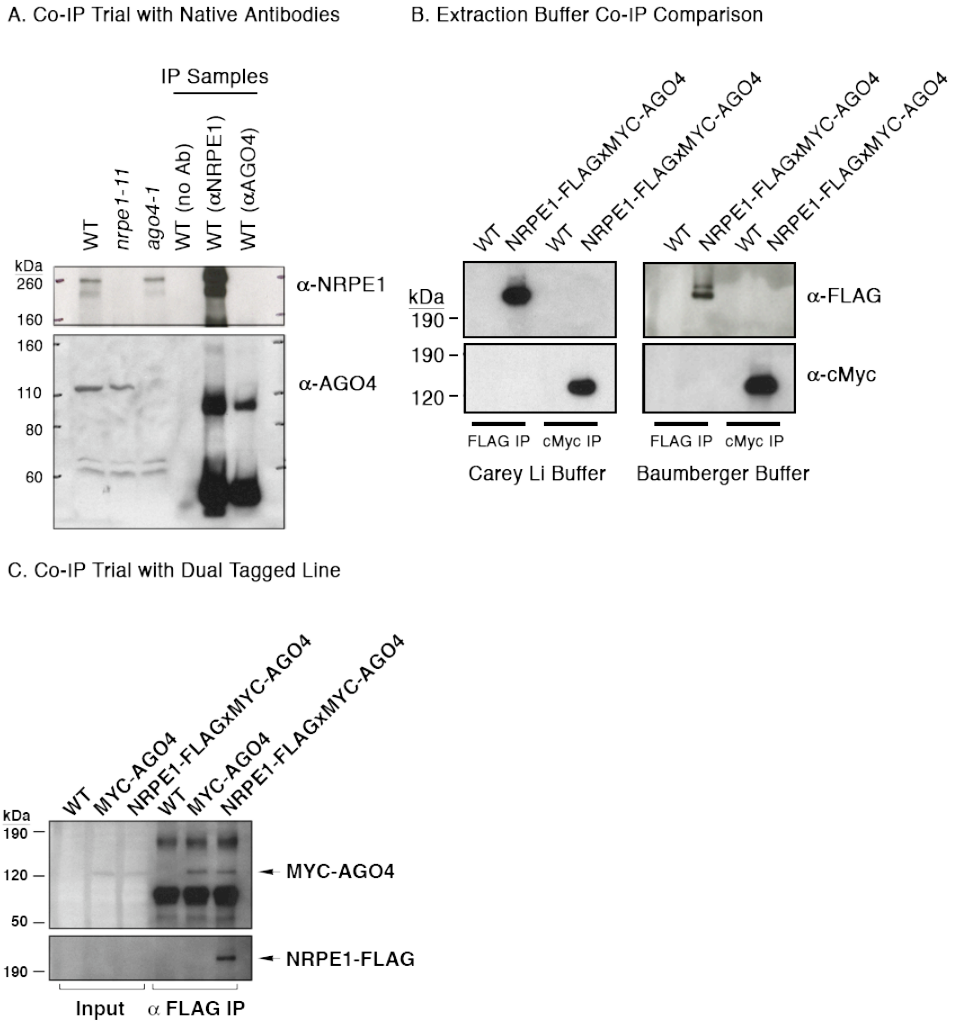


Figure S7. Failure to verify reported NRPE1-AGO4 interaction *in vivo*. (A) Western blot analysis showing lack of co-immunoprecipitation between NRPE1 and AGO4 using native antibodies. Wild type, *nrpe1-11* and *ago4-1* total protein extract controls demonstrate the specificity of these antibodies. (B) A transgenic line bearing both MYC-AGO4 and NRPE1-FLAG genomic constructs was generated by crossing lines from Li et al (2006) and Pontes et al (2006). The possibility exists that the NRPE1-AGO4 interaction is sensitive to buffer conditions so a side-by-side comparison was performed with the extraction buffer and techniques used in the originating report (Li et al, 2006) and the buffer and techniques typically used in the Pikaard lab (Baumberger et al, 2005 with modifications in this manuscript). Reciprocal co-IPs were performed with FLAG and cMyc resin under both conditions. Interaction between NRPE1 and AGO4 was not observed in either immunoprecipitate with either buffer. (C) Western blot analysis showing non-specific IP of MYC-AGO4 with anti-FLAG resin from whole plant extract. This is the only case where an apparent interaction was observed between NRPE1 and AGO4. The result cannot be trusted, though, since the control sample showed immunoprecipitation of MYC-AGO4 with the anti-FLAG resin.

Table S1. Primers used in this study.

Target	Primer	Sequence (5' to 3')	Application
NRPD1 Δ1337-1453	NRPD1-F NRPD1 d1337-R	CAC CGG TGT CTC ACA TTC CAA AGT CCC C CCA TGT AAA GAT CGT TCT AAG CAG TGA CAT AGG AAT	Generate genomic NRPD1 Δ1337-1453 clone; deletes DeCL domain
NRPE1 Δ1251-1976	NRPE1-F NRPE1 d1251-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GAT AAA GAA GAA ACA GAT GTG TAC AGC TTC CTT	Generate genomic NRPE1 Δ1251-1976 clone; deletes entire CTD
NRPE1 Δ1426-1976	NRPE1-F NRPE1 d1426-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CCA CGA TTT GTC TGA AAC AGA TTT GTG TCC	Generate genomic NRPE1 Δ1426-1976 clone; deletes all repeats, DeCL and QS-rich domains
NRPE1 Δ1566-1976	NRPE1-F NRPE1 d1566-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CCC CAT ACC CCA ACC AGC AGG	Generate genomic NRPE1 Δ1566-1976 clone; deletes 4 repeats, DeCL and QS-rich domains
NRPE1 Δ1651-1976	NRPE1-F NRPE1 d1651-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GTC TTC TGC AGT GGG ACT TGG C	Generate genomic NRPE1 Δ1651-1976 clone; last repeat at C-terminus; deletes DeCL and QS-rich domains
NRPE1 Δ1736-1976	NRPE1-F NRPE1 d1736-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A CTC AGA GGT GAA TGA GTC CAA GCG	Generate genomic NRPE1 Δ1736-1976 clone; deletes DeCL and QS-rich domains
NRPE1 Δ1851-1976	NRPE1-F NRPE1 d1851-R	CAC CGC GTA CTA CAA ACG GAA ACG GTC A GAA TTC ATT GAC AAG TAC TTT ACG AAA CCT	Generate genomic NRPE1 Δ1851-1976 clone; deletes QS-rich domain
NRPE1 Δ1251-1426	d1251-1426 mut-F d1251-1426-F d1251-1426 mut-R d1251-1426-R	GTG TAC AGC TTC CTT GAC AAA AAG AAC TGG GGA ACT GAA TCA GC GAC AAA AAG AAC TGG GGA ACT GAA TCA GC AAG GAA GCT GTA CAC ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C	Generate genomic NRPE1 Δ1251-1426 clone using SLIM strategy (Chiu et al., 2004); deletes linker between domain H and CTD internal repeats
NRPE1 Δ1251-1651	d1251-1651 mut-F d1251-1651-F d1251-1651 mut-R d1251-1651-R	GTG TAC AGC TTC CTT AAG GAT ACC AAT GAG GAT GAT AGA AAT CCG TG AAG GAT ACC AAT GAG GAT GAT AGA AAT CCG TG AAG GAA GCT GTA CAC ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C ATC TGT TTC TTC TTT ATC ATC TAG ACC AGT CTG C	Generate genomic NRPE1 Δ1251-1651 clone using SLIM strategy (Chiu et al., 2004); deletes linker and CTD internal repeats
NRPE1 Δ1426-1651	d1426-1651-F d1426-1651-R	GTT TCA GAC AAA TCG TGG AAG GAT ACC AAT GAG CTC ATT GGT ATC CTT CCA CGA TTT GTC TGA AAC	Generate genomic NRPE1 Δ1426-1651 clone using Stratagene strategy; deletes CTD repeats
NRPE1 aa1234-1842	NRPE1 1234-F NRPE1 1842-R	CAC CAA AGA GAC TGG TCT AGA TGA TAA AGA AGA AAC AGA TG TTA GAA TTC TTC AGC ACG GTC AGG GT	cDNA clone of NRPE1 CTD (-QS domain) used for bacterial expression and transgenics
NRPE1 aa1426-1651	NRPE1 1426-F NRPE1 1651-R	CAC CAT GTG GGA CAA AAA GAA CTG GGG AAC TG TCA GTC TTC TGC AGT GGG ACT TGG C	cDNA clone of NRPE1 repeats used for bacterial expression and transgenics
NRPE1 aa1426-1851	NRPE1 1426-F NRPE1 1851-R	CAC CAT GTG GGA CAA AAA GAA CTG GGG AAC TG TCA AGG TTT CGT AAA GTA CTT GTC AAT GAA TTC	cDNA clone of NRPE1 repeats and DeCL used for transgenics
NRPE1 aa1851-1977	NRPE1 1851-F NRPE1 1977-R	CAC CAT GCC TCG GCC TAG CGG AAA CAG TTA TGT CTG CGT CTG GGA CGG	cDNA clone of NRPE QS-rich domain used for bacterial expression and transgenics
NRPD1 aa1337-1453	NRPD1 1337-F NRPD1 1453-R	CAC CAA AAA CAT CGA GTT GCT TTC CCA GTC ATT G TCA CGG GTT TTC GGA GAA ACC AC	cDNA clone of NRPD1 DeCL domain used for transgenics
NRPE1 aa1251-1425, 1652-1977	NRPE1 1251-F NRPE1 1977-R	CAC CCT TCA AAT GGT CAT ATC CAC GAC AAA CGC TTA TGT CTG CGT CTG GGA CGG	cDNA clone of NRPE1 repeat internal deletion used for bacterial expression; cloned from NRPE1 Δ1426-1651-HA total RNA

Table S1. Primers used in this study (continued).

AtSN1	AtSN1-F AtSN1-R	AGG ATT TAT TTC AAT CCA CGA ACC T CGA CTC CCA TAA GTA ACG AGT TG	Chop-PCR (Herr et al., 2005)
At2g19920	AtSN1 control-F AtSN1 control-R	CTC TGG GTT ACC TTT CAG GAA TCA G CTA AAT TGA AGA GCT TAC CTG CTT G	Chop-PCR control (Herr et al., 2005)
AtSN1	AtSN1 RT-F AtSN1 RT-R	ACC AAC GTG CTG TTG GCC CAG TGG TAA ATC AAA ATA AGT GGT GGT TGT ACA AGC	RT-PCR (Herr et al., 2005)
solo LTR	solo LTR-F solo LTR-R	ATC AAT TAT TAT GTC ATG TTA AAA CCG ATT G TGT TTC GAG TTT TAT TCT CTC TAG TCT TCA TT	RT-PCR (Wierzbicki et al., 2008)
Actin	Actin-F Actin-R	TCA TAC TAG TCT CGA GAG ATG ACT CAG ATC ATG TTT GAG TCA TTC TAG AGG CGC GCC ACA ATT TCC CGT TCT GCG GTA G	RT-PCR (Herr et al., 2005)
GAPA	GAPA-F GAPA-R	GGT AGG ATC GGG AGG AAC GAT AAC CTT CTT GGC ACC AG	RT-PCR, glyceraldehyde 3-phosphate dehydrogenase A (Kanno et al., 2005)

EK HAN TAN

1014 E Southdowns Dr, Bloomington IN 47401

(215) 629-6317 ehtan@wustl.edu

EDUCATION

Washington University in St Louis, MO **2005 – 2011**

- **Ph.D.,** Molecular Genetics and Genomics

Duke University, Durham, NC **2001 – 2005**

- **B.S.,** Cum Laude – Double Major with Distinctions in Biology & Visual Arts, Minor in Chemistry

HONORS & AWARDS

- Best Poster Award – DBBS Molecular Genetics Genomics retreat **2007**
- The Mary Duke Biddle Foundation Visual Art Award **2005**
- Undergraduate Research Grant for Distinction Project in Visual Arts **2004**
- Dean's List **2001, 2004, 2005**
- Malaysian Public Services Department Scholarship **1999 – 2005**

RESEARCH & WORK EXPERIENCES

- Talk at Molecular Genetics & Genomics retreat (St Louis, MO) **2010**
  "The role of a RPB9-like subunit of Pol II, IV and V in RNA directed DNA Methylation"
- Talk at Arabidopsis Supergroup meeting (Bloomington, IN) **2009**
  "Major roles for minor subunits of RNA silencing RNA Polymerases"
- Attended the International Congress of Plant and Molecular Biology (St Louis, MO) **2009**
- Undergraduate Mentor, Indiana University **2009 – 2010**
  Laid groundwork and provided guidance on a LUCIFERASE-based suppressor screen for the C-terminal domain of NRPE1 using the *ros1-1* mutant
- Poster & Abstract at Keystone Symposia (Snowmass, CO) **2008**
  Mol. Basis for Chromatin/Epigenetic Phenomena – *"Role of DNA demethylation in the epigenetic on/off switch regulating ribosomal RNA genes"*
- Attended Midwest Meeting on Chromatin, Transcription & Nuclear Dynamics **2007**
- Teaching Assistant: DNA Manipulations Lab, Washington University in St Louis **2007**
  Helped design experiments for the course, office hours & grading
- Lab assistant/Independent Study: Prof. Xinnian Dong, Duke University **2002 – 2005**
  Distinction project: *"Cloning and characterization of SSN2, an activator of plant Immune Response"*

PUBLICATIONS

- **Tan, E.H.,** Ream, T.S., and Pikaard, C.S. (2011) Differential roles for RPB9 subunit variants of RNA polymerase II, IV and V in *Arabidopsis*. *In submission to Genes and Development*.
- Mosher, A. R., **Tan, E.H.,** Shin, J., Fischer, R.L., Pikaard, C.S., Baulcombe, D.C. (2011) Genomic imprinting of Pol IV-dependent siRNAs is regulated by a novel mechanism. *Submitted to PloS One.*
- Haag, J.R., Gu., J., Pontes, O., **Tan, E.H.,** and Pikaard, C.S. (2011) Functional analysis of NRPD1 and NRPE1 C-terminal domains required for RNA-Directed DNA Methylation. *In preparation*.
- Ream, T.S., **Tan, E.H.**, and Pikaard, C.S. (2011) TIBS review on Pol IV and V. *In preparation*.
- Song, J., Durrant, W.E., Wang, S., Yan, S., **Tan, E.H.**, Dong, X. 2011. DNA repair proteins are directly involved in regulation of gene expression during plant Immune Response. *Cell Host & Microbe*, 9(2): 115-124.

RESEARCH FOCUS

- Expertise in molecular biology:
  - Cloning; RNA/DNA manipulations; Site-directed mutagenesis; Southern; Western; Northern Blotting; Antibody design/affinity purification & Quantitative PCR techniques
  - Chromatin Immunoprecipitation (including work with various crosslinkers)
  - Recombinant protein expression/purification from *E. coli* (affinity tag and HPLC methods), plant transformation, complementation & protein complex purification using epitope tags
  - Immunofluorescence, DNA FISH & confocal microscopy experience, PERL programming

PH.D. ABSTRACT

Among eukaryotes, plants have the distinction of encoding multisubunit RNA polymerases used exclusively for RNA directed DNA Methylation (RdDM) in addition to Pol I, II, and III. In *Arabidopsis thaliana*, Pol IV is required for the biogenesis of 24nt siRNAs whereas Pol V transcription is needed for cytosine methylation of the DNA sequences corresponding to these siRNAs. The ancestry of Pol IV and V can be traced back to Pol II, and the extant Pol II, IV and V still share multiple non-catalytic subunits encoded by the same genes. Genetic analysis of non-catalytic subunits that are highly similar reveals that these subunits are not necessarily redundant. For instance, NRPB9b but not its 97% similar paralog, NRPB9a is required for RdDM. Likewise, Pol IV and Pol V-specific 7th largest subunits are very similar yet have different involvements in RdDM. In some of the non-catalytic subunit mutants of Pol IV, 24nt siRNA accumulation is not dramatically reduced, yet RNA silencing is disrupted. This contrasts with Pol IV catalytic subunit mutants in which siRNA biogenesis and RdDM are coordinately disrupted. Taken together, these results suggest that Pol IV might possess functions in RdDM that are in addition to, and separable from siRNA biogenesis. Differences in Pol V subunit composition based on the use of non-catalytic subunit variants might also have functional consequences for RdDM. The evidence we have suggests that alternative non-catalytic subunits in Pol IV and V are likely to influence interactions with other proteins for RdDM.

REFERENCES

- **Craig Pikaard** (Thesis advisor) (812) 855-1874 cpikaard@indiana.edu
- **Doug Chalker** (314) 935-8838 dchalker@biology2.wustl.edu

- **Joe Jez** (314) 935-4899 jjez@wustl.edu