8-30-2012

Analysis of DIE5 and LIA5 reveals the importance of DNA repair in programmed DNA rearrangement of Tetrahymena thermophila

Wan-Yi: Annie) Shieh
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Developmental, Regenerative, and Stem Cell Biology

Dissertation Examination Committee:
Douglas L. Chalker, Chair
Yehuda Ben-Shahar
Sarah C. R. Elgin
Susana Gonzalo-Hervas
Joseph M. Jez
Kristen Kroll
Petra Anne Levin

Analysis of *DIE5* and *LIA5* reveal the importance of DNA repair in programmed DNA rearrangements of *Tetrahymena thermophila*

by
Annie WY Shieh

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 2012

Saint Louis, Missouri

ABSTRACT

During its somatic nuclear differentiation, the single cell eukaryote *Tetrahymena thermophila* undergoes genome-wide programmed DNA rearrangement to eliminate transposon-like elements from its future soma. This process involves small RNA-directed heterochromatin formation followed by extensive nuclear reorganization to form subnuclear domains. While more has been known about small RNAs and heterochromatin, the mechanisms and players involved in the process of nuclear reorganization and the subsequent removal of transposon-like elements from the somatic genome are just starting to unravel. My thesis work centers on the study of two novel nuclear proteins Die5p (Chapter 2) and Lia5p (Chapter 3) and their roles in DNA rearrangement. These essential proteins function downstream of small RNA targeted heterochromatin establishment. While Lia5p is essential for nuclear reorganization to form distinct subnuclear structures, Die5p is a protein conserved across ciliate species and appears to be important for the integrity of the differentiating genome. Maintaining genome integrity during somatic nuclear differentiation has proven to be an active process. Similar to V(D)J recombination during mammalian B and T cell maturation, programmed DNA rearrangement in *Tetrahymena* induces global DNA damage that requires proper response and repair. Through the study of *LIA5* and *DIE5*, we show that nuclear reorganization during *Tetrahymena* DNA rearrangement is intimately associated with the response to DNA damage. Furthermore, we implicate a chromodomain protein Pdd1 as a component of the DNA damage response system, thus providing evidence to support the link between heterochromatin and DNA repair during the reprogramming of *Tetrahymena* somatic genome.

ACKNOWLEDGMENTS

I would like to thank all the individuals that have in one way or another contributed to the preparation and completion of my dissertation. First and foremost, my deepest gratitude to my thesis advisor Doug Chalker who has guided me through my research, encouraged me when things got rough and challenged me to become an independent scientist.

Next, I would like to thank the past and present members of the Chalker lab for all the lively discussions regarding science, life, and much more. Thanks for all the good laughs and drama that turn even the roughest days into a fun one. I will never forget our 'near reality show' experiences. Thank you all for creating a great environment to work in.

Finally, a big thanks to my family and Sidney Wang for their understanding and never-ending support through all the ups and downs that graduate school had created.

TABLES OF CONTENTS

FIGURES AND TABLES

Chapter 1

Introduction

I. Perspective

Chromosomes and protein components of the eukaryotic nucleus exist in non-random distribution. Transcriptionally active euchromatin and relatively inert heterochromatin are partitioned into separate nuclear domains. Furthermore, various sub-nuclear structures such as nucleoli, transcription foci, cajal bodies, splicing speckles etc. have been described (Spector, 2003, 2006). These observations suggest that the patterned organization serves to compartmentalize functional activities in the nucleus (reviewed in Stein et al., 2009). The changes in nuclear landscape observed during human stem cell differentiation (Butler et al., 2009) further highlights the importance of nuclear reorganization in epigenetically regulating developmental specific genes. Elucidating the dynamics and the mechanisms that drive nuclear organization is therefore fundamental to understanding nuclear function.

Due to its unique biology, ciliates have proven to be a prominent model organism that has provided many pivotal insights to the fundamentals of biology. Nobel prize winning discoveries of ribozymes and telomere function are just a few examples. More recently, the study of programmed DNA rearrangement has contributed to our understandings in small RNA-directed heterochromatin formation (Chalker and Yao, 2011). This evolutionary conserved process has been shown to be crucial for proper inheritance and maintenance of genome integrity across metazoans. Moreover, massive chromatin repositioning that accompanies heterochromatin formation during ciliate programmed DNA rearrangement provides a unique opportunity to study nuclear organization (Chalker, 2008). In attempt to study the mechanisms and players involved in

nuclear organization in *Tetrahymena*, we uncovered that this process bridges small RNA-mediated heterochromatin formation and DNA damage repair.

Ciliate genome reprogramming during sexual reproduction involves extensive excision events that remove considerable portion of the germline genome from its developing soma. Domesticated transposases have been recently implicated in such process, followed by DNA double strand break (DSB) repair to reconstitute functioning somatic genome (Baudry et al., 2009; Cheng et al., 2010; Kapusta et al., 2011; Lin et al., 2012). Not only are these important nuclear events integrated for the process of ciliate somatic nuclear differentiation, increasing evidence in other organisms have also implicated an intimate relationship between the major components of small RNA biogenesis, heterochromatin, transposase activity and DNA damage repair. Our understanding of the interplay between these processes is at its infancy.

The scope of my thesis encompasses the studying of two genes, *LIA5* and *DIE5* and their roles in programmed DNA rearrangement during *Tetrahymena* sexual reproduction. The initial goals of these studies were to provide further understanding of how cells reorganize their nuclear content and package unwanted chromatin for silencing during genome reprogramming. However, in the process of elucidating the functions of these two essential genes, the results of my thesis projects uncovered the importance of a heterochromatin component in responding to DNA damage for repair. In turn, we highlight the importance of DNA damage response in the process of *Tetrahymena* DNA rearrangement. Additionally, we provide further evidence to support the connection between heterochromatin and DNA damage repair, and speculate its biological significance in regulating transposable elements during genome reprogramming.

II. Heterochromatin and silencing

In eukaryotes, DNA is wrapped around a histone octamer and organized into higher order chromatin structures. Intricate networks of remodelers and enzymes in turn help to epigenetically compact and organize genomes into discrete nuclear domains. A diverse set of nuclear structures have been described, each associating with a different nuclear activity (Stein et al., 2009). Heterochromatin and euchromatin represent one of the earliest descriptions of nuclear compartmentalization. While the less condensed euchromatin is considered to be more accessible for active transcription, heterochromatin is compact and generally associated with transcriptional silencing. Although these compartments had been generally thought to separate transcriptional activity, and that heterochromatin is associated with gene silencing, increasing evidence reveal that heterochromatin is required for activating expression of certain genes (Lu et al., 2000; Yasuhara and Wakimoto, 2006).

Histones and their modifications play essential roles in regulating heterochromatin. Because different modifications and variants of histones have been found to be differentially enriched on euchromatin relative to heterochromatin, a histone code hypothesis had been proposed as a means to regulate the genome (Jenuwein and Allis, 2001). Major hallmarks of heterochromatin include hypoacetylated histones, tri-methylation of histone 3 lysine 27 (H3K27), and di- and tri-metylation of lysine 9 (H3K9). On the other hand, transcriptionally active chromatin is often hyperacetylated and methylated at H3K4. Heterochromatin protein 1 (HP1) is a major component of heterochromatin. Originally identified in *Drosophila melanogaster*, HP1 has been found

to be highly conserved from yeast to human (Baudry et al., 2009). This protein contains an amino-terminal chromodomain that can interact with methylated H3K9. Similar chromodomain proteins are known to interact with methylated H3K27 (Lachner et al., 2001). In addition, the C-terminal chromoshadow domain allows self-dimerization, creating an interface for interaction with other proteins (Brasher et al., 2000; Mendez et al., 2011). Each of these interacting proteins is believed to contribute to a different aspect of heterochromatin.

Heterochromatin is preferentially associate with genomic regions containing repetitive elements such as transposons and satellite DNA that are found in the centromeres and telomeres (Birchler et al., 2000; Martens et al., 2005). Proper establishment and maintenance of heterochromatin is critical for maintaining genome integrity, as it is essential for silencing potentially harmful DNA elements as well as ensuring proper centromere and telomere functions (Peng and Karpen, 2009). Although some *cis*-acting sequences can promote the formation of facultative heterochromatin, cells combat the repetitive nature of constitutive heterochromatin (such as those found in the telomere and pericentric regions of eukaryotes) by involving small RNA mediated silencing mechanisms. This process has been most extensively described in *Schizosaccharomyces pombe* (Volpe et al., 2002; Verdel et al., 2004), however various evidence have shown that the involvement of RNAi machinery in heterochromatin and gene silencing is conserved across a diverse set of organisms (Malone et al., 2005; Kuramochi-Miyagawa et al., 2008; Fagegaltier et al., 2009; Wang and Elgin, 2011). Mechanisms and players involved differ slightly depending on the lineage and the chromosomal context, however, the similarity of the fundamentals and basic mechanisms

are clear. This process involves the production of small interfering RNAs by the RNAse III family enzyme – Dicer. The small RNAs assemble with effector Argonaut family proteins, which can be divided into two clades: AGO and PIWI proteins. In the case of transcriptional silencing, the small RNAs and Argonaut, together with other factors form the RNA inducing transcriptional silencing (RITS) complex. RITS was first described in *S. pombe* (Volpe et al., 2002), where it has been shown to direct the deposition of H3K9 methylation by histone methyltransferases on sequences that possess homology to the small RNAs loaded into the complex. In *Tetrahymena*, a RITS like complex has also been shown to direct H3K27 methylation (Liu et al., 2007). Methylated histones recruit chromodomain proteins such as HP1 (*Dorsophila* and mammals) or Swi6 (*S. pombe*), consequently nucleating the generation of higher order chromatin structure (determined by sequence complementarity of the small RNAs). Further recruitment of additional chromatin proteins and subsequent propagation in turn leads to the spreading of heterochromatin and silencing of the effective region.

In addition to maintaining genome integrity, heterochromatin has impacts on other nuclear activities since the higher order chromatin structure limits the accessibility of the machineries that act on chromosomes. Relevant to my thesis is the impact of chromatin structure on DNA damage repair. It has been shown that repair within heterochromatin is less efficient than in euchromatin. Particularly, HP1 mediated compaction of chromatin is inhibitory for repair (Goodarzi et al., 2008, 2009). Consequently, mammalian HP1-β is phosphorylated and transiently released from the damaged sites to facilitate chromatin decondensation and repair (Ayoub et al., 2008). Countering this observation, however, HP1 has also been found to localize to sites of

DNA damage. Interestingly, this recruitment is dependent on the chromoshadow domain and independent of the chromodomain. Furthermore, the disruption of HP1 inhibits repair (Ayoub et al., 2009; Luijsterburg et al., 2009). Clearly HP1 exhibits dynamic mobility during DNA damage. The exact role of HP1 in repair remains to be determined, however, it has been proposed that HP1 may act as a component of DNA damage response (DDR) (Luijsterburg et al., 2009) (discussed in the next section). To explain the contrasting observations regarding HP1 recruitment and eviction during DNA damage, a bimodal theory has been proposed to suggest that although HP1 interaction with H3K9 and H3K27 methylation via its chromodomain is transiently released to allow chromatin decondensation, it is subsequently recruited to the site of damage via its chromoshadow domain and may play an active role in the process of repair (Ayoub et al., 2009).

III. DNA damage repair

Double-strand breaks (DSBs) are among the most deleterious DNA lesions. Incorrect repair can cause inversion, translocation, deletion and chromosome fusion. To prevent deleterious consequences, cells have evolved an elaborate system to repair this damage. DSBs are typically repaired by either non-homologous end joining (NHEJ) or homologous recombination (HR). The choice of which repair pathway is used depends upon the cell cycle or developmental stage. Furthermore, different organisms often favor one pathway over the other. For instance, NHEJ is the predominant DSB repair pathway in mammalian cells and is active throughout cell cycle. On the other hand, HR functions in late S-G2 phase and is the dominant repair pathway in budding yeasts.

In higher eukaryotes, NHEJ pathway involves the binding of broken ends by Ku80/Ku70 heterodimers, which in turn recruits the DNA-dependent protein kinase (DNA-PK) complex. Subsequently, the broken ends undergo processing, and are rejoined by the LigaseIV complex, which includes the catalytic domain Ligase IV and its co-factor XRCC4. Cellular response to DNA damage extends beyond simply processing and rejoining the broken ends. To ensure timely detection and repair, upon insult, cells immediately activate a signal transduction network to convey damage. These signals are collectively known as the DNA damage response (DDR) and are important for not only facilitating damage repair, but also regulating cell cycle or inducing apoptosis when damage persists.

DDR is initiated by binding of sensor proteins to DNA breaks. This in turn activates signal transduction proteins, which leads to covalent modification and activation of numerous substrates. These substrates can be proteins that further amplify damage signal, or function directly in DSB repair. One of the key events that occur during DDR is the rapid phosphorylation of histone H2A variant H2AX by ATM and related kinases (ATR and DNA-PK). Phosphorylated H2AX (γH2AX) is critical for DNA damage repair, and appears to act as an interacting platform or recognition signal for repair proteins. Histone modification and recruitment of massive protein complexes to DNA lesion triggers nuclear reorganization and changes in chromatin structure. In fact, one visual response to DNA damage is the formation of nuclear foci where DNA repair occurs (reviewed in Yin and Bassing, 2008; Bao, 2011; Cann and Dellaire, 2011; Lukas et al., 2011). DNA repair foci is an example of compartmentalizing nuclear activities.

IV. Programmed DNA damage

Although potentially harmful to cells, many essential cellular processes involve programmed introduction of DSBs and their subsequent repair. One of the well-known examples is in the programmed DNA rearrangement to generate immunoglobulin and T-cell receptor diversity of vertebrate adaptive immune systems. In this process, programmed DSBs is introduced for V(D)J recombination by the Rag domesticated transposase. Rejoining of the coding regions is then carried out by the NHEJ pathway, involving Ku70/Ku80 heterodimers and the DNA Ligase IV complex (reviewed in Gellert, 2002; Kapitonov and Jurka, 2005; Soulas-Sprauel et al., 2007).

Studies of programmed DNA rearrangements have revealed profound insights into the mechanisms of DNA repair and recombination. Most of these DNA rearrangements involve a few loci, thus affect a limited portion of the genome within the organisms in which they occur. However, for some eukaryotes, programmed DNA rearrangements occur genome-wide. Such large-scale genome reorganizations have been known since Theodor Boveri described chromatin diminution in the parasitic nematodes in 1887. Other whole genome rearrangement phenomena have been describe in hagfish, copepods, ciliates, and most recently, the sea lamprey (Beermann, 1977; Nakai et al., 1991; Tobler and Müller, 2001; Smith et al., 2009, 2010). In my dissertation research, I have investigated the regulation of DNA rearrangement in the ciliated protozoan *Tetrahymena thermophila*.

Like the process of V(D)J recombination, DNA rearrangement in ciliates involve the cutting and rejoining of DNA. Such event triggers programmed DNA damage that requires proper response and repair. Differing from V(D)J recombination, however, the

developmental program in ciliates involves a genome-wide excision event that triggers extensive DNA damage. The presence of phosphorylated H2AX and recent discovery that NHEJ pathway components are essential provide evidence that link ciliate programmed DNA rearrangement to damage and DSB repair.

V. *Tetrahymena* somatic nuclear differentiation involves genome-wide programmed DNA rearrangement

During reprogramming of its somatic genome, ciliate *Tetrahymena thermophila* undergoes genome-wide DNA rearrangement to eliminate transposon-like elements and many non-coding sequences. This process involves massive nuclear reorganization to partition involved chromatin and machineries into distinct subnuclear structures. Since this major event during *Tetrahymena* sexual reproduction can be induced and highly synchronized, it provides a unique opportunity to study the dynamics and mechanisms of nuclear compartmentalization.

The single cell eukaryote *Tetrahymena thermophila* organizes two morphologically and functionally distinct nuclei within the same cell (Figure 1). These two nuclei serve the analogous roles as the germline and the soma of metazoans. Accordingly, they perform their most critical functions to different degrees during the two main stages of *Tetrahymena* cell cycle: vegetative growth and sexual reproduction (via conjugation). The germline micronucleus is diploid, containing five chromosome pairs that are maintained in a transcriptionally silenced state during vegetative growth. In contrast, the somatic macronucleus is polygenomic, containing nearly 200 different chromosomes from which all gene expression occurs to support growth. The 200

macronuclear chromosomes comprise a fragmented subset of the genome represented in the five micronuclear chromosomes.

During sexual reproduction (Figure 2), micro- and macronuclei differentiate from a common zygotic genome derived from the germline of the previous generation. The zygotic genome is formed from the fusion of haploid nuclei of two mating partners that are generated by micronuclear meiosis during the early stages of conjugation. A selected meiotic product in each partner undergoes an additional nuclear division to provide stationary and migratory 'gametic' nuclei, the later of which are exchanged between partners. The fusion of stationary and exchanged haploid nuclei produces genetically identical zygotic nuclei in each partner. These zygotic nuclei complete two rounds of mitotic division to generate the four precursor (anlage) nuclei, two that remain germline in character and two that differentiate into new macronuclei. Thus the micronuclear anlagane are maintained in a transcriptionally quiescent state while the macronuclear anlage differentiates, which induces the activation of gene expression.

The differentiation of the somatic macronuclei requires genome-wide DNA rearrangements. These rearrangement events can be categorized as two major types, chromosome breakage and internal DNA elimination. The germline-derived chromosomes are fragmented into at least 180 Chromosome Breakage Sequences (CBS), a highly conserved 15bp sequence found at all know fragmentation sites. Breakage at the CBS is coupled with *de novo* telomere addition to stabilize the shortened somatic chromosomes. One purpose of fragmentation may be to facilitate chromosome partitioning during the amitotic division of macronuclei (these nuclei do not undergo conventional mitosis and divide without formation of spindles). Concurrent with

chromosome breakages, ~50Mbp of the germline-derived DNA is eliminated from an estimated 5000-6000 loci, which represent >30% of the original germline genome (Yao and Gorovsky, 1974; reviewed in Yao and Chao, 2005; Chalker and Yao, 2011). The sequences eliminated, termed Internal Eliminated Sequences (IESs), range in size from 300bp to >20kbp and are predominantly found in intergenic regions. They are efficiently excised and their macronuclear-destined flanking DNA is rejoined. The newly fragmented and rearranged chromosomes are then subject to several rounds of amplification, resulting in a mature, polyploid somatic genome, which is transcriptionally competent to support vegetative growth in this new generation.

The ~50Mbp sequences removed from the developing somatic nucleus are primarily repetitive sequences that resemble transposable elements (Wuitschick et al., 2002; Fillingham et al., 2004) and other 'junk DNA' in eukaryotes. While other eukaryotes normally silence these elements by packaging them into heterochromatin, *Tetrahymena* faithfully eliminates these sequences during each round of sexual reproduction. It has long been a mystery how many diverse sequences that share little similarity to one another can coordinately undergo excision. Early experiments that identified essential cis-acting sequences furthered this conundrum as no two IESs studied shared common cis-regulatory sequences. The discovery about a decade ago that the recognition of sequence to be eliminated utilizes small RNAs to target specific genomic regions for removal from the developing genome dispelled some of this mystery. More significantly, these studies revealed that DNA elimination is a means of defense against the mobility of transposable elements and showed that the mechanism of IES excision

shares similarity with that by which other eukaryotes package silenced DNA into heterochromatin (reviewed in Slotkin and Martienssen, 2007).

VI. Programmed DNA rearrangement involves small RNA-directed heterochromatin formation (Figure 3)

The mechanism by which *Tetrahymena* DNA elimination achieves sequence specificity is through the action of small RNAs. These small RNAs are produced by bi-directional transcription of the germline genome during meiosis. Although the micronucleus is silent during vegetative growth, upon induction of conjugation, the germline nucleus produces bi-directional transcripts, which are subsequently processed by dsRNA-specific, Dicer-like ribonuclease (Dcl1) into 27-30nt small RNAs, called scan (scn)RNAs (Malone et al., 2005; Mochizuki, 2005). These small RNAs are assembled with a PIWI family protein Twi1, which is required for their stability (Mochizuki et al., 2002). The scnRNAs that are complementary to IESs are then selectively transported into the developing somatic nucleus to guide the marking of the homologous IESs for elimination. Supporting the importance of these RNAs in DNA rearrangements, mutant strains lacking either *DCL1* or *TWI1* fail to accumulate small RNAs, consequently, fail to eliminate IESs and arrest in development prior to completion of conjugation.

RNA-guided genome rearrangement is the ciliate equivalent of small RNA-directed heterochromatin formation. In the developing somatic nucleus, the IESs recognized by complementary scnRNAs are marked with histone modifications that are hallmarks of heterochromatin, histone H3K27 and H3K9 methylation. These chromatin modifications are established by small RNA recruitment of Enhancer of Zeste Like

methyltransferase (Ezl1). Defects in scnRNA production or accumulation cause failed or improper establishment of these histone modifications. Furthermore, like *DCL1* and *TWI1* mutants, cells lacking *EZL1* fail to eliminate IESs and arrest at the same developmental stage (Taverna et al., 2002; Liu et al., 2007).

Not only are IES marked with heterochromatin-specific modifications, but they are also compartmentalized into nuclear substructures. Just prior their elimination, the marked sequences targeted for elimination are found to concentrate in electron dense chromatin structures that bear similarity to heterochromatin found in other eukaryotes (Smothers et al., 1997). It has been proposed that these nuclear foci are the site of IES elimination. During the final stages of nuclear differentiation, disappearance of these condensed structures is accompanied by elimination of the IESs from the future soma.

The involvement of the programmed DNA degradation (Pdd) proteins further links IES elimination with heterochromatin. Pdd1 and Pdd3 contain chromodomains (Madireddi et al., 1996; Nikiforov et al., 2000). Like Drosophila heterochromatin protein 1 (HP1) and Polycomb (Pc) (Paro and Hogness, 1991), Pdd1 and Pdd3 chromodomain bind methylation marks on histone H3 (Taverna et al., 2002; Liu et al., 2007). During conjugation, these proteins are recruited to the IESs in the developing somatic nucleus through their interaction with methylated H3K27 and H3K9. They are found to localize to the IES elimination foci. Due to its abundance and easy visualization, Pdd1 is often used as a marker for foci formation. Like cells lacking other genes required for the process of IES elimination, *PDD1* mutants fail to excise IESs and arrest at the same developmental stage.

Partitioning 50Mbp sequences from 6000 loci into countable distinct nuclear foci necessitates massive nuclear reorganization. The extent of this reorganization event is especially apparent from the dynamic localization of the abundant conjugation protein Pdd1 (Chalker, 2008). Pdd1 localization is initially dispersed throughout the developing genome, but together with the IESs, Pdd1 becomes concentrated into condensed nuclear foci at the onset of IES excision (Madireddi et al., 1996; Smothers et al., 1997). It is likely that such nuclear reorganization serves to compartmentalize DNA rearrangement activity into sub-nuclear domains to ensure efficient regulation. However, the exact function and the mechanism of nuclear reorganization that lead to the formation of these heterochromatin-like structures are not well understood.

VII. Linking DNA elimination to transposon control

As mentioned earlier, many IESs share sequence similarity to transposable elements. Most IESs are repeated sequences found in the micronuclear genome. Eliminated sequences, such as the Tel1 (Cherry and Blackburn, 1985) and REP (Fillingham et al., 2004) elements are clearly related to transposons. The similarity of some IESs to transposons led to the hypothesis that IES excision evolved from the mechanism of transposon insertion. Examination of excision intermediates supported this idea. IES deletion produces intermediates with four-base 5' protruding ends, which resembles the transposition products of some transposons such as Tn7 (Saveliev and Cox, 1995, 1996). For these reasons, it seemed likely that the eliminated sequences in *Tetrahymena* are derived from transposons during evolution, even though many share no transposon homology. The relatively recent discovery that the excision of IESs involves a

domesticated transposase has provided important insight into the mechanism of this process and its evolution.

Domesticated transposases have been identified in diverse eukaryotic lineages. These transposases have been shown to serve critical functions that are essential for the fitness of their hosts. Two well-known examples are RAG1 from Transib transposons that functions in V(D)J recombination of vertebrate immunoglobulin genes (Agrawal et al., 1998; Kapitonov and Jurka, 2005) and Cenp-B from pogo DNA tranposons, involved in the formation of centromere in some eukaryotes (Smit and Riggs, 1996; Casola et al., 2008). Recent efforts in *Tetrahymena* and its distant cousin *Paramecium* have identified the involvement of the domesticated *piggyBac* Transposase in ciliate DNA rearrangement (Baudry et al., 2009; Cheng et al., 2010) In *Tetrahymena*, *TPB2* (*Tetrahymena piggyBac* transposase 2) produces an enzyme with a conserved DDD motif that is capable of catalyzing endonuclease activity *in vitro*. During DNA rearrangement, Tpb2 protein co-localizes with chromodomain protein Pdd1 to the subnuclear DNA rearrangement foci. Disruption of Tpb2 expression inhibits the formation of these foci and abolishes proper DNA rearrangement. These observations suggest that ciliates have adapted to the invasion of transposable elements by employing domesticated transposase to silence them.

As mentioned earlier, ciliates have recruited components of the Non Homologous End Joining pathway to repair the DNA lesions created by Tpb2 during programmed rearrangement. Recent efforts have separately implicated heterochromatin and small RNAs to DNA damage response in other organisms. HP1 recruitment to the site of DNA damage has been found to be essential for repair (Ayoub et al., 2009; Luijsterburg et al.,

2009). Furthermore it has been shown in Arabidopsis, zebrafish and mammalian cell lines that major components of the RNAi machinery and small RNAs generated from the site of DNA damage are crucial for DDR (Francia et al., 2012; Wei et al., 2012). These observations suggest that the interplay between small RNAs and heterochromatin goes beyond gene silencing and leave us pondering upon the evolutionary significance regarding hosts' adaptation in combating transposon invasions. *Tetrahymena thermophila* integrates these pathways into one process – programmed DNA rearrangement, making it a unique model for further studying the interaction between these pathways.

References:

Agrawal, A., Eastman, Q.M., and Schatz, D.G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature *394*, 744–751.

Ayoub, N., Jeyasekharan, A.D., Bernal, J.A., and Venkitaraman, A.R. (2008). HP1-β mobilization promotes chromatin changes that initiate the DNA damage response. Nature *453*, 682–686.

Ayoub, N., Jeyasekharan, A.D., and Venkitaraman, A.R. (2009). Mobilization and recruitment of HP1: a bimodal response to DNA breakage. Cell Cycle *8*, 2945–2950.

Bao, Y. (2011). Chromatin response to DNA double-strand break damage. Epigenomics *3*, 307–321.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Betermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes & Development *23*, 2478–2483.

Beermann, S. (1977). The diminution of Heterochromatic chromosomal segments in Cyclops (Crustacea, Copepoda). Chromosoma *60*, 297–344.

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

Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. Embo J. *19*, 1587–1597.

Butler, J.T., Hall, L.L., Smith, K.P., and Lawrence, J.B. (2009). Changing nuclear landscape and unique PML structures during early epigenetic transitions of human embryonic stem cells. J. Cell. Biochem. *107*, 609–621.

Cann, K.L., and Dellaire, G. (2011). Heterochromatin and the DNA damage response: the need to relax. Biochem. Cell Biol. *89*, 45–60.

Casola, C., Hucks, D., and Feschotte, C. (2008). Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol. Biol. Evol. *25*, 29–41.

Chalker, D.L. (2008). Dynamic nuclear reorganization during genome remodeling of Tetrahymena. Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research *1783*, 2130–2136.

Chalker, D.L., and Yao, M.-C. (2011). DNA elimination in ciliates: transposon domestication and genome surveillance. Annu. Rev. Genet. *45*, 227–246.

Cheng, C.Y., Vogt, A., Mochizuki, K., and Yao, M.C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Molecular Biology of the Cell *21*, 1753–1762.

Cherry, J.M., and Blackburn, E.H. (1985). The internally located telomeric sequences in the germ-line chromosomes of Tetrahymena are at the ends of transposon-like elements. Cell *43*, 747–758.

Fagegaltier, D., Bougé, A.-L., Berry, B., Poisot, E., Sismeiro, O., Coppée, J.-Y., Théodore, L., Voinnet, O., and Antoniewski, C. (2009). The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. Proc. Natl. Acad. Sci. U.S.A. *106*, 21258–21263.

Fillingham, J.S., Thing, T.A., Vythilingum, N., Keuroghlian, A., Bruno, D., Golding, G.B., and Pearlman, R.E. (2004). A non-long terminal repeat retrotransposon family is restricted to the germ line micronucleus of the ciliated protozoan Tetrahymena thermophila. Eukaryotic Cell *3*, 157–169.

Francia, S., Michelini, F., Saxena, A., Tang, D., Hoon, M. de, Anelli, V., Mione, M., Carninci, P., and Fagagna, F. d'Adda di (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature.

Gellert, M. (2002). V(D)J recombination: RAG proteins, repair factors, and regulation. Annu. Rev. Biochem. *71*, 101–132.

Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Löbrich, M., and Jeggo, P.A. (2008). ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. Mol. Cell *31*, 167–177.

Goodarzi, A.A., Noon, A.T., and Jeggo, P.A. (2009). The impact of heterochromatin on DSB repair. Biochem. Soc. Trans. *37*, 569–576.
Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. *3*, e181.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. PLoS Genet. *7*, e1002049.

Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. *22*, 908–917.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Lin, I.-T., Chao, J.-L., and Yao, M.-C. (2012). An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. Mol. Biol. Cell *23*, 2213–2225.

Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes & Development *21*, 1530–1545.

Lu, B.Y., Emtage, P.C., Duyf, B.J., Hilliker, A.J., and Eissenberg, J.C. (2000). Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. Genetics *155*, 699–708.

Luijsterburg, M.S., Dinant, C., Lans, H., Stap, J., Wiernasz, E., Lagerwerf, S., Warmerdam, D.O., Lindh, M., Brink, M.C., Dobrucki, J.W., et al. (2009). Heterochromatin protein 1 is recruited to various types of DNA damage. The Journal of Cell Biology *185*, 577–586.

Lukas, J., Lukas, C., and Bartek, J. (2011). More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. Nat. Cell Biol. *13*, 1161–1169.

Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., and Allis, C.D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell *87*, 75–84.

Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Mol. Cell. Biol. *25*, 9151–9164.

Martens, J.H.A., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. Embo J. *24*, 800–812.

Mendez, D.L., Kim, D., Chruszcz, M., Stephens, G.E., Minor, W., Khorasanizadeh, S., and Elgin, S.C.R. (2011). The HP1a disordered C terminus and chromo shadow domain cooperate to select target peptide partners. Chembiochem *12*, 1084–1096.

Mochizuki, K. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes & Development *19*, 77–89.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Nakai, Y., Kubota, S., and Kohno, S. (1991). Chromatin diminution and chromosome elimination in four Japanese hagfish species. Cytogenet. Cell Genet. *56*, 196–198.

Nikiforov, M.A., Gorovsky, M.A., and Allis, C.D. (2000). A Novel Chromodomain Protein, Pdd3p, Associates with Internal Eliminated Sequences during Macronuclear Development inTetrahymena thermophila. Molecular and Cellular Biology *20*, 4128–4134.

Paro, R., and Hogness, D.S. (1991). The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. Proc. Natl. Acad. Sci. U.S.A. *88*, 263–267.

Peng, J.C., and Karpen, G.H. (2009). Heterochromatic genome stability requires regulators of histone H3 K9 methylation. PLoS Genet. *5*, e1000435.

Saveliev, S.V., and Cox, M.M. (1995). Transient DNA breaks associated with programmed genomic deletion events in conjugating cells of Tetrahymena thermophila. Genes Dev. *9*, 248–255.

Saveliev, S.V., and Cox, M.M. (1996). Developmentally programmed DNA deletion in Tetrahymena thermophila by a transposition-like reaction pathway. The EMBO Journal *15*, 2858.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272–285.

Smit, A.F., and Riggs, A.D. (1996). Tiggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. U.S.A. *93*, 1443–1448.

Smith, J.J., Antonacci, F., Eichler, E.E., and Amemiya, C.T. (2009). Programmed loss of millions of base pairs from a vertebrate genome. Proc. Natl. Acad. Sci. U.S.A. *106*, 11212–11217.

Smith, J.J., Stuart, A.B., Sauka-Spengler, T., Clifton, S.W., and Amemiya, C.T. (2010). Development and analysis of a germline BAC resource for the sea lamprey, a vertebrate that undergoes substantial chromatin diminution. Chromosoma *119*, 381–389.

Smothers, J.F., Mizzen, C.A., Tubbert, M.M., Cook, R.G., and Allis, C.D. (1997). Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures. Development *124*, 4537–4545.

Soulas-Sprauel, P., Rivera-Munoz, P., Malivert, L., Le Guyader, G., Abramowski, V., Revy, P., and de Villartay, J.-P. (2007). V(D)J and immunoglobulin class switch recombinations: a paradigm to study the regulation of DNA end-joining. Oncogene *26*, 7780–7791.

Spector, D.L. (2003). The dynamics of chromosome organization and gene regulation. Annu. Rev. Biochem. *72*, 573–608.

Spector, D.L. (2006). SnapShot: Cellular bodies. Cell *127*, 1071.

Stein, G.S., Zaidi, S.K., Stein, J.L., Lian, J.B., van Wijnen, A.J., Montecino, M., Young, D.W., Javed, A., Pratap, J., Choi, J.-Y., et al. (2009). Organization, integration, and assembly of genetic and epigenetic regulatory machinery in nuclear microenvironments: implications for biological control in cancer. Ann. N. Y. Acad. Sci. *1155*, 4–14.

Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in< i> Tetrahymena</i>. Cell *110*, 701–711.

Tobler, H., and Müller, F. (2001). Chromatin Diminution. In Encyclopedia of Life Sciences, John Wiley & Sons, Ltd, ed. (Chichester: John Wiley & Sons, Ltd),.

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

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

Wang, S.H., and Elgin, S.C.R. (2011). Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc. Natl. Acad. Sci. U.S.A. *108*, 21164–21169.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.-G., and Qi, Y. (2012). A Role for Small RNAs in DNA Double-Strand Break Repair. Cell *149*, 101–112.

Wuitschick, J.D., Gershan, J.A., Lochowicz, A.J., Li, S., and Karrer, K.M. (2002). A novel family of mobile genetic elements is limited to the germline genome in Tetrahymena thermophila. Nucleic Acids Res. *30*, 2524–2537.

Yao, M.-C., and Chao, J.-L. (2005). RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. Annu. Rev. Genet. *39*, 537–559.

Yao, M.C., and Gorovsky, M.A. (1974). Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis. Chromosoma *48*, 1–18.

Yasuhara, J.C., and Wakimoto, B.T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. Trends Genet. *22*, 330–338.

Yin, B., and Bassing, C.H. (2008). The sticky business of histone H2AX in V(D)J recombination, maintenance of genomic stability, and suppression of lymphoma. Immunologic Research *42*, 29–40.

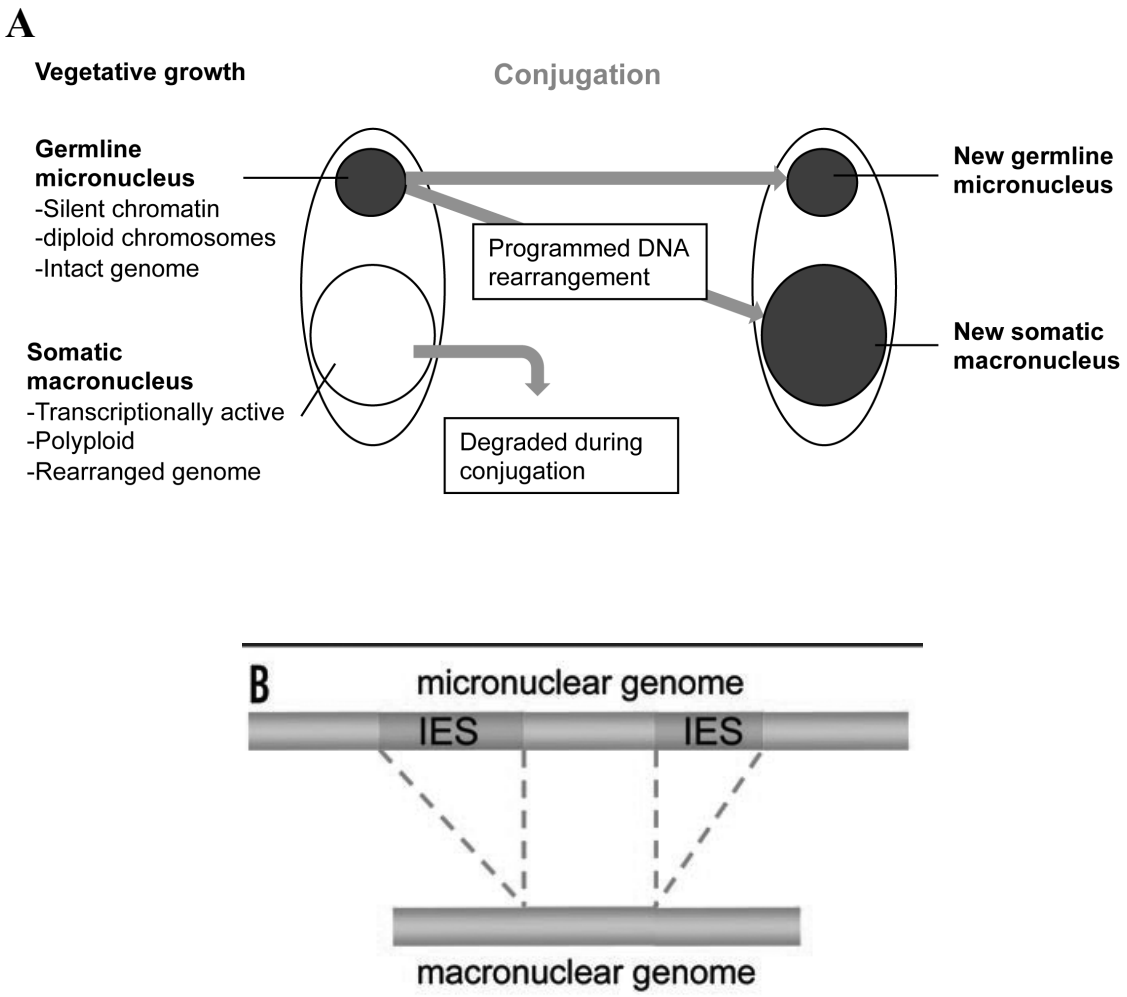

Figure 1 – *Tetrahymena* nuclear dimorphism
(A) *Tetrahymena* contains two functionally and morphologically distinct nuclei within each cell. The germline micronucleus contains 5 diploid chromosomes that are transcriptionally silent during vegetative growth. The macronucleus houses the somatic genome that is transcriptionally active to support vegetative growth. This genome is polyploid and highly fragmented. During sexual reproduction (conjugation), the parental macronucleus is lost. The germline micronucleus undergoes division and cross-fertilization with its mating partner to give rise to both the germline and the somatic nucleus of the next generation. (B) Somatic nuclear differentiation involves genome wide programmed DNA rearrangement that involves chromosome fragmentation and the removal of ~50Mbp transposon-like elements termed Internal Eliminated Sequences (IESs).

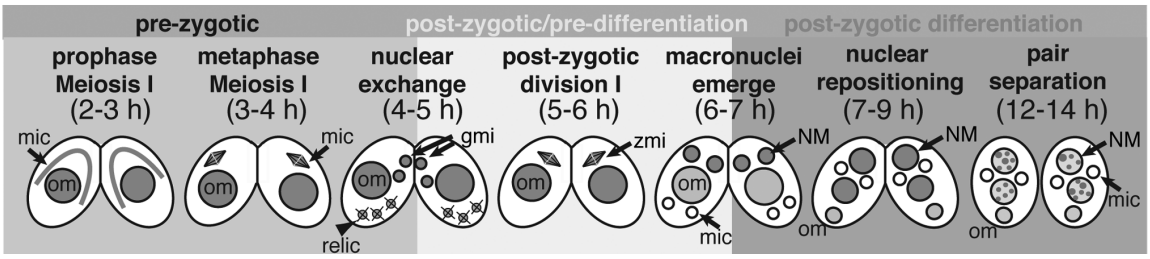

Figure 2 – Major nuclear events during *Tetrahyman* conjugation
Conjugation is initiated by pairing of mating compatible cells. The onset of mating directs the germline micronucleus of each cell to enter meiosis. One of four meiotic products is selected to undergo replication and division to produce two identical gametic micronuclei (gmi) in each mating partner. One from each cell is exchanged and fuses with the partner's stationary copy to produce genetically identical zygotic nuclei. After karyogamy, two additional rounds of mitosis produce the progenitors of the new somatic and germ-line nuclei. Shortly after the second nuclear division, two enlarge and begin to differentiate into macronuclei while the remaining two are preserved silent, one of which is selected as the future germline. Adapted from (Chalker, 2008)

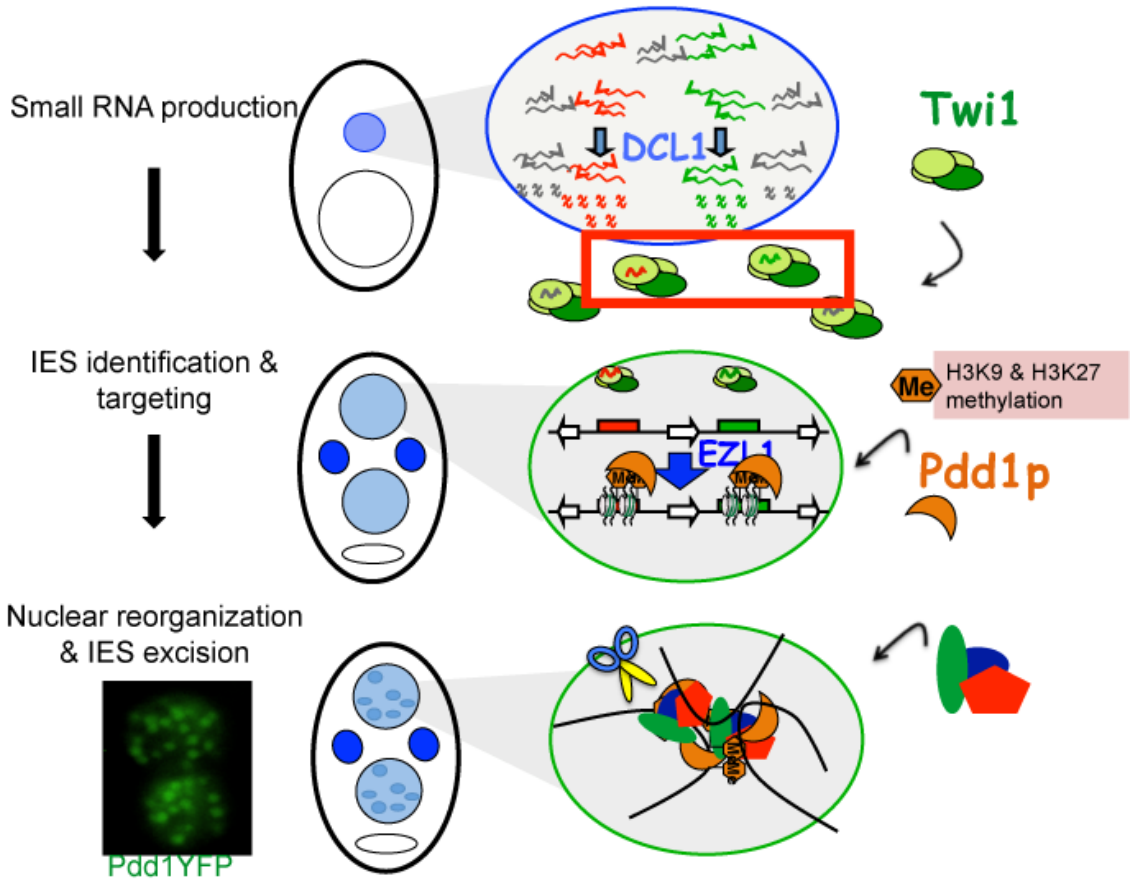

Figure 3 – Model for small RNA directed IES elimination
Progression through development (arrow) and major molecular events are labeled in the left column. Nuclear morphology of the cells at respective developmental stages are depicted next to its corresponding arrows. Small purple circle and the larger white circles represent the parental micronucleus and macronucleus, respectively. The blue small and large circles represent the developing micronucleus and macronucleus, respectively. A fluorescent image and a schematic representation of the nuclear reorganization foci are shown. The right-most column depicts the molecular events of IES elimination described as follows: **small RNA production**: parental micronucleus produces bi-directional transcripts that are processed by Dicer-like ribonuclease (DCL1) into doubled stranded small RNAs. These small RNAs include ones that are complementary to IESs as well as other sequences. Small RNAs are loaded onto Piwi family protein Twi1. The pool of Twi1-RNA complex that contains small RNAs complementary to IESs are selected and transported into the developing macronucleus. **IES targeting**: Twi1-RNA complex identifies the IESs in the developing macronucleus and lead to the recruitment of Enhancer of zest-like methyltransferase (Ezl1). Ezl1 catalyzes H3 K9 and K27 methylation on histones associated with the IESs. These methyl marks recruit chromodomain proteins such as Pdd1. **Nuclear reorganization**: IESs and Pdd proteins are assembled into subnuclear foci, followed by the excision of IESs.

Chapter 2

The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both *Paramecium* and *Tetrahymena*

Matsuda, A., Shieh, A.W, Chalker, D.L., and Forney, J.D.

Eukaryotic Cell, 2010

Summary & Perspective

While genome rearrangements of ciliates have been studied for over three decades, the machineries that carry out these processes are rather poorly described. To uncover genes encoding proteins involved in programmed DNA rearrangement, a screen was performed in *Paramecium* by Atsushi Matsuda and Jim Forney at Purdue University to identify genes highly expressed when these processes occur. This screen identified the developmental specific nuclear protein DIE5 (Defective IES Excision 5) that is conserved between *Paramecium* and *Tetrahymena*. Collaboration between the Forney and Chalker labs to investigate the role of DIE5 in ciliate programmed DNA rearrangement revealed that this gene is essential in both species. Although conserved between evolutionarily distant ciliate species *Tetrahymena* and *Paramecium*, DIE5 has no clear homologs in other organisms. While the exact function of DIE5 remains to be elucidated, we showed that *Tetrahymena* DIE5 (TtDIE5) is required for maintaining genome integrity of the developing somatic nucleus. The lack of TtDIE5 resulted in failure to complete development and caused differentiating macronuclei to eventually lose detectable DNA content even though the nuclear envelope remained intact. This phenotype was at the time unique to *DIE5*. Recent advances in the field since the publication of this paper may shed lights into the significance of this finding. In particular, it has been shown that the major players of the NHEJ DNA repair pathway Ku80 is essential for maintaining macronuclear genome integrity during *Tetrahymena* DNA rearrangement. Knockdown of Ku80 expression resulted in a loss of DNA content from the developing macronucleus as reported for DIE5 mutant. Similarly, DNA damage repair has been implicated in *Paramecium* DNA rearrangement, as repair proteins Ligase IV and XRCC4 are required

for the process. It is likely that DIE5 is a component of a conserved process that mends broken DNA ends during rearrangement. The role of DIE5 in DNA rearrangement is worth revisiting.

Contributions to the work

A. Matsuda and A.W. Shieh contributed equally to this work. Half of the paper describes the identification and role of DIE5 in *Paramecium* while the other half reports the characterization of the Tetrahymena homolog. The identification of the protein and the characterization of its function in *Paramecium* was done by Matsuda and Forney. Except for the initial northern blot analysis, I performed all experiments characterizing DIE5 in *Tetrahymena* with advice and support from D. L. Chalker.

EUKARYOTIC CELL, July 2010, p. 1087–1099 Vol. 9, No. 7
1535-9778/10/$12.00 doi:10.1128/EC.00379-09

The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both *Paramecium* and *Tetrahymena*▽

Atsushi Matsuda,[1]†‡ Annie Wan-Yi Shieh,[2]‡ Douglas L. Chalker,[2] and James D. Forney[1]*

Department of Biochemistry, Purdue University, 175 S. University Street, West Lafayette, Indiana 47907-2063,[1] and Biology Department, Washington University in St. Louis, Campus Box 1137, One Brookings Drive, St. Louis, Missouri 63130[2]

Received 19 December 2009/Accepted 14 May 2010

Development in ciliated protozoa involves extensive genome reorganization within differentiating macronuclei, which shapes the somatic genome of the next vegetative generation. Major events of macronuclear differentiation include excision of internal eliminated sequences (IESs), chromosome fragmentation, and genome amplification. Proteins required for these events include those with homology throughout eukaryotes as well as proteins apparently unique to ciliates. In this study, we identified the ciliate-specific *D*efective in *I*ES *E*xcision 5 (*DIE5*) genes of *Paramecium tetraurelia* (*PtDIE5*) and *Tetrahymena thermophila* (*TtDIE5*) as orthologs that encode nuclear proteins expressed exclusively during development. Abrogation of PtDie5 protein (PtDie5p) function by RNA interference (RNAi)-mediated silencing or TtDie5p by gene disruption resulted in the failure of developing macronuclei to differentiate into new somatic nuclei. *Tetrahymena* Δ*DIE5* cells arrested late in development and failed to complete genome amplification, whereas RNAi-treated *Paramecium* cells highly amplified new macronuclear DNA before the failure in differentiation, findings that highlight clear differences in the biology of these distantly related species. Nevertheless, IES excision and chromosome fragmentation failed to occur in either ciliate, which strongly supports that Die5p is a critical player in these processes. In *Tetrahymena*, loss of zygotic expression during development was sufficient to block nuclear differentiation. This observation, together with the finding that knockdown of Die5p in *Paramecium* still allows genome amplification, indicates that this protein acts late in macronuclear development. Even though DNA rearrangements in these two ciliates look to be quite distinct, analysis of *DIE5* establishes the action of a conserved mechanism within the genome reorganization pathway.

The biology of ciliates offers an extreme case of differential regulation of separate copies of the genome (see reference 34). These protists possess two morphologically and functionally distinct types of nuclei coexisting within a common cell (reviewed in references 20 and 43). The somatic macronucleus is polygenomic and transcriptionally active, whereas the germinal micronucleus is diploid and transcriptionally inert throughout the vegetative growth cycle.

When ciliates undergo development during the sexual phase of the life cycle, the existing somatic macronuclei disappear and new micro- and macronuclei arise from germ line-derived precursors. These genetically identical precursor nuclei are formed upon the conjugation of two mating-compatible partners, which induces meiosis of germ line micronuclei. The products of meiosis and a postmeiotic division are two haploid gametic nuclei in each conjugate, one of which is exchanged between partners. The migratory pronucleus fuses with the stationary copy to form a diploid zygotic nucleus that subsequently divides and differentiates. The new micronuclei are maintained in a silent state, while the differentiating macronuclei acquire extensive chromatin modifications that mediate regulated gene expression. In some ciliate species, nuclear differentiation can alternatively occur through a self-fertilization process called autogamy, which occurs without cell pairing and exchange of genetic material, but nonetheless starts with meiosis and leads to the production of new micro- and macronuclei.

Macronuclear differentiation includes genome-wide DNA rearrangements that extensively remodel the developing somatic chromosomes. The processes of chromosome fragmentation and DNA elimination dramatically alter the genome found in the mature somatic macronucleus after conjugation. Different ciliate species eliminate anywhere from 15% to 90% of the germ line (micronucleus-limited) DNA during this process (reviewed in references 43 and 57). In addition to these physical alterations, the genome is endoreplicated to tens or even thousands of copies per macronucleus, a ploidy level that varies between species.

The DNA segments eliminated from the developing macronucleus are called internal eliminated sequences (IESs), which can vary in structure both within a species and between different species. In *Paramecium tetraurelia*, IESs are generally short (26 to 883 bp) and have a modestly conserved 8-bp inverted repeat that has some similarity to the termini of the mariner/Tc1 superfamily of transposable elements (22). Excision of these IESs is precise, leaving a single copy of the 5′-TA-3′ dinucleotide in the macronucleus-retained sequences. The mechanism generates double-stranded breaks at both ends, followed by joining of the macronucleus-destined DNA (see references 5 and 19). On the other hand, IESs in another

* Corresponding author. Mailing address: Department of Biochemistry, Purdue University, 175 S. University Street, West Lafayette, IN 47907-2063. Phone: (765) 494-1632. Fax: (765) 494-7897. E-mail: forney@purdue.edu.
† Present address: Kobe Advanced ICT Center, National Institute of Communication Technology, Kobe 651-2492, Japan.
‡ A.M. and A.W.-Y.S. contributed equally to this work.
▽ Published ahead of print on 21 May 2010.

ciliate, *Tetrahymena thermophila*, are relatively large (0.6 to >20 kb) and show substantial heterogeneity in their excision boundaries. The elimination of the *Tetrahymena* IESs shows some similarity to the imprecise elimination of transposon-like sequences from the developing genome of *Paramecium* (24). Unlike the *Paramecium* IESs, those in *Tetrahymena* show no obviously conserved sequence characteristics at their excision boundaries. The model proposed for their excision also differs from that of *Paramecium*. Evidence from the characterization of rearrangement intermediates suggests that a double-stranded break at one end of the IES is followed by transesterification of the freed end to the other boundary, generating the macronuclear junction (45, 46).

Despite substantial variability in the size, structure, and excision mechanisms of IESs in different ciliate species, there is a growing body of molecular evidence that IES excision in both *Tetrahymena* and *Paramecium* involves an RNA interference (RNAi)-like pathway (17, 34, 36, 38, 56). This pathway is more extensively characterized for *Tetrahymena*. A large pool of germ line-derived small (27- to 30-nucleotide [nt]) RNAs is generated by cleavage of RNA produced by extensive, bidirectional transcription within meiotic micronuclei (11, 36). These "scan RNAs" (scnRNAs), generated by the Dicer-like protein Dcl1p (30, 37), direct methylation of histone H3 on lysine 9 (K9) and/or lysine 27 (K27), marking the IESs at the beginning of nuclear differentiation for later removal from the genome (26, 50). A role for small RNAs has also been identified in *Paramecium* and possibly in *Stylonychia*, but it is less clear if these act by directing specific chromatin modifications on or near IESs (17, 21, 25). Further studies and comparative analyses should reveal additional biochemical pathways that mediate these remarkable genome reorganization processes. For example, a recent study discovered a domesticated piggyBac transposase in *Paramecium* that is required to remove the IESs from the developing macronuclei and reported that a similar, developmentally expressed transposase gene is contained within the *Tetrahymena* genome (4).

In this report, we show that a gene named *D*efective *I*ES *E*xcision 5 (*DIE5*) is expressed during sexual reproduction and is required for removal of germ line-specific sequences from the *Paramecium* genome. Although significant sequence identity with *DIE5* could not be found in most other eukaryotes, we identified a candidate homologue in *Tetrahymena thermophila* (*TtDIE5*) that is also developmentally regulated. Disruption of *TtDIE5* revealed that it is also required for IES excision and chromosome breakage. Further investigations demonstrated that loss of TtDie5p does not appear to disturb well-characterized steps in macronuclear development, such as small-RNA accumulation and formation of DNA rearrangement foci. Comparison of our results for both species revealed a conserved role for Die5p in the formation of the ciliate macronucleus yet also uncovered significant differences in the action of Die5p between these organisms. We believe this type of comparative study has the power to discern key conserved features of ciliate genome reorganization.

MATERIALS AND METHODS

Cell lines and culture. *Paramecium tetraurelia* stock d4-110 (*hr-b*/*hr-b*) was used to generate cultures undergoing synchronized conjugation for RNA isolation and whole-cell PCR analysis. Strains d4-502 (*pwA-502*/*pwA-502*; *nd6-1*/*nd6-1*) and a3093 (*pwB-96*/*pwB-96*; *nd9-c*/*nd9-c1*) (from Mihoko Takahashi, University of Tsukuba) were used for genetic analysis. Elsewhere, nd6 (*nd6-1*/*nd6-1*) was used as a wild-type control. Paramecia were cultured at 27°C as described by Sonneborn (49) in 1.25 to 2.5 g Austrian winter pea (Outsidepride) in 800 ml double-distilled water (ddH_2O) buffered with K-DS (4 mM sodium citrate, 2.8 mM sodium phosphate dibasic, 1.2 mM potassium phosphate monobasic, 1.5 mM calcium chloride) supplemented with 1.25 mg/liter stigmasterol and inoculated with *Klebsiella pneumoniae* 1 to 2 days prior to use (53). *Tetrahymena thermophila* stocks B2086, CU427, and CU428.1 were cultured axenically in Neff's (0.5% dextrose, 0.25% yeast extract, 0.25% proteose peptone, 3.3 mM $FeCl_3$) or SPP (0.2% dextrose, 0.1% yeast extract, 1% proteose peptone, 0.003% Sequestrene) medium at 30°C as described previously (18, 42).

For conjugation, mating-reactive *Paramecium* cells in starvation medium were mixed at a density of >2,000 cells/ml and incubated at 27°C. Conjugating cells were enriched for by the methods described by Yang and Takahashi (55) and Vosskühler and Tiedtke (54), which resulted in cultures 80 to 99% pure for conjugating *Paramecium*. Conjugating *Tetrahymena* cells were prepared for mating by washing out growth medium and culturing them in 10 mM Tris-HCl (pH 7.4) for >6 h. Mixing cultures of complementary mating types at equal cell density produced cultures with mating efficiencies of 80 to 95% (29).

Total RNA isolation. Total RNA was isolated from 50 to 100 ml of *Paramecium* cell culture (100 to 1,000 cells/ml) with an RNeasy minikit (Qiagen) supplemented by a QIAshredder for cell homogenization and an RNase-free DNase set (Qiagen) for elimination of genomic DNA. All products were used according to the manufacturer's directions. RNA was extracted from 10 to 20 ml *Tetrahymena* cell culture (2×10^5 cells/ml), using RNAzol extraction (14). Total RNA was used in reverse transcription (RT)-PCR to monitor expression in wild-type and Δ*DIE5* lines, using *TtDIE5* oligos 5′-GTTTATGTTTTCTAATTGAGCTTT-3′ and 5′CTGGTATAATCATTAATGCTCG-3′ or HHP1 (*Tetrahymena* HP1 gene) oligonucleotides 5′-GGAGCTTCAACTCATTAAACACG-3′ and 5′-TCGGGAGAAGCATACTTAGCA-3′, which amplify 252-bp or 371-bp cDNA products, respectively.

Microinjection and observation of GFP fluorescence. *Paramecium* plasmids containing the green fluorescent protein (GFP)-*DIE5* fusions were derived from pZC′ΔRI (kindly provided by Eric Meyer, CNRS, Paris, and Jean Cohen, CNRS, Gif-sur-Yvette). Plasmid p5AGN5At contains a 652-bp upstream region, the GFP gene, and the full-length open reading frame (ORF) of *DIE5a* (we designated the gene identified by differential display as *DIE5a* and its paralog as *DIE5b*) followed by the 330-bp *DIE5a* downstream region. Plasmid sequences are available on request. Approximately 2 pl of plasmid solution (~5 μg/μl) in distilled water was injected into the macronuclei of *Paramecium* cells as previously described (32). GFP-expressing cells were fixed and stained with propidium iodide (Vector Laboratories, Burlingame, CA) as described previously (32). Confocal microscopy was performed with a Bio-Rad MRC 1024 UV/Vis system.

To examine Die5p localization in *Tetrahymena*, the *TtDIE5* coding region (genome coordinates CH445530:323,516 to 324,279) was amplified and cloned into the pENTR-D plasmid to create pENTR-*DIE5*, which is compatible with Gateway recombination cloning (Invitrogen). Subsequently, LR Clonase II was used to recombine the *DIE5* coding sequence into a destination vector containing an MTT1-inducible GFP expression cassette cloned upstream of a cycloheximide-resistant *rpl29* allele. This construct was linearized with HindIII in the flanking *rpl29* sequences and introduced into starved *Tetrahymena* cells by biolistic transformation. Transformants were selected in SPP medium containing 12.5 μg/ml cycloheximide. To induce GFP-*DIE5* expression, 0.08 μg/ml $CdCl_2$ was added to mating cells 3.5 h postmixing. Cells at 6 h to 14 h postmixing were fixed in 2% paraformaldehyde, counter stained with DAPI (4′,6-diamidino-2-phenylindole), and visualized, using a Nikon model Eclipse E600 microscope outfitted with a QImaging Retiga EX CCD camera driven by Openlab image acquisition software (Improvision).

RNAi and phenotypic observation. A cDNA fragment corresponding to the region of *DIE5a* between a HincII site and the polyA addition site was cloned into pL4440 (52). (Note that a 23-bp segment in this region perfectly matches *DIE5b* and therefore likely silences both paralogs). RNAi experiments were performed by feeding *Paramecium Escherichia coli* producing double-stranded RNA as previously described (http://*Paramecium*.cgm.cnrs-gif.fr/RNAi/). Conjugation or autogamy of RNAi-treated cells was induced within 48 h of this feeding. Conjugating pairs were isolated in fresh culture medium for phenotypic analysis. For genetic studies, exconjugants were isolated and grown separately for ~10 cell divisions prior to observation of phenotypes. To observe the phenotypes of the F_2 generation, ~10 starved F_1 cells were transferred to fresh culture fluid and allowed to grow for additional cell divisions, and their phenotype was scored after autogamy (self-fertilization). Nuclear DNA was stained with propidium

iodide in Vectashield after cells were fixed with 4% paraformaldehyde as described previously (32).

Micronuclear and macronuclear Δ*DIE5* strains. A *Tetrahymena DIE5* knockout construct was generated, using a MultiSite Gateway cloning kit (Invitrogen). *DIE5* upstream (814 bps; nts 322753 to 323567 of contig CH445530) and downstream (761 bps; nts 324295 to 325056 of contig CH445530) flanking sequences were amplified by PCR and cloned into the Gateway donor vectors, pDONR-P4-P1R and pDONR-P2R-P3, respectively, using BP recombinase. The *MTT1-NEO* (2,079 bps) selection cassette derived from pMNBL (47) was amplified and cloned into donor plasmid pENTR-D by a topoisomerase-mediated reaction. These three donor plasmids were mixed in equal molar ratios with the destination vector pDEST-R4-R3 and LR Clonase Plus, and the resulting recombination created the gene disruption vector pKO-*TtDIE5*.

Plasmid pKO-*TtDIE5* was digested with StuI and introduced by biolistic transformation into either starved CU427 and CU428 populations (macronuclear transformation) or mating B2086 and CU428 populations (germ line transformation) 2.5 to 3.5 h postmixing as described previously (7, 8). Putative transformants with *DIE5* disrupted (Δ*DIE5*) within either their macronuclei or both their macro- and micronuclei were selected by growth in the presence of paromomycin. Micronuclear knockouts were verified by crossing original transformants to CU427 to test for segregation of the *MTT1-NEO* cassette among the cycloheximide-resistant progeny. The heterozygous micronuclear Δ*DIE5* transformants were crossed to the star strains B*VI and B*VII to generate homozygous micronuclear Δ*DIE5* and Δ*DIE5* micronuclei/wild-type macronuclei heterokaryons. For both somatic and germ line transformants, which initially contained a mixture of wild-type and Δ*DIE5* alleles in their macronuclei, cells were subcloned and cultured in growth medium containing increasing concentrations of paromomycin until only mutant alleles remained, thus producing complete macronuclear-knockout strains. The elimination of the *DIE5* gene in the macronucleus and micronucleus was confirmed by PCR and Southern blot hybridization analysis as described previously (30).

Whole-cell PCR amplification. PCR amplifications were performed on whole-cell *Paramecium* as described previously (32) (see Fig. 3).

Northern and Southern blot analyses. DNA from *Paramecium* cultures was isolated as previously described (23). *Tetrahymena* genomic DNA was isolated from 1×10^6 to 2×10^6 cells, using a Wizard genomic DNA isolation kit (Promega), followed by resuspension in 10 mM Tris-HCl by incubation at 65°C for 1 h or at 4°C overnight. Northern and Southern blots were performed as described previously (30). Probes for Northern hybridization were derived from cloned cDNA fragments containing whole *Paramecium* or *Tetrahymena DIE5* ORFs. Plasmids used to generate Southern probes were pSA2.1HP for the 2.1-kb HincII and PstI or 1.5-kb BglII-PstI fragments of the macronuclear *A-51* allele or p4578c containing a 787-bp fragment of the *A-51* allele generated by PCR, using a forward primer (5′-GGATCTGTTGATCAACTAG-3′) and a reverse primer (5′-CTGATAGCGTATTTGGATTAG-3′) with total genomic DNA from exconjugant cells (see probes in Fig. 3A). This reaction amplifies the circularized IES4578 of the *A-51* allele that is present transiently in the genomic DNA of cells during sexual reproduction. To examine the *Tetrahymena DIE5* locus in the knockout lines, isolated genomic DNA was digested with BstBI and separated on 0.9% agarose gel at 40 V overnight. The probe used for analysis of the knockout lines was isolated from pDONR-Die5, created for generating the pKO-TtDie5 plasmid. To assess failure of chromosome breakage, total genomic DNA isolated from wild-type or Δ*DIE5* cells after 16 h of mating was digested with EcoRI, fractionated, and probed with a 0.8-kbp fragment that spans the EcoRI site at position 335013 of chromosomal scaffold CH445662.

Fluorescence microscopy. For examining the nuclear morphology of knockout strains and for cellular localization of GFP-Die5p, cells were fixed in 2% paraformaldehyde and stained with DAPI (1 μg/ml) for 10 to 30 min. Cells were then immobilized under 22- by 22-mm coverslips in 5 μl of 2% methylcellulose. For the visualization of DNA elimination structures, an integrative PDD1-YFP fusion construct was introduced into *Tetrahymena* cells by biolistic transformation. Conjugating transformants induced with 0.05 μg/ml $CdCl_2$ were fixed with 2% paraformaldehyde at 14 h postmixing and were counter-stained with DAPI. For histone modification analysis, 9-h conjugating cells were fixed with Schaudinn's fixative (2 parts saturated mercuric chloride to 1 part 95% ethanol) and dehydrated with methanol. The cells were then rehydrated with Tris-buffered saline (TBS) and blocked in 1% bovine serum albumin (BSA) plus 0.01% Tween 20. Anti-H3K9me2 rabbit polyclonal (Upstate Biotechnology) and anti-H3K27me3 mouse monoclonal (Abcam) antibodies were used at 1:500 dilution for immunostaining. Secondary antibodies used were Alexa Fluor 488-conjugated anti-rabbit and anti-mouse antibodies (1:1,000; Invitrogen).

Nucleotide sequence accession numbers. The nucleotide sequences of the *Paramecium DIE5a* and *DIE5b* genes are present in the GenBank database under accession numbers 124427424 and 124429605, respectively. The *Tetrahymena DIE5* gene is found on genomic scaffold scf_8254365 under GenBank accession number CH445530. The *Tetrahymena DIE5* gene is designated TTHERM_00686240, and the protein ID in GenBank is EAS04981.1. Preliminary *Paramecium* and *Tetrahymena* genome sequence data were obtained from Genoscope (http://www.genoscope.cns.fr/) and the J. Craig Venter Institute (formerly the Institute for Genomic Research; http://www.jcvi.org/), respectively (2, 13).

RESULTS

***Paramecium DIE5* encodes a novel, developmentally expressed nuclear protein.** As the genome remodeling that creates the somatic macronucleus is a major event in *Paramecium* development, we used differential display to identify proteins expressed exclusively during conjugation, as these are candidates that promise to be important for this nuclear differentiation (32). Upon further examination of the expression of individual candidate genes by Northern blot analysis, one in particular exhibited a dramatic increase in its mRNA abundance at 13 h into conjugation (Fig. 1A), and overall its expression closely corresponded with the known timing of IES excision (10 to 22 h, with a peak at 14 h) (6, 23). We also detected a low level of expression in starved cell populations, but this is likely derived from developing cells spontaneously undergoing autogamy (self-fertilization).

We named this promising candidate *D*efective *I*ES *E*xcision 5 (*DIE5*), due to its gene knockdown phenotype described below. *DIE5* is predicted to encode a 199-amino-acid (aa) protein with a molecular mass of 24 kDa. A nearly identical protein is encoded by a second locus in the *Paramecium* genome (http://www.genoscope.cns.fr/externe/GenomeBrowser/Paramecium/). This copy is clearly a *DIE5* paralog derived from the recent whole-genome duplication in the *Paramecium* lineage (2), as the two genes are remarkably similar (86% nucleotide identity; 98% amino acid identity) and both loci contain homologous copies of the neighboring gene, *NMD3*, present in the same orientation. Therefore, following the nomenclature convention for *Paramecium* genes (1), we designated the gene identified by differential display as *DIE5a* and its paralog as *DIE5b*. (We refer to both genes below simply as *DIE5*, given that 196 of 199 aa are conserved between them, and we used the *DIE5a* sequence in the design of all functional experiments described.) Analysis of the *Paramecium tetraurelia* Die5p amino acid sequence revealed no conserved protein domains or motifs except for two classical nuclear localization signals (NLSs); however, an identifiable homologue was found in the genome of the ciliate, *Tetrahymena thermophila* (described below). (When necessary to distinguish the *Paramecium* and *Tetrahymena* genes, we add the prefix *Pt* or *Tt* before *DIE5*). Thus, *DIE5* appears to be a novel gene that is conserved within the oligohymenophora lineage.

The two putative NLSs, along with the timing of expression, suggested to us that *DIE5* may encode a nuclear protein that participates in macronuclear differentiation. To further investigate this possibility, we examined Die5p localization by fusing GFP to its N terminus in a transgene expressed from the *DIE5* promoter. *Paramecium* cells containing this transgene showed no detectable GFP fluorescence during logarithmic growth, starvation, or the early stages of conjugation. However, consistent with *DIE5* mRNA expression (Fig. 1A), GFP-Die5p

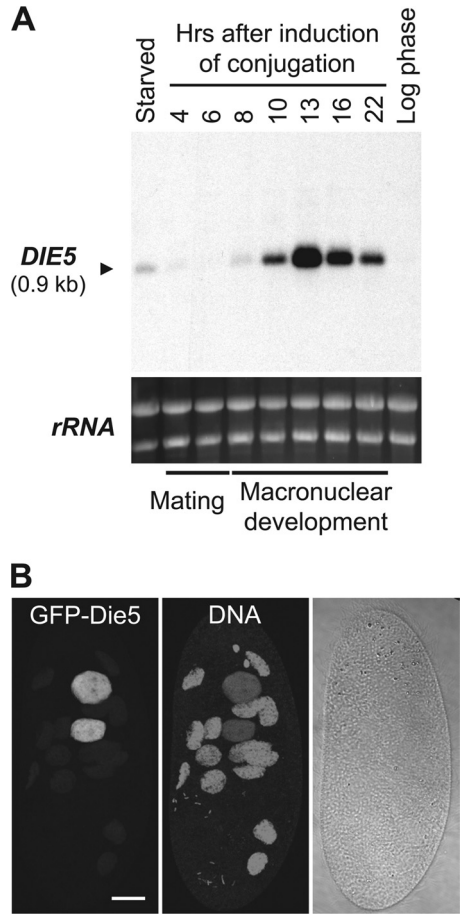

FIG. 1. *DIE5* is a developmentally regulated gene encoding a nuclear protein in *Paramecium*. (A) Northern blot of total RNA (20 μg per lane) from *Paramecium* probed with *Paramecium DIE5a*. Ethidium bromide staining of rRNA was used as a loading control. (B) Cellular localization of GFP-Die5a expressed from an extrachromosomal DNA driven by the *DIE5a* promoter. Fluorescent images are projections of optical sections obtained by confocal microscopy. The GFP fluorescence exclusively localized to the new macronuclei, which show weaker propidium iodide staining (DNA) than the old macronuclear fragments. The image corresponds to a cell approximately 14 h after the start of mating. The bar corresponds to 20 μm.

was observed in cells by 7 h after the initiation of mating. The fusion protein was found to localize exclusively within the developing macronuclei but not in the fragments of old macronuclei of conjugating or autogamous cells (Fig. 1B and data not shown). Thus, *PtDIE5* encodes a developmentally expressed protein that appears to act in differentiating macronuclei.

Silencing of *PtDIE5* inhibits formation of a functional macronucleus. To investigate whether the expression timing and localization of Die5p is indicative of a function in macronuclear development, we knocked down *DIE5* expression, using RNAi-mediated gene silencing, by feeding *Paramecium* cells *E. coli* expressing double-stranded RNA corresponding to a fragment of *DIE5* (16). To assess the level of knockdown achieved by this approach, we monitored protein expression of GFP-Die5p in transformed cells, using GFP-specific antibodies, which provided a proxy for endogenous expression. Western blot analysis showed as substantial reduction in GFP-Die5p levels, to ~35% of that observed upon feeding cells *E. coli* transformed with the empty RNAi vector (data not shown).

To determine whether *DIE5* is essential to complete development, we knocked down expression in mating cells that were genetically marked to allow true progeny to be distinguished from the parental lines. The two parental cell lines used in this experiment were each homozygous for a recessive allele at different loci (*pwA* or *pwB*; see Materials and Methods). Since successful conjugation of these cells generates F_1 progeny that are heterozygous at all loci, these F_1 exhibit a wild-type phenotype (genotype *PWA*/*pwA*, *PWB*/*pwB*). Both *DIE5* RNAi-treated and control cultures (i.e., cells fed bacteria containing the empty RNAi vector) produced high percentages of viable cells (78 and 97%, respectively), but only the control cells gave rise to true progeny that were phenotypically wild-type, whereas all viable cells of *DIE5* RNAi-treated cells exhibited the parental mutant phenotypes (Fig. 2A). Thus, loss of Die5p resulted in failure to form new macronuclei.

We visually followed the *DIE5* RNAi-treated conjugants throughout development to ascertain whether the lack of sexual progeny was caused by a failure in prezygotic events (e.g., meiosis, nuclear exchange, or karyogamy) or in events associated with postzygotic differentiation of the new macronucleus. Control matings that produced wild-type cells generated exconjugants with two macronuclear anlagen that eventually segregated to the daughter cells at the first postmating cell division (~18 h after induction of conjugation). In contrast, cytological observations of *DIE5* RNAi-treated cells (examined 30 and 36 h after induction of conjugation) showed a dramatic reduction in cells with differentiating macronuclei, as only 2 to 8% of exconjugants had two new macronuclei, while 12 to 23% had no macronuclear anlagen at all (Fig. 2B and C). The no-macronucleus cells are expected to include those that did not survive conjugation (lethal). The increase in no-macronucleus cells at 54 h may also include cells that entered autogamy shortly after conjugation. This is not possible in a normal mating but could occur in cells that have undergone macronuclear regeneration. Despite these nuclear abnormalities, most *DIE5* RNAi-treated conjugants proceeded through the first postconjugative cell division. These results are substantially different from those for silencing *UBA2* (which encodes a SUMO-activating enzyme), where most cells were arrested with two micronuclei and two macronuclei (32). Together with the genetic analysis of exconjugants, these results reveal that knockdown of *DIE5* expression blocks the completion of macronuclear development.

Despite the defects observed upon *DIE5* knockdown, these cells still exhibited a high level of viability upon exit from conjugation. This observation suggests that *DIE5* RNAi treatment likely induced the alternative developmental pathway of parental macronuclear regeneration (MR), which can occur in *Paramecium* when new macronuclei fail to form. This pathway has been observed upon silencing of other genes during conjugation (32, 41). Normally in wild-type cells, old macronuclear fragments remain in the exconjugants and are transcriptionally active for several postconjugative cell divisions. DNA replication no longer occurs, though, and these fragments are typically lost within 8 to 10 cell divisions (either actively or by

A

RNAi	Survival (%)	F_1 phenotype	
		Wild type	Mutant
Vector	97	33	2
DIE5	78	0	28

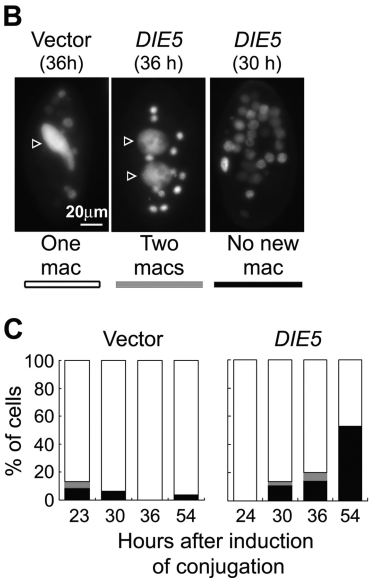

FIG. 2. Silencing of *DIE5* disrupts nuclear events during conjugation. (A) Progeny from conjugation treated with RNAi. Survival (%) and phenotype of the marker genes are shown. Both parental strains for conjugation were homozygous for Mendelian recessive mutant alleles of different marker genes (*pwA* and *pwB*; required for ciliary reversal); thus, successful conjugation should produce the wild-type phenotype of backward swimming, while failure in either nuclear exchange, fertilization, or formation of the new macronucleus should result in the mutant phenotype, i.e., no backward swimming upon stimulation. (B) Representative DAPI-stained *Paramecium* exconjugants with normal and defective cytological phenotypes observed after RNAi treatments. Each picture contains a single cell with or without the new macronucleus (arrowheads) and old macronuclear fragments. Phenotypic classes were assigned to white, gray, or black, as indicated below the pictures and plotted in the graphs in panel C. mac, macronucleus.

dilution). When the new macronucleus is incapable of division after sexual reproduction, one or more of the parental macronuclear fragments regenerate into a single macronucleus that is again capable of DNA replication, amitotic division, and transcription. After MR, the macronuclear genotype is the same as that of the parental lines (e.g., a mutant for *pwA* or *pwB*), while the micronuclear genotype is heterozygous, as nuclear exchange occurs between conjugates. To determine whether MR had occurred, several cells that survived *DIE5* knockdown during conjugation were followed into autogamy to reveal their micronuclear genotypes. The resulting F_2 lines of the *DIE5* RNAi F_1 survivors produced wild-type (as well as mutant) progeny, which demonstrates that the F_1 cell lines contained heterozygous micronuclei. Thus, the prezygotic events of conjugation (i.e., meiosis and nuclear exchange) must have occurred normally in the *DIE5* RNAi-treated cells, and the surviving F_1 were the result of MR.

***DIE5* is required for *Paramecium* IES excision.** The failure to form mature macronuclei after *DIE5* RNAi treatment despite successful formation of the zygotic nucleus raised the possibility that this protein is important for the removal of IESs. To examine the effect of *DIE5* knockdown on IES excision, we first employed a PCR-based approach to specifically amplify micronucleus-derived sequences. By locating one PCR primer for each amplicon within an IES and its partner in the flanking macronucleus-destined DNA, we could distinguish DNA in the macronuclear anlagen (and new micronuclei) from the abundant, rearranged DNA of the old macronuclear fragments. The abundance of the resulting PCR products (named "pp1," "pp2," and "pp3" in Fig. 3A) should increase as anlagen DNA is amplified and decrease as the IESs containing the primers are excised (Fig. 3B, vector lanes). In addition, as small IESs are contained within the PCR amplicons, we could follow their fates, as their excision prior to removal of the IESs containing the primer sites generated smaller products observed as faster-migrating bands in the gels (pp1s, pp2s, and pp3s in Fig. 3A and B). RNAi of *DIE5* showed a gradual increase in the amount of full-length (IES-containing) PCR product over the developmental time course, indicating that loss of *DIE5* does not inhibit developmental DNA amplification (Fig. 3B). Nevertheless, we saw no evidence of shorter products, which are readily detectable in control cells, that would indicate IES excision had occurred.

To further demonstrate that loss of *DIE5* blocked IES excision, we used Southern blot hybridization to analyze the rearrangement status of the new macronuclei. For this study, total DNA was isolated from large cultures of postautogamous cells at a point when about 50% of well-fed control cells (treated with the empty RNAi vector) had undergone the first cell division. The DNA was digested with SspI, which has frequent recognition sites in IESs but only one in the coding region of the *A-51* allele. The probe containing macronuclear DNA from the A gene (HincII-PstI in Fig. 3A) detected only the macronuclear form of this region (indicated by the 5.7-kbp band in Fig. 3C) in DNA from control cells (Fig. 3C, labeled vector). In contrast, two additional bands of 1.3 and 2.0 kb were observed in DNA from *DIE5* RNAi-treated cells. These are the sizes expected for amplification of the unprocessed A-51 gene (Fig. 3C). Furthermore, using a probe for IES4578 (Fig. 3A) that contained only micronucleus-limited DNA, we detected a single 500-bp DNA fragment, which corresponds to the IES-containing locus in DNA from *DIE5* RNAi-treated cells but not from control cells (Fig. 3C). The unprocessed DNA can be observed in these experiments because amplification of DNA continues in the developing macronuclei despite the absence of IES excision. Together the results show that knockdown of *DIE5* expression inhibits IES excision in the developing macronuclei of *Paramecium*, yet the DNA is amplified to levels comparable to that for normal developing macronuclei.

***Tetrahymena DIE5* is required for macronuclear differentiation.** The PtDie5p sequence contained no conserved domains that offered clues to its biochemical function; however, we did find a putative *DIE5* homologue (*TtDIE5*) encoded within the *Tetrahymena thermophila* genome. These two ciliate proteins are similar in size (199 and 207 aa, respectively) and share 21%

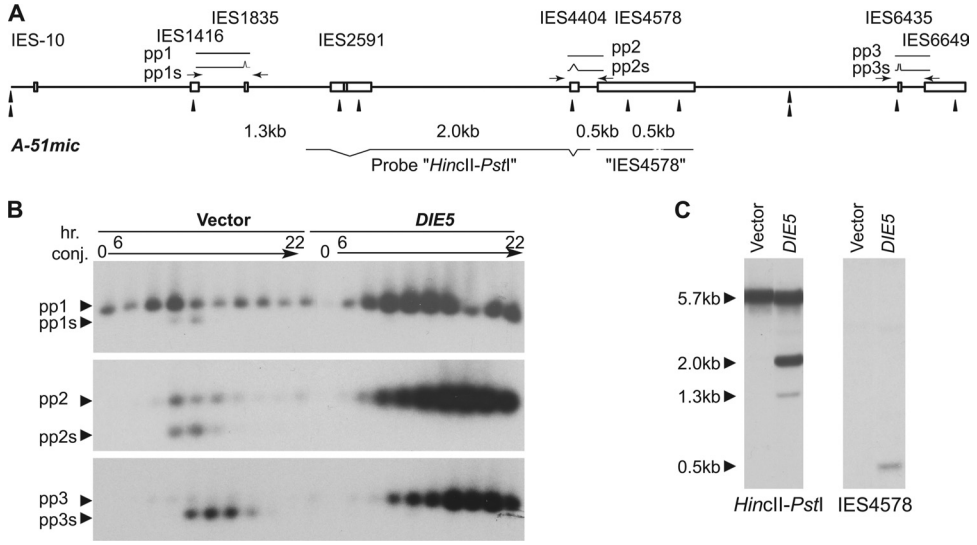

FIG. 3. Silencing of *DIE5* inhibits excision of IESs. (A) A partial map of the micronuclear version of the *A-51* allele showing locations of IESs (boxes), SspI recognition sites (arrowheads), and sizes of fragments generated by SspI digestion. Positions of expected PCR products (pp1 to pp3) and probes for Southern hybridization are also indicated. Arrows show positions of the primers relative to IESs not drawn to scale. (B) Whole-cell semiquantitative PCRs of RNAi-treated exconjugants, using one primer in the macronuclear sequence and the other primer inside the IESs. Each lane represents whole-cell PCR products taken at 2-h intervals from 6 to 22 h after induction of conjugation. The predicted PCR products correspond to pp1 to pp3 in panel A. Due to excision of smaller IESs during rearrangement, two bands are expected for each primer set. (C) SspI-digested total genomic Southern blots of RNAi-treated exautogamous cells probed with either a HincII-PstI fragment or IES4578, as indicated in panel A. Total DNA (~10 μg) was isolated from an exautogamous cell culture when about 50% of control cells (RNAi using empty vector) had undergone the first cell division. Most IESs in the micronuclear version of the *A-51* allele contain SspI sites (single arrow heads in panel A), while only one site is present in the macronuclear-destined sequence of the *A-51* allele (the double arrow head in panel A). Thus, for probe HincII-PstI, 1.3- and 2.0-kb fragments are expected for unprocessed DNA, while a 5.7-kb fragment is expected for the processed DNA, including abundant old macronuclear DNA in exconjugants. Probe IES4578 contains only the IES sequence and detects 0.5-kb fragments if the IES is present at high levels in exautogamous cells.

amino acid identity throughout their coding regions (Fig. 4A). In addition, when the sequences were analyzed for predicted secondary structures using PSIPRED (33), a common domain structure that predicted beta sheets and alpha helices in the form β1α1α2β3β4 was revealed within the first 110 aa of both proteins (data not shown). Intriguingly, our Northern blot analysis revealed that *TtDIE5* is expressed exclusively during conjugation, providing data to support that this *Tetrahymena* gene may have a function similar to that of its *Paramecium* counterpart. *TtDIE5* expression was first detected 4 h into conjugation, which corresponds to the end of meiosis, and peaked at 6 h, when new macronuclei first emerge. Expression levels declined slightly at 8 h but continued at a low level until the end of macronuclear development (about 15 h after cells first paired) (Fig. 4B). The expression pattern we observed is nearly identical to recently published microarray data (35).

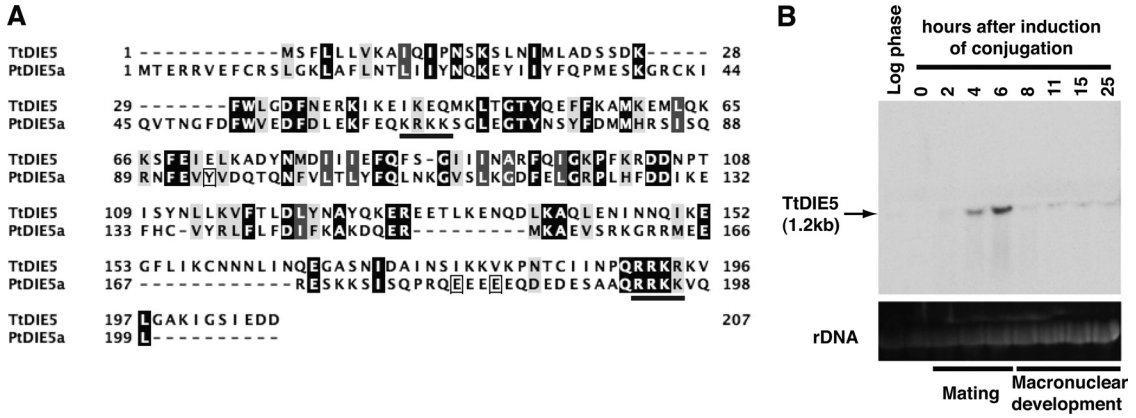

FIG. 4. *Tetrahymena* homologue (TtDIE5) is developmentally expressed. (A) Pairwise alignment of *Paramecium* Die5 protein from paralog a (PtDIE5a) and *Tetrahymena* Die5p sequences. The boxes around three amino acids indicate differences between the two *Paramecium* sequences (the differences are C, D, D, respectively). The underlined regions indicate potential nuclear localization signals in the *Paramecium* sequence. (B) Northern blot analysis of total RNA (20 μg per lane) extracted from growing (log), starved, or conjugating cells (between 2 and 25 h after mixing populations of compatible mating types). Ethidium bromide staining of rRNA was used as a loading control.

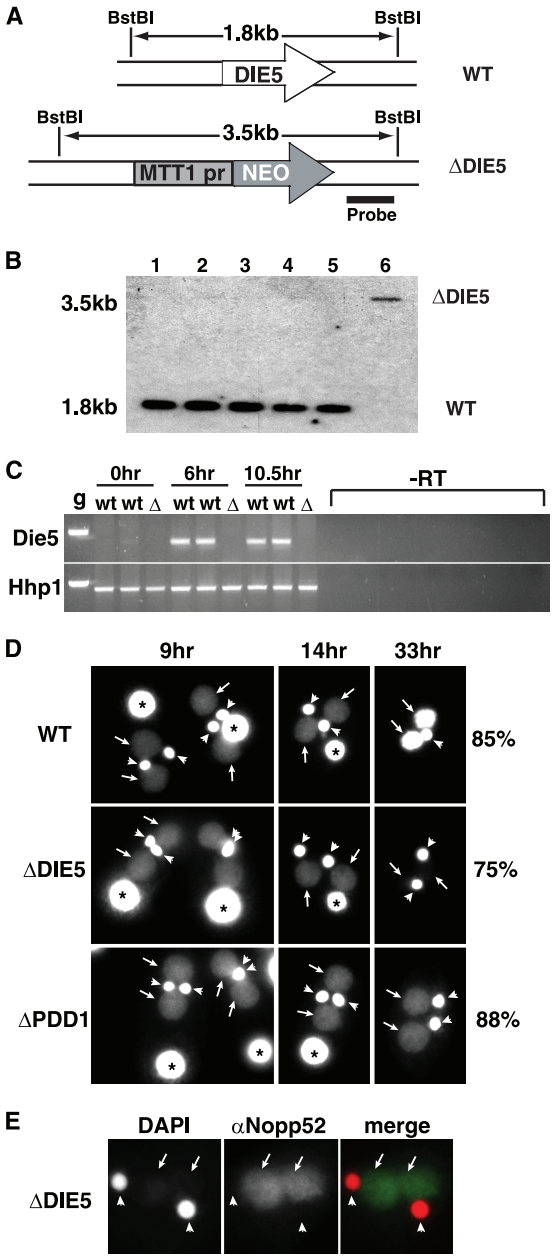

FIG. 5. *TtDIE5* zygotic expression is essential to complete conjugation. (A) Diagram of the gene disruption construct showing replacement of the coding sequence with the *neo3* selectable cassette (47). Restriction enzyme sites (BstBI) and the region corresponding to the radiolabeled probe fragment used for Southern blot analysis are indicated. WT, wild type; MTT1pr, metallothionein gene 1 promoter. (B) Southern blot hybridization of DNA isolated from the wild type (lane 1), *DIE5* micronuclear-knockout lines (lanes 2 to 5), and the *DIE5* complete knockout line (lane 6) using the probe shown in panel A. Longer exposure of the blot reveals the 3.5-kb band in *DIE5* micronuclear-knockout samples corresponding to the two copies of *DIE5* disrupted by the *neo3* cassette in their micronuclei. (C) RT-PCR for expression of *DIE5* transcripts in conjugating wild-type (wt) and *DIE5* complete knockout (Δ) cells at indicated time points. The bottom panel shows control RT-PCR with *HHP1* primers. Both *DIE5* and *HHP1* primers span an intron of their respective genes. g, *Tetrahymena* genomic DNA used as a control for amplification. (D) Fluorescent images of representative DAPI-stained wild-type (WT), micronuclear-knockout (Δ*DIE5*), and Δ*PDD1* strains at 9 h, 14 h, and 33 h postmixing. For each mating strain, the percentage of cells exhibiting their respective arrest phenotype at 33 h is indicated on the right. Asterisks, arrows, and arrowheads indicate old/parental macronuclei, new macronuclei, and the micronuclei, respectively. (E) The nuclear envelope remains intact in arrested Δ*DIE5* micronuclear-knockout cells (ΔDIE5). Postconjugative Δ*DIE5*mic cells were fixed with 2% paraformaldehyde and stained with αNopp52 antibody and DAPI. New macronuclei and micronuclei are indicated as described for panel D.

Our RNAi knockdown experiments showed that Die5p is essential for macronuclear differentiation in *Paramecium*. If *TtDIE5* is a homologue of the *Paramecium* protein, it is likely to have an essential role in *Tetrahymena* development as well. To investigate this possibility and learn more about Die5p function, we used homologous gene replacement to disrupt *TtDIE5*. A knockout construct (Fig. 5A) consisting of the *neo3* selectable cassette (45) flanked on each side by DNA sequences from immediately upstream and downstream of the *TtDIE5* locus was introduced by biolistic transformation into starved or conjugating *Tetrahymena* cells to disrupt the macronuclear or micronuclear copies, respectively. Paromomycin-resistant transformants were obtained by both strategies and were subsequently cultured to generate full macronuclear knockouts or homozygous micronuclear-knockout lines (see below and Materials and Methods). Genetic crosses of strains lacking all macronuclear *DIE5* copies produced viable progeny (data not shown), indicating that *DIE5* expression prior to transcriptional activation of developing macronuclei is not required to complete conjugation. In contrast, attempts to generate complete knockout strains (lacking *DIE5* in both micro- and macronuclei) by crossing heterozygous germ line (micronuclear) knockout strains were unsuccessful. One-quarter of the progeny resulting from these crosses should have been homozygous knockouts (in both nuclei), yet only homozygous wild type or heterozygous knockout strains were found among the >30 viable progeny screened (data not shown). This finding provided the first indication that TtDie5p is critical for development, as is the *Paramecium* protein.

As strains lacking all macronuclear *DIE5* copies grew normally and produced viable progeny when mated, the inability of heterozygous *DIE5* micronuclear knockouts to produce homozygous knockout progeny must result from the loss of zygotic *DIE5* expression during macronuclear differentiation. To allow us to further investigate Die5p's role during *Tetrahymena* development, we generated homozygous micronuclear-knockout heterokaryon strains of different mating types by performing genomic exclusion crosses between heterozygous micronuclear knockouts and two different star strains, B*VI and B*VII, which have defective micronuclei. These abortive matings resulted in the transfer of a haploid micronucleus from the *DIE5* knockout to its star strain partner without inducing new macronuclear development, such that after pair separation and micronuclear endoreplication, both exconjugants were homozygous in their germ line. The resulting paromomycin-sensitive exconjugants (the star strain partner) with homozygous knockout micronuclei and wild-type macronuclei were identified by genomic locus PCR (data not shown) and Southern blot analysis (Fig. 5B) and then verified by genetic crosses (data not shown). Herein we refer to these cell lines as *DIE5* micro-

nuclear knockouts or Δ*DIE5*mic (these cells were used for most experiments; therefore, all figures labeled ΔDIE5 refer to micronuclear knockouts). Phenotypic assortment of the paromomycin-resistant, homozygous, micronuclear-knockout exconjugants allowed us to generate complete (micro-/macronuclear) knockout strains (referred to herein as *DIE5* complete knockouts). We confirmed the loss of all *DIE5* copies and expression during conjugation by Southern blot analyses and RT-PCR (Fig. 5B and C), respectively. The ability to generate *DIE5* complete knockouts confirms that Die5p is dispensable for vegetative growth.

Δ*DIE5*mic strains are paromomycin sensitive but are homozygous for the *DIE5*::*NEO3* allele in their silent micronuclei, so that when mated, their progeny, if viable, would be paromomycin-resistant. To measure the ability of these germ line knockout strains to produce Δ*DIE5* progeny, $>10^6$ postconjugative cells were cultured in growth medium containing paromomycin (plus $CdCl_2$). In most trials, all cells died upon the addition of the drug, providing further support that zygotic expression of *DIE5* is essential for the completion of conjugation. We did obtain paromomycin-resistant cells for some matings (up to two survivors per 10^6 mating pairs). We interpret this to mean that some cells express sufficient Die5p from their wild-type parental macronuclei to provide for the essential functions of this protein late in macronuclear differentiation. To further test this possibility, we created an N-terminal GFP-*TtDie5p* transgene expressed from the strong, cadmium-inducible *MTT1* promoter, integrating the construct upstream of the macronuclear *rpL29* genomic locus in the Δ*DIE5*mic strains. Expression of this transgene by the addition of cadmium during conjugation increased progeny survival by 3 to 4 orders of magnitude. Thus, inducing additional Die5p from the parental macronucleus partially rescued the loss of *DIE5* zygotic expression from macronuclear anlagen, which provided clear evidence that loss of *DIE5* expression is the cause of lethality upon mating Δ*DIE5*mic cells and that the GFP-TtDie5p fusion is functional. We were also able to increase survival of Δ*DIE5*mic conjugates by introducing an rDNA-based expression vector by electroporation that carries the GFP-*TtDIE5* transgene at ~9 h into conjugation, an observation which further supports that this protein is required relatively late in macronuclear development.

Tetrahymena conjugation can be readily staged by the configuration of nuclei within mating pairs or exconjugants (31). To begin to ascertain why mating Δ*DIE5*mic cells fail to produce viable progeny, we fixed and DAPI-stained cells to monitor their development. For most of conjugation, the progression of Δ*DIE5*mic conjugates was indistinguishable from that of wild-type mating cells as they completed meiosis, karyogamy of gametic nuclei, and the subsequent nuclear divisions to generate macronuclear anlagen (Fig. 5D). Mating pairs separated, and their nuclear morphology appeared normal until ~14 h of conjugation, after which it became evident that most knockout cells arrested their development. The last visible event to be triggered during conjugation is the elimination of one of the two micronuclei in each conjugant. Wild-type cells remain with the characteristic nuclear configuration of one micronucleus and two new macronuclei (Fig. 5D, 33 h) until they are fed, at which point they divide their one remaining micronucleus and undergo a specialized postconjugative cytokinesis to partition one micro- and one macronucleus to each daughter. In contrast, the Δ*DIE5*mic conjugates failed to trigger micronuclear resorption, arresting with two micro- and macronuclei, and were unable to divide when returned to growth medium.

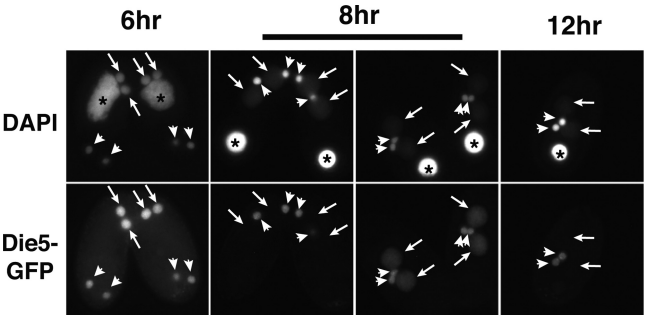

FIG. 6. Cellular localization of GFP-Die5p in conjugating *Tetrahymena* cells at 6 h, 8 h, and 12 h. The cells were fixed with 2% paraformaldehyde and counterstained with DAPI. Asterisks, arrows, and arrowheads indicate old/parental macronuclei, new macronuclei, and the micronuclei, respectively.

The two micro-/two macronuclei arrest phenotype has been described for knockouts of several genes required for programmed DNA rearrangement, including Δ*PDD1* strains (Fig. 5D, bottom panels). Such mutant strains not only fail to eliminate one micronucleus, but unlike wild-type cells, stop amplification of the DNA in the new macronuclei (Fig. 5D, compare the DAPI-staining intensity of Δ*PDD1* cells to that of the wild type at 33 h). Intriguingly, not only did Δ*DIE5*mic exconjugates stop anlagen DNA amplification, but the majority of cells (75%) actually lost most of the DNA content of the anlagen, as indicated by the loss of DAPI staining between 14 and 33 h postmixing (Fig. 5D). The macronuclear structure appeared to remain intact, as immunofluorescence to detect a nucleolar protein, NOPP52, revealed the integrity of the nuclear compartment (Fig. 5E). We are unsure of the mechanism of DNA loss, as attempts to detect chromosome degradation by the presence of unprotected ends revealed that unlike the degrading old macronucleus, these new macronuclei do not label in terminal deoxynucleotidyltransferase-mediated dUTP-biotin nick end labeling (TUNEL) assays (data not shown). Nevertheless, the loss of macronuclei in the Δ*DIE5*mic exconjugates is consistent with the *Paramecium DIE5* RNAi phenotype in which anlagen form but fail to become the macronuclei of the surviving exconjugants.

As the *DIE5* expression profile and knockdown/knockout phenotypes were very similar in both *Paramecium* and *Tetrahymena*, we expected that *TtDIE5* would also encode a protein localized primarily to macronuclear anlagen. However, we were surprised when we examined the localization of the GFP-TtDie5p fusion used in the rescue experiments described above. Although GFP-TtDie5p was detected in the developing macronuclei during early stages of macronuclear development (Fig. 6, 6-h to 8-h arrows), this localization was rapidly lost as conjugation progressed, before the period when the essential zygotic expression would occur. In contrast, we observed the GFP fluorescence within micronuclei through all stages of conjugation (Fig. 6, arrowheads). Therefore, while Die5p is a

developmentally expressed nuclear protein in both ciliates, their localization patterns suggest some differentiation in their action.

Zygotic expression of *DIE5* is required for programmed DNA rearrangement in *Tetrahymena*. Transcription of *TtDIE5* was induced by 4 h into conjugation (Fig. 4A) and peaked at 6 h; thus, it was somewhat unexpected that we found that the macronuclear copies were dispensable. Additionally, complete-knockout cells lacking *DIE5* from both macro- and micronuclei exhibit the same phenotype as the germ line knockouts, which further confirms that the critical *TtDIE5* expression occurs from macronuclear anlagen. This suggests that *DIE5* functions at late stages of *Tetrahymena* conjugation when DNA rearrangement occurs. To determine whether zygotic Die5p is required for *Tetrahymena* DNA rearrangements, we monitored IES excision and chromosome breakage in Δ*DIE5*mic exconjugants. To assess the process of IES excision, we examined the elimination of the M IES, which is a >1-kbp sequence located on micronuclear chromosome 4 (3, 9). This IES has two equally used left deletion boundaries that are 300-bp apart, such that successful IES excision generates two alternative products through the elimination (Δ) of either 0.6 kbp or 0.9 kbp from the locus (Fig. 7A). These two forms can be easily distinguished by a PCR-based IES excision assay, using primers designed to amplify across the IES. By crossing Δ*DIE5*mic lines together or with wild-type strain B2086, each of which contains only the MΔ0.9-kbp rearranged form in their macronuclei, we could test for appearance of the MΔ0.6-kbp deletion as an indicator of new rearrangement in the anlagen. Whereas the slower-migrating PCR product indicative of the MΔ0.6-kbp deletion was detected when each Δ*DIE5*mic line was crossed to B2086, we saw no evidence of M-element rearrangement in crosses of the two Δ*DIE5*mic strains (Fig. 7B).

Since RNAi knockdown of *Paramecium DIE5* blocked chromosome fragmentation (data not shown) as well as IES excision (Fig. 3), we tested whether Δ*TtDIE5*mic knockouts also fail to fragment chromosomes. We isolated DNA from postconjugative Δ*DIE5*mic cells and used Southern blot analysis to examine chromosome fragmentation at a site which lies 2.2 kbp downstream of the *LIA1* gene (Fig. 7C) (30). During conjugation, breakage occurs at this sequence, followed by the addition of 300 bp to 400 bp of telomeric DNA during growth. Restriction digestion of genomic DNA with EcoRI allows simultaneous detection of the unprocessed micronuclear form of this locus at 10.5 kbp, the fragmented chromosome of parental macronuclei with fully elongated telomeres that migrates at 2.5 to 2.6 kbp, and the 2.2-kbp form that results from *de novo* breakage and minimal telomere addition (Fig. 7C). Whereas this 2.2-kbp fragment was easily observed in genomic DNA samples from wild-type matings, the DNA from Δ*DIE5*mic conjugants showed no evidence of chromosome fragmentation, as the predominant band observed was ~2.5 kbp, which represents DNA from parental macronuclei of unmated cells in the population (Fig. 7D). These results together with experiments with *Paramecium* (data not shown) indicate that *DIE5* is required for IES excision and chromosome fragmentation in both ciliates.

Critical events leading to DNA rearrangement are unaffected in Δ*DIE5*. IES excision is guided by small RNAs. In *Tetrahymena*, it is known that these small RNAs target H3K9 and K27 methylation to IES chromatin, which is bound by the Pdd1 and Pdd3 chromodomain-containing proteins that are reorganized into distinct foci that are hypothesized to be the site of IES excision (reviewed in reference 10). The timing of

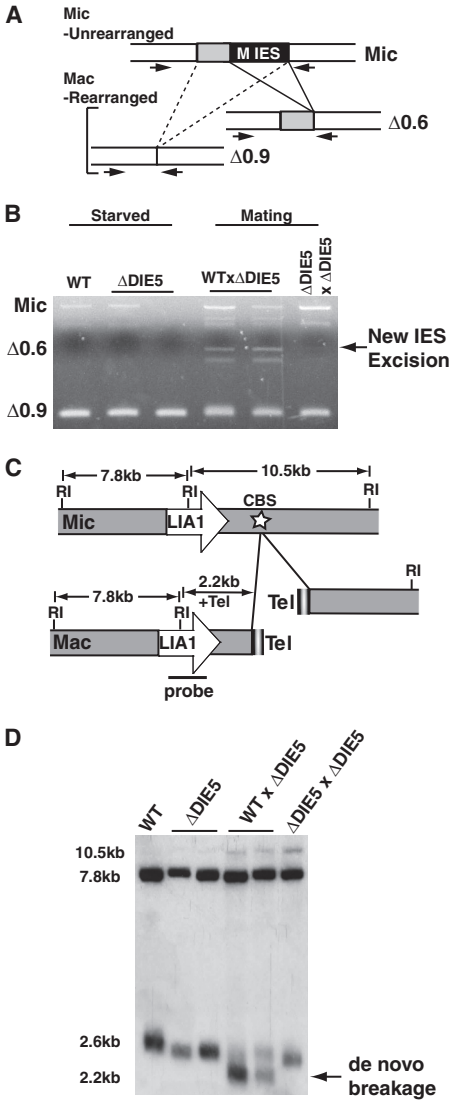

FIG. 7. Germ line *TtDIE5* is required for *Tetrahymena*-programmed DNA rearrangement. Total genomic DNA was isolated from starved wild-type (WT) and *DIE5* micronuclear-knockout strains (Δ*DIE5*), as well as from cells 16 h after crosses of *DIE5* micronuclear knockouts to the wild type (WT × ΔDIE5) and two *DIE5* micronuclear knockouts (Δ*DIE5* × Δ*DIE5*). The DNA was used for PCR-based IES excision assays (A and B) or Southern blot analyses for chromosome breakage (C and D). (A) Schematic of PCR-based IES excision assay strategy. Arrows denote forward and reverse primers used to amplify across the M element. Alternative rearrangement products resulting from deletion of 0.6-kbp (Δ0.6) or 0.9-kbp (Δ0.9) are shown. (B) M-element excision PCR. Arrow indicates new IES excision. (C) Diagram shows the macronuclear chromosomal scaffold surrounding the *LIA1* gene, which lies within 2.2 kbp of a chromosomal-breakage sequence (CBS) (white star). Relevant EcoRI (RI) restriction sites used for the Southern blot analysis are shown. The probe spans the central EcoRI site and detects a 7.8-kbp fragment common to both nuclei as well as to either the 10.5-kbp micronucleus-specific fragment or a 2.5- to 2.6-kbp macronucleus-specific fragment (2.2 kbp of unique sequence plus 300 to 400 bp of telomeric DNA). Tel, telomere. (D) Southern blot analysis to assess chromosome breakage. Arrow indicates the product of *de novo* breakage.

these events has been well described, and proteins required for many of these steps have been identified (11, 26, 36, 50). To investigate the role of *DIE5* in *Tetrahymena* DNA rearrangement, we examined these events to determine whether they are perturbed by the loss of *DIE5*.

Production of developmental-specific small RNAs occurs during meiosis, before the emergence of zygotic expression. As zygotic, not somatic, expression of *DIE5* is essential for conjugation in *Tetrahymena*, we expected that *DIE5* would not be required for their biogenesis or accumulation. We isolated small RNAs from conjugating complete-*DIE5*-knockout cells, and as predicted, we found that the levels of total small RNAs and those homologous to the M IES observed were comparable in wild-type and Δ*DIE5*mic conjugating cells (data not shown).

We also examined whether H3K9 and H3K27 methylation was perturbed by the loss of *DIE5*. Although establishment of these marks occurs during the time that we observed GFP-TtDie5p to localize to developing macronuclei, deposition of these modifications in conjugating *DIE5* knockout cells appeared to be unaffected (Fig. 8A). Therefore, Die5p is not required for the establishment of these heterochromatic modifications in *Tetrahymena*, further suggesting that the essential function of *DIE5* occurs at later stages of conjugation.

The chromodomain-containing protein Pdd1p is an essential component of the DNA rearrangement machinery and a major constituent of DNA rearrangement foci (12, 27). Disruption of genes required for IES excision can block the formation of these foci (26, 44). To investigate chromatin reorganization upon loss of Die5p, we examined the localization of a Pdd1p-YFP fusion expressed in Δ*DIE5*mic cells. Even though these mutant lines fail to excise IESs, Pdd1p foci in the developing macronuclei appear to form, as we observed in wild-type conjugants, albeit with a slight delay in their maturation (Fig. 8B). Given that all events prior to IES excision described to date appear normal in Δ*DIE5* knockouts, we suspect that Die5p acts after these known events.

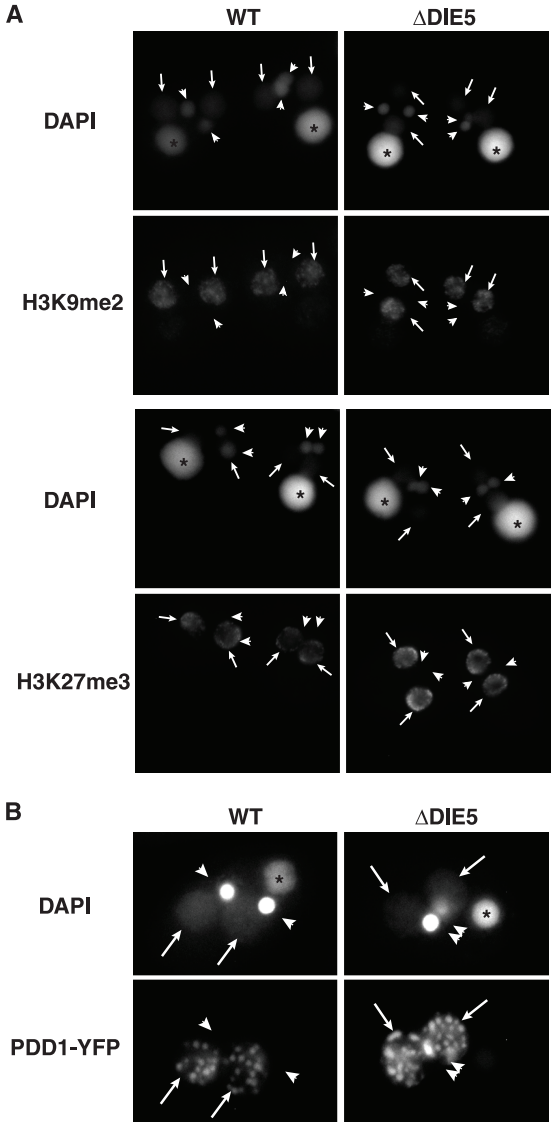

FIG. 8. Germ line knockout of *TtDIE5* does not inhibit critical events leading to IES excision. (A) H3K27me3 and H3K9me2 histone mark deposition is unaffected in *TtDIE5* germ line knockouts. Nine-hour-conjugating wild type (WT) and micronuclear-knockout (Δ*DIE5*) cells were fixed with Schaudinn's fixative and stained with either H3K27me3 or H3K9me2 antibodies at 9 h into conjugation. The cells were counter stained with DAPI. (B) Formation of PDD1 foci is not disrupted in Δ*DIE5mic* cells. Conjugating wild-type and *DIE5* micronuclear-knockout cells (Δ*DIE5*) expressing an inducible *PDD1*-YFP transgene were fixed at 14 h with 2% paraformaldehyde and counter stained with DAPI. Asterisks, arrows, and arrowheads indicate old/parental macronuclei, new macronuclei, and the micronuclei, respectively.

DISCUSSION

Orthologous *DIE5* genes in *Paramecium* and *Tetrahymena* are essential for macronuclear development. Research over the past 10 years has revealed that some key components of the genome reorganization pathway in ciliates are well-known proteins in other eukaryotic organisms. Examples include proteins of the RNA interference pathway, histone-modifying enzymes, and transposases (4, 26, 30, 36). Other studies have identified proteins with recognizable domains, and yet clear orthologs cannot be identified, even in other ciliate genomes (58). *DIE5* is a member of a third group, proteins that are both conserved among ciliates (*Paramecium* and *Tetrahymena*) but novel to this class of organisms. This group may contain core components that account for the unusual precision and efficiency of ciliate genome reorganization. Our results provide strong evidence that *Paramecium* and *Tetrahymena* encode orthologous Die5 proteins. Although the amino acid identity between the proteins is modest (21%), each is the top reciprocal BLAST hit in comparisons of their respective genomes. Additionally, the amino acid identity is not centered in a single region, as expected for a shared domain; rather, it is spread across the entire coding region. Both are small, nuclear proteins (199 aa and 207 aa in *Paramecium* and *Tetrahymena*, respectively) expressed exclusively during conjugation. Finally, RNAi knockdown (*Paramecium*) and gene disruption (*Tetrahymena*) demonstrate that these Die5 proteins are essential for genome rearrangements in their respective species. To-

gether the data demonstrate that *DIE5* encodes a conserved component of the macronuclear development pathway in ciliates.

Die5 is one of a few conserved ciliate proteins known to be required for DNA rearrangements. Previous studies have identified several developmentally regulated proteins that are required for genome rearrangements in *Paramecium* or *Tetrahymena*. These include chromatin-associated proteins (27, 28, 39, 40, 48, 58), a transposase (4), components of the RNAi pathway (30, 36, 37) and SUMO pathways (32), and a putative RNA binding protein (41). Of these identified components, only the RNAi-associated proteins, piggyMac transposase, and SUMO components show clear homologues between these different ciliates. In contrast, the Pdd and Lia (*l*ocalized *i*n macronuclear *a*nlagen) DNA rearrangement proteins of *Tetrahymena* do not have obvious homologues in *Paramecium* (58). This search for homologues is complicated by the significant sequence diversity between *Paramecium* and *Tetrahymena*, a point that was evident even before complete-genome sequences were available (reviewed in reference 15). Thus, failure to detect homologues based on primary sequence data is not definitive, and one must consider the possibility that protein three-dimensional structure is maintained with minimal sequence identity. Functional homology is the most relevant criteria, but few of the proteins involved in these rearrangements have known biochemical roles. For those that do, such as the *Tetrahymena* chromodomain-containing Pdd1 and Pdd3 proteins that have been shown to associate with IES chromatin-containing histone H3 methylated on lysine 27 and/or lysine 9, respectively, evidence for an analogous role in *Paramecium* DNA rearrangements is lacking (26, 50). The Nowa proteins in *Paramecium* are required for elimination of germ line transposons and a subset of IES that are controlled by maternal effects (41). The N-terminal domains contain repeated elements with similarity to RNA binding motifs in other species, and evidence of nucleic acid binding activity was obtained. While there are *Tetrahymena* proteins, including CnjBp (41, 51), that share similar glycine-rich repeats, none are clearly identifiable as homologues. Whereas the Pdd, Lia, and Nowa proteins reinforce the divergence between the DNA rearrangement machinery in the two species, Die5p reveals a novel connection between the two systems.

***DIE5* is required late in macronuclear development.** Although the novel sequence of Die5p limits speculation on its biochemical function, the molecular events that are disrupted (or not disrupted) by inhibiting *DIE5* expression in two species argue for late action in macronuclear development. First of all, prezygotic events of meiosis and pronuclear exchange occur normally upon *DIE5* RNAi silencing in *Paramecium*. The survivors of conjugation upon *DIE5* knockdown are heterokaryons with heterozygous micronuclei (with alleles from both parents), but they have macronuclei regenerated from fragments of their parental macronuclei. This phenotype has been observed upon knockdown of other genes involved in macronuclear development (32, 41). Likewise, early-stage events of scnRNA biogenesis (data not shown) and accumulation of specific chromatin modifications (Fig. 8) associated with DNA elimination appear unaltered upon disruption of *DIE5* in *Tetrahymena.* This is in contrast to disruption of RNAi components (e.g., *DCL1* or *TWI1*) that are defective in these events.

Even relatively late-stage events prior to IES excision were not disrupted upon loss of *DIE5* expression. Macronuclear anlagen form and substantial DNA amplification occurs in RNAi-treated *Paramecium* even though IES excision is inhibited. In *Tetrahymena* Δ*DIE5*mic conjugants, Pdd1p foci formed as expected for wild-type cells (Fig. 8B). These observations, together with the fact that we saw no phenotype when *DIE5* was disrupted only from the parental macronucleus, suggest that its essential role occurs after activation of zygotic expression of the macronuclear anlagen. Most striking is our observation that *Tetrahymena* heterokaryons with wild-type macronuclei but homozygous for the Δ*DIE5* allele in their micronuclei fail to excise IESs and arrest prior to the elimination of one of the two micronuclei. These results clearly demonstrate a late role for Die5p in macronuclear differentiation.

While loss of *DIE5* expression blocked IES excision in both ciliates studied here, we did observe some differences in terminal phenotypes, in particular the degree of anlagen DNA amplification. *DIE5* RNAi-treated *Paramecium* appeared to extensively amplify their anlagen DNA, whereas Δ*DIE5*mic *Tetrahymena* cells arrested their anlagen differentiation at a low amplification level. We favor the explanation that this most likely represents distinct differences in nuclear differentiation events (e.g., macronuclear regeneration occurs in *Paramecium* but not in *Tetrahymena*) in these evolutionarily diverse organisms. Nevertheless, we cannot rule out that such phenotypic differences may result from low levels of PtDie5p remaining after RNAi silencing. What is clear is that RNAi silencing that was robust enough to prevent IES excision in *Paramecium* to the same extent as was observed in *Tetrahymena* Δ*DIE5*mic conjugants elicited differential effects on anlagen DNA amplification.

The surprising observation that anlagen DNA is degraded in *Tetrahymena* Δ*DIE5* conjugants has not been observed in other DNA rearrangement mutants (12, 30, 36, 40). The mechanism of degradation is not understood, and we were unable to demonstrate labeling in TUNEL assays. It is interesting to note that anlagen DNA degradation appears to be the outcome in *Paramecium DIE5* RNAi silencing, but as mentioned previously, these cells survive by regenerating a functional macronucleus from an old macronuclear fragment. This is possible in *Paramecium* because macronuclear fragments persist throughout conjugation and continue transcription for multiple cell divisions after conjugation. In contrast, degradation of the old macronucleus in *Tetrahymena* is complete prior to formation of the new genome and leaves no recovery pathway if development is defective. Further analysis will be required to establish whether the degradation of *Tetrahymena* anlagen DNA has a similar molecular basis as the fate of *Paramecium* anlagen in *DIE5*-silenced cells.

The significance of TtDie5p localization to the micronuclei but not anlagen at late time points is unclear. This may reflect the limitation of using a macronuclear GFP-*DIE5* transgene that cannot be expressed after degradation of the old macronucleus. Alternatively, the localization could be evidence of an unexpected role of Die5p from the newly formed micronucleus. Unusual observations such as these underscore the importance of investigating the role of conserved proteins such as Die5p in the DNA rearrangement process to elucidate the evolution and regulation of this massive genome-remodeling

process and its connections to large-scale genome reorganizations in other species.

ACKNOWLEDGMENTS

We thank Charles H. Rexer for initial work toward characterizing *TtDIE5*. We also thank Genoscope and the J. Craig Venter Institute for preliminary *Paramecium* and *Tetrahymena* sequence data, respectively.

A.M. and J.D.F. were supported by a National Science Foundation grant (MCB-9506009). D.L.C. and A.W.-Y.S. were supported by National Institutes of Health research grant GM06593.

REFERENCES

1. **Allen, S. L., M. I. Altschuler, P. J. Bruns, J. Cohen, F. P. Doerder, J. Gaertig, M. Gorovsky, E. Orias, and A. Turkewitz.** 1998. Proposed genetic nomenclature rules for *Tetrahymena thermophila*, *Paramecium primaurelia* and *Paramecium tetraurelia* on behalf of the Seventh International Meeting on Ciliate Molecular Biology Genetics Nomenclature. Genetics **149:**459–462.
2. **Aury, J. M., O. Jaillon, L. Duret, B. Noel, C. Jubin, B. M. Porcel, B. Segurens, V. Daubin, V. Anthouard, N. Aiach, O. Arnaiz, A. Billaut, J. Beisson, I. Blanc, K. Bouhouche, F. Camara, S. Duharcourt, R. Guigo, D. Gogendeau, M. Katinka, A. M. Keller, R. Kissmehl, C. Klotz, F. Koll, A. Le Mouel, G. Lepere, S. Malinsky, M. Nowacki, J. K. Nowak, H. Plattner, J. Poulain, F. Ruiz, V. Serrano, M. Zagulski, P. Dessen, M. Betermier, J. Weissenbach, C. Scarpelli, V. Schachter, L. Sperling, E. Meyer, J. Cohen, and P. Wincker.** 2006. Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature **444:**171–178.
3. **Austerberry, C. F., C. D. Allis, and M. C. Yao.** 1984. Specific DNA rearrangements in synchronously developing nuclei of Tetrahymena. Proc. Natl. Acad. Sci. U. S. A. **81:**7383–7387.
4. **Baudry, C., S. Malinsky, M. Restituito, A. Kapusta, S. Rosa, E. Meyer, and M. Betermier.** 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev. **23:**2478–2483.
5. **Bétermier, M.** 2004. Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate Paramecium. Res. Microbiol. **155:**399–408.
6. **Bétermier, M., S. Duharcourt, H. Seitz, and E. Meyer.** 2000. Timing of developmentally programmed excision and circularization of Paramecium internal eliminated sequences. Mol. Cell. Biol. **20:**1553–1561.
7. **Bruns, P. J., and D. Cassidy-Hanley.** 2000. Biolistic transformation of macro- and micronuclei. Methods Cell Biol. **62:**501–512.
8. **Cassidy-Hanley, D., J. Bowen, J. H. Lee, E. Cole, L. A. VerPlank, J. Gaertig, M. A. Gorovsky, and P. J. Bruns.** 1997. Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics **146:**135–147.
9. **Cassidy-Hanley, D., M. C. Yao, and P. J. Bruns.** 1994. A method for mapping germ line sequences in Tetrahymena thermophila using the polymerase chain reaction. Genetics **137:**95–106.
10. **Chalker, D. L.** 2008. Dynamic nuclear reorganization during genome remodeling of Tetrahymena. Biochim. Biophys. Acta **1783:**2130–2136.
11. **Chalker, D. L., and M. C. Yao.** 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in Tetrahymena thermophila. Genes Dev. **15:**1287–1298.
12. **Coyne, R., M. Nikiforov, J. Smothers, C. Allis, and M. Yao.** 1999. Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. Mol. Cell **4:**865–872.
13. **Eisen, J. A., R. S. Coyne, M. Wu, D. Wu, M. Thiagarajan, J. R. Wortman, J. H. Badger, Q. Ren, P. Amedeo, K. M. Jones, L. J. Tallon, A. L. Delcher, S. L. Salzberg, J. C. Silva, B. J. Haas, W. H. Majoros, M. Farzad, J. M. Carlton, R. K. Smith, Jr., J. Garg, R. E. Pearlman, K. M. Karrer, L. Sun, G. Manning, N. C. Elde, A. P. Turkewitz, D. J. Asai, D. E. Wilkes, Y. Wang, H. Cai, K. Collins, B. A. Stewart, S. R. Lee, K. Wilamowska, Z. Weinberg, W. L. Ruzzo, D. Wloga, J. Gaertig, J. Frankel, C. C. Tsao, M. A. Gorovsky, P. J. Keeling, R. F. Waller, N. J. Patron, J. M. Cherry, N. A. Stover, C. J. Krieger, C. del Toro, H. F. Ryder, S. C. Williamson, R. A. Barbeau, E. P. Hamilton, and E. Orias.** 2006. Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. PLoS Biol. **4:**e286.
14. **Fan, Q., R. Sweeney, and M. C. Yao.** 2000. Creation and use of antisense ribosomes in Tetrahymena thermophila. Methods Cell Biol. **62:**533–547.
15. **Frankel, J.** 2000. Cell biology of Tetrahymena thermophila. Methods Cell Biol. **62:**27–125.
16. **Galvani, A., and L. Sperling.** 2002. RNA interference by feeding in Paramecium. Trends Genet. **18:**11–12.
17. **Garnier, O., V. Serrano, S. Duharcourt, and E. Meyer.** 2004. RNA-mediated programming of developmental genome rearrangements in Paramecium tetraurelia. Mol. Cell. Biol. **24:**7370–7379.
18. **Gorovsky, M. A., M. C. Yao, J. B. Keevert, and G. L. Pleger.** 1975. Isolation of micro- and macronuclei of Tetrahymena pyriformis. Methods Cell Biol. **9:**311–327.
19. **Gratias, A., and M. Betermier.** 2003. Processing of double-strand breaks is involved in the precise excision of paramecium internal eliminated sequences. Mol. Cell. Biol. **23:**7152–7162.
20. **Jahn, C. L., and L. A. Klobutcher.** 2002. Genome remodeling in ciliated protozoa. Annu. Rev. Microbiol. **56:**489–520.
21. **Juranek, S. A., S. Rupprecht, J. Postberg, and H. J. Lipps.** 2005. snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates. Eukaryot. Cell **4:**1934–1941.
22. **Klobutcher, L. A., and G. Herrick.** 1995. Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and *Euplotes* Tec transposons. Nucleic Acids Res. **23:**2006–2013.
23. **Ku, M., K. Mayer, and J. D. Forney.** 2000. Developmentally regulated excision of a 28-base-pair sequence from the Paramecium genome requires flanking DNA. Mol. Cell. Biol. **20:**8390–8396.
24. **Le Mouël, A., A. Butler, F. Caron, and E. Meyer.** 2003. Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. Eukaryot. Cell **2:**1076–1090.
25. **Lepère, G., M. Nowacki, V. Serrano, J. F. Gout, G. Guglielmi, S. Duharcourt, and E. Meyer.** 2009. Silencing-associated and meiosis-specific small RNA pathways in Paramecium tetraurelia. Nucleic Acids Res. **37:**903–915.
26. **Liu, Y., S. D. Taverna, T. L. Muratore, J. Shabanowitz, D. F. Hunt, and C. D. Allis.** 2007. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes Dev. **21:**1530–1545.
27. **Madireddi, M. T., R. S. Coyne, J. F. Smothers, K. M. Mickey, M.-C. Yao, and C. D. Allis.** 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in *Tetrahymena*. Cell **87:**75–84.
28. **Madireddi, M. T., M. Davis, and D. Allis.** 1994. Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in *Tetrahymena*. Dev. Biol. **165:**418–431.
29. **Malavé, T. M., and J. D. Forney.** 2004. Identification of a developmentally regulated translation elongation factor 2 in Tetrahymena thermophila. Gene **326:**97–105.
30. **Malone, C. D., A. M. Anderson, J. A. Motl, C. H. Rexer, and D. L. Chalker.** 2005. Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Mol. Cell. Biol. **25:**9151–9164.
31. **Martindale, D. W., C. D. Allis, and P. Bruns.** 1982. Conjugation in Tetrahymena thermophila: a temporal analysis of cytological stages. Exp. Cell Res. **140:**227–236.
32. **Matsuda, A., and J. D. Forney.** 2006. The SUMO pathway is developmentally regulated and required for programmed DNA elimination in Paramecium tetraurelia. Eukaryot. Cell **5:**806–815.
33. **McGuffin, L. J., K. Bryson, and D. T. Jones.** 2000. The PSIPRED protein structure prediction server. Bioinformatics **16:**404–405.
34. **Meyer, E., and D. L. Chalker.** 2007. Epigenetics of ciliates, p. 127–150. *In* C. D. Allis, T. Jenuwein, D. Reinberg, and M.-L. A. E. Caparros (ed.), Epigenetics. Cold Spring Harbor Press, Cold Spring Harbor, NY.
35. **Miao, W., J. Xiong, J. Bowen, W. Wang, Y. Liu, O. Braguinets, J. Grigull, R. E. Pearlman, E. Orias, and M. A. Gorovsky.** 2009. Microarray analyses of gene expression during the Tetrahymena thermophila life cycle. PLoS One **4:**e4429.
36. **Mochizuki, K., N. A. Fine, T. Fujisawa, and M. A. Gorovsky.** 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell **110:**689–699.
37. **Mochizuki, K., and M. A. Gorovsky.** 2005. A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes Dev. **19:**77–89.
38. **Mochizuki, K., and M. A. Gorovsky.** 2004. Small RNAs in genome rearrangement in Tetrahymena. Curr. Opin. Genet. Dev. **14:**181–187.
39. **Nikiforov, M., M. Gorovsky, and C. Allis.** 2000. A novel chromodomain protein, Pdd3p, associates with internal eliminated sequences during macronuclear development in *Tetrahymena thermophila*. Mol. Cell. Biol. **20:**4128–4134.
40. **Nikiforov, M., J. Smothers, M. Gorovsky, and C. Allis.** 1999. Excision of micronuclear-specific DNA requires parental expression of Pdd2p and occurs independently from DNA replication in *Tetrahymena thermophila*. Genes Dev. **13:**2852–2862.
41. **Nowacki, M., W. Zagorski-Ostoja, and E. Meyer.** 2005. Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in Paramecium tetraurelia. Curr. Biol. **15:**1616–1628.
42. **Orias, E., E. P. Hamilton, and J. D. Orias.** 2000. Tetrahymena as a laboratory organism: useful strains, cell culture, and cell line maintenance. Methods Cell Biol. **62:**189–211.
43. **Prescott, D. M.** 1994. The DNA of ciliated protozoa. Microbiol. Rev. **58:**233–267.
44. **Rexer, C. H., and D. L. Chalker.** 2007. Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila. Eukaryot. Cell **6:**1320–1329.
45. **Saveliev, S. V., and M. M. Cox.** 1996. Developmentally programmed DNA

deletion in *Tetrahymena thermophila* by a transposition-like reaction pathway. EMBO J. **15:**2858–2869.
46. **Saveliev, S. V., and M. M. Cox.** 2001. Product analysis illuminates the final steps of IES deletion in Tetrahymena thermophila. EMBO J. **20:**3251–3261.
47. **Shang, Y., X. Song, J. Bowen, R. Corstanje, Y. Gao, J. Gaertig, and M. A. Gorovsky.** 2002. A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in Tetrahymena thermophila. Proc. Natl. Acad. Sci. U. S. A. **99:**3734–3739.
48. **Smothers, J. F., M. T. Madireddi, F. D. Warner, and C. D. Allis.** 1997. Programmed DNA degradation and nucleolar biogenesis occur in distinct organelles during macronuclear development in *Tetrahymena*. J. Eukaryot. Microbiol. **44:**79–88.
49. **Sonneborn, T. M.** 1970. Gene action in development. Proc. R. Soc. Lond. B Biol. Sci. **176:**347–366.
50. **Taverna, S. D., R. S. Coyne, and C. D. Allis.** 2002. Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell **110:** 701–711.
51. **Taylor, F. M., and D. W. Martindale.** 1993. Retroviral-type zinc fingers and glycine-rich repeats in a protein encoded by cnjB, a Tetrahymena gene active during meiosis. Nucleic Acids Res. **21:**4610–4614.
52. **Timmons, L., and A. Fire.** 1998. Specific interference by ingested dsRNA. Nature **395:**854.
53. **Tsukii, Y.** 1994. Evolution of mitochondrial DNA in Paramecium. Jpn. J. Genet. **69:**685–696.
54. **Voskühler, C., and A. Tiedtke.** 1993. Magnetic separation of phagosomes of defined age from Tetrahymena thermophila. J. Eukaryot. Microbiol. **40:**556–562.
55. **Yang, X., and M. Takahashi.** 1999. Disturbance of the determination of germinal and somatic nuclei by heat shock in Paramecium caudatum. J. Eukaryot. Microbiol. **46:**49–55.
56. **Yao, M. C., and J. L. Chao.** 2005. RNA-guided DNA deletion in Tetrahymena: an RNAi-based mechanism for programmed genome rearrangements. Annu. Rev. Genet. **39:**537–559.
57. **Yao, M. C., S. Duharcourt, and D. L. Chalker.** 2002. Genome-wide rearrangements of DNA in ciliates, p. 730–758. *In* N. Craig, R. Craigie, M. Gellert, and A. Lambowitz (ed.), Mobile DNA II. Academic Press, New York, NY.
58. **Yao, M. C., C. H. Yao, L. M. Halasz, P. Fuller, C. H. Rexer, S. H. Wang, R. Jain, R. S. Coyne, and D. L. Chalker.** 2007. Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena. J. Cell Sci. **120:**1978–1989.

Chapter 3

The transposon-derived protein *LIA5* is required to induce DNA damage response during programmed DNA rearrangement in *Tetrahymena thermophila*

Shieh, A.W. and Chalker, D.L.

Manuscript in preparation

Summary & Perspective

In attempt to uncover machineries that are involved in *Tetrahymena* DNA rearrangement, a cytological screen was conducted to identify proteins that are localized to the developing somatic nucleus during this process. Such a screen identified *LIA5* along with four other novel proteins (Yao et al., 2007), one of which (*LIA1*) have previously been shown to be essential for conjugation (Rexer and Chalker, 2007). Although Lia5 protein has no obvious homologs in other organisms, it exhibits protein architecture that is common to transposon-derived proteins. Consistent with this notion, we showed that Δ*LIA5* blocks the induction of DNA breaks associated with the excision of eliminated sequences. This process has been shown to require the domesticated *piggyBac* transposase (Baudry et al., 2009; Cheng et al., 2010). Investigating the role of *LIA5* thus provided another glimpse into how transposons have helped shape eukaryotic genomes. *LIA5* is also essential for the massive chromatin reorganization event that accompanies DNA rearrangement. In the absence of Lia5p, components of heterochromatin (histone H3 K9 and K27 methylation and chromodomain protein – Pdd1p) are present, but fail to assemble into foci. We found that foci assembly coincides with Pdd1 dephosphorylation and that both events fail in *ΔLIA5* cell. Furthermore, foci formation and Pdd1 dephosphorylation can be rescued by ectopically inducing DNA damage by UV irradiation. Taken together, we unraveled a relationship between DNA repair and DNA elimination foci, and implicated the chromodomain protein Pdd1 as a component of DNA damage response.

References

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Betermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes & Development *23*, 2478–2483.

Cheng, C.Y., Vogt, A., Mochizuki, K., and Yao, M.C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Molecular Biology of the Cell *21*, 1753–1762.

Rexer, C.H., and Chalker, D.L. (2007). Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila. Eukaryotic Cell *6*, 1320–1329.

Yao, M.-C., Yao, C.-H., Halasz, L.M., Fuller, P., Rexer, C.H., Wang, S.H., Jain, R., Coyne, R.S., and Chalker, D.L. (2007). Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena. J. Cell. Sci. *120*, 1978–1989.

INTRODUCTION

DNA damage threatens genome integrity, posing danger to the health of cells and organisms. DNA double-strand breaks (DSB) are among the most deleterious DNA lesions. They occur frequently, either as a consequence of environmental stresses or strain from essential cellular processes, including transcription and DNA replication. DSB are also introduced as part of intrinsic cellular programs. Spo11 induced breaks trigger homologous recombination during meiosis, and the Rag1/2 recombinase initiates immunoglobulin gene rearrangement during vertebrate lymphocyte maturation.

Given their prevalence and severity, if left unattended, it is not surprising that cells have multiple means to mend these lesions. DSBs are repaired by two major pathways – Homologous Recombination (HR) and Non-Homologous End Joining (NHEJ) (see Kanaar et al. 2008). HR is used primarily when an undamaged donor strand is available to template repair (e.g. repair of stalled replication forks). The Rad51 protein is a major player in this pathway, binding to single-stranded DNA after exonucleolytic processing of the damaged DNA. NHEJ is the major pathway for repairing non-replication associated breaks. Catalysis of NHEJ repair involves the binding of broken ends by the Ku70/Ku80 heterodimer, which results in the recruitment and activation of the DNA-PK complex. After processing, the broken ends are rejoined by DNA ligase IV in association with its partner XRCC4.

Upon sensing lesions in DNA, cells respond by transducing a cascade of signals to induce repair. This is collectively referred to as the DNA damage response (DDR). This process activates effector proteins that ensures proper amplification and transmission of the repair signal to facilitate repair, as well as evokes cellular responses

to either stall damaged cells in their cell cycle or trigger apoptosis in cells that fail to resolve their DNA breaks. Some of the earliest events of DDR include the phosphorylation of Histone variant H2AX (γH2AX) and the formation of DNA repair foci. DNA repair foci represent the ordered assembly of repair factors at the sites of the lesions to effect the healing of the damage DNA (reviewed in Misteli and Soutoglou 2009). These events suggest that extensive chromatin remodeling occurs upon DNA damage.

DNA repair is influenced by nuclear architecture (see Misteli and Soutoglou 2009). Evidence suggest that the repair of DSB occurs with slower kinetics in heterochromatin compared to euchromatin (Kim et al. 2007). Furthermore, a major heterochromatin component, heterochromatin protein 1 (HP1) has been shown to play a critical role in DNA damage repair (reviewed in Ayoub et al. 2009; Ball and Yokomori 2009; Dinant and Luijsterburg 2009). While some studies suggest that HP1 mobilization facilitates repair by allowing accessibility of repair machineries to damage sites, a more direct role for the actual process of repair have been implicated. In fact, it has been suggested that HP1 acts as an essential component of the DDR. The exact role of heterochromatin component(s) in DNA repair is yet unclear, however, one can envision the biological significance in the involvement of specialized mechanism to repair damage in heterochromatin domains. As these domains are rich in repetitive sequences, it is necessary to carefully regulate repair to prevent improper recombination between distal homologous sequences, which could lead to inversions or translocations of chromosomal sequences. Enlisting heterochromatin component proteins in assisting repair may help prevent such deleterious effects.

The programmed genome rearrangements of the ciliate *Tetrahymena thermophila* provide an opportunity to examine the interplay between heterochromatin and DNA repair. During *Tetrahymena* somatic nuclear differentiation, nearly 50Mb of germline-derived DNA are packaged as heterochromatin and eliminated by site-specific recombination. *Tetrahymena* are single cell eukaryotes that exhibit nuclear dimorphism, where two morphologically distinct nuclei contain different copies of the genome that individually act as the germline and the soma (Karrer 2000). The germline micronucleus houses a diploid genome that is transcriptionally silent during vegetative growth, divides mitotically, and exists to maintain and transmit genetic information to sexual progeny. Conversely, the somatic macronucleus is responsible for all gene expression necessary to support growth. This somatic genome is polyploid and highly fragmented. The macronucleus is a terminally differentiated nucleus, which divides amitotically, and is lost during sexual reproduction when a new macronucleus is formed from the parental germline.

During sexual reproduction, micro- and macronuclei differentiate from a common zygotic genome, derived from cross-fertilization of meiotic products from a mating partner's germline micronucleus. As macronuclei differentiate, 5000-6000 dispersed loci are identified and targeted for elimination. In addition the germline-derived chromosomes undergo chromosome breakage (at ~180 site) coupled with *de novo* telomere addition. Eliminating 50Mb of DNA removes germline-specific transposon-like elements and many non-coding sequences, which are termed Internal Eliminated Sequences (IESs). The recognition of IESs involves small RNA directed heterochromatin formation. The small RNAs are generated during meiosis, early into *Tetrahymena* conjugation, by

processing of bi-directional transcripts into ~28nt scan (scn)RNAs by the Dicer Like 1(Dcl1) protein (Mochizuki et al. 2002; Malone et al. 2005; Mochizuki and Gorovsky 2005). Later in the early differentiating macronuclei, the small RNAs homologous to IESs target their complementary loci for the deposition of Histone H3K9 and H3K27 methylation (Taverna et al. 2002; Liu et al. 2007). These marks recruit chromodomain-containing proteins such as Pdd1p and Pdd3p (Madireddi et al. 1996; Nikiforov et al. 2000). The IES heterochromatin then assembles with additional factors that lead to the generation of IES–containing, heterochromatin-like nuclear foci (Yao et al. 2007).

Clearly, the silencing of IESs during *Tetrahymena* nuclear reprogramming resembles the way other metazoans silence transposable elements (see Chalker and Yao 2011). In fact, ciliate IESs may have derived from transposons and other invading elements during evolution. Consistent with this notion, a domesticated *piggyBac* transposase is employed by ciliates as the excisase that cuts out IESs from their developing somatic genome. Furthermore, much like the process of V(D)J recombination, which also employs a domesticated transposase – RAG, IES excision in *Tetrahymena* involves globally induced DNA damage, as is apparent from the presence of γ-H2AX in the developing MAC during IES excision. The involvement of DNA damage is further supported by recent findings that DNA LigaseIV/XRCC4 complex and Ku80, all are major NHEJ pathway components, are essential for DNA rearrangement in *Paramecium* and *Tetrahymena*, respectively.

In this paper, we investigate the role of a developmental specific nuclear protein Lia5p in programmed DNA rearrangement. We show that *LIA5* is required for the formation of IES excision foci and is likely involved in the process of programmed DNA

damage that is associated with DNA rearrangement. Furthermore, our work implicates the essential chromodomain containing protein – Pdd1p as a component of DNA damage response.

RESULTS

Lia5p, a transposon-related protein is essential for development

Previously, we identified several *LIA* (Localized In macronuclear Anlagen) proteins that were expressed exclusively during macronuclear differentiation and showed that at least one of these, Lia1p, was required for the associated programmed genome rearrangements (Rexer and Chalker, 2007; Yao et al., 2007). The LIA proteins had no clear orthologs and few conserved domains making the prediction of function challenging. *LIA5* was initially described as encoding a 1048 amino acid (aa) glutamine-rich protein containing a putative FVYE or PHD-type zinc finger, but more recent analyses have revealed that this motif shows similarity to a zinc ribbon domain (pfam13842: Tnp_zf-ribbon_2) commonly found at the C-terminus of transposon-derived proteins. Furthermore, the central region of Lia5p shares similarity with the IS4 transposase family (pfam13843: DDE_Tnp_1_7), which includes the *Tetrahymena piggyBac* transposase (Tbp2p) (Cheng et al., 2010) (Figure 1A). Despite this structural similarity, alignment of Lia5p to Tpb2p and other predicted transposases showed that Lia5p apparently lacks the DDD catalytic triad found in active transposase (Figure 1B). Domestication of transposon-derived proteins, including Rag1/2 and Tbp2p, has created novel pathways acting on eukaryotic chromosomes. Our data suggest that Lia5p may have similarly evolved from domestication of a transposable element protein.

While *LIA5* expression occurs exclusively during development, peaking between 6-8 hours after initiation of conjugation (Yao et al., 2007), its mRNA can be detected as early as 2 hrs (Figure 1D). To determine the timing of protein accumulation, we tagged the endogenous gene on its amino terminus with a hemagglutinin (HA) epitope and examined its expression. HA-Lia5p could not be detected until 8hrs into conjugation indicating that it does not accumulate until new macronuclei form (Figure 1E). Together with our previous observation that GFP-Lia5p localizes to the developing macronuclei (Yao et al., 2007) these data led us to suspect that Lia5p participates in the genome reorganization of the developing somatic macronucleus.

To determine whether Lia5p is essential for macronuclear differentiation, we deleted all copies of *LIA5* from both the micro- and macronuclear genome by homologous gene replacement with the *neo3* paramomycin-resistance cassette (Shang et al., 2002) (Figure 2A). Creation of these *LIA5* knockout (Δ*LIA5*) cell lines was verified by Southern blot analysis, and the loss of all expression was confirmed using rtPCR (Figure 2B, C). When Δ*LIA5* strains were mated, no viable progeny were produced indicating that this gene has an essential role during conjugation.

To determine the stage of development that knockout cells failed to complete, we examined the nuclear morphology of the cells throughout conjugation. Δ*LIA5* cells progressed through early stages (meiosis, nuclear exchange and karyogamy, and formation of new macronuclei) at a rate that was similar to that of wild type (wt) cells with no obvious developmental delays (data not shown). Wild type mating cells complete conjugation after pair separation by eliminating one of two micronuclei, producing cells that have two newly differentiated macronuclei and one micronucleus. These cells are

poised to divide the one remaining micronucleus and undergo cytokinesis once fed. We found that Δ*LIA5* cells arrested as exconjugants, prior to elimination of one micronucleus. Furthermore, the new macronuclei formed in mutant conjugants failed to fully amplify their genomic DNA as indicated by the weak intensity of DAPI staining relative to wt. This two macronuclei/two micronuclei terminal arrest phenotype has been commonly observed in mutants lacking genes (such as *DCL1, PDD1* and *LIA1*) that are required for programmed DNA rearrangements (Coyne et al., 1999; Mochizuki et al., 2002; Malone et al., 2005; Rexer and Chalker, 2007)

***LIA5* is required for DNA elimination and chromosome fragmentation**

Approximately 50 Mb of germ line-derived DNA are eliminated from nearly 6000 loci during differentiation of the somatic genome. When cells are unable to complete these DNA rearrangements, it triggers the developmental arrest that we observed for *ΔLIA5* cells. We therefore examined loci that undergo either DNA elimination or chromosome breakage to determine whether *LIA5* is required for these somatic genome remodeling events.

The M IES is a well-characterized eliminated sequence (Figure 3A). We isolated single exconjugants from either wt or Δ*LIA5* mating populations and used nested PCR to assess this element's rearrangement status. Each successful rearrangement of the M IES generates one of two alternative products, removing either a 0.6kb or a 0.9kb fragment. PCR using primers flanking the IES can detect both of these rearranged products as well as any unrearranged loci. M IES rearrangement was readily detected in wt exconjugants as the predominant PCR products are less that 600 bp. In contrast, in Δ*LIA5* mated cells

we detected accumulation of a larger product migrating at the size expected for the unrearranged, germ line form (Figure 3B). Thus Δ*LIA5* cells are unable to excise this IES from their developing macronuclei.

In addition to IES excision, *Tetrahymena* macronuclear differentiation involves breakage of chromosomes at ~180 loci followed by *de novo* telomere addition. Although the connection between chromosome breakage and IES excision is poorly understood, strains lacking genes that are required for IES excision fail to fragment chromosomes as well. To test whether *LIA5* is required for this process, we examined the chromosome breakage site found just downstream of the *LIA1* gene (Malone et al., 2005). DNA isolated from post-conjugation populations of wt or mutant cells was digested with EcoRI and analyzed by Southern blot using a *LIA1*-specific radiolabeled probe. In wt exconjugant populations, *de novo* breakage is readily observed as a ~2.2kb fragment, which migrates faster than the mature macronuclear form (at ~2.5-2.6kb), which has fully elongated telomeres (the major form detected in vegetatively growing cells and observed in the post-conjugative populations due to unmated cells in the populations tested) (Figure 3D). No evidence of chromosome breakage is observed in Δ*LIA5* or *ΔDCL1* mutant populations, indicated both by the absence of the 2.2kb fragment and increased abundance of the unrearranged micronuclear form migrating at 10.5kb. Thus *LIA5* is required for both IES excision and chromosome breakage.

Δ*LIA5* strains establish heterochromatin, but fail to reorganize it into nuclear foci

IESs are targeted for elimination from developing macronuclei using a mechanism of small RNA directed establishment of heterochromatin modifications

(Mochizuki et al., 2002; Taverna et al., 2002; Liu et al., 2007). To determine whether Lia5p acts in the establishment of heterochromatin modifications on IESs or in downstream events, we examined the ability of Δ*LIA5* cells to complete the critical steps in the macronuclear development. Wt cells generate germline specific scnRNAs during meiosis that direct histone H3 K9 and K27 methylation to homologous IES in differentiating macronuclei (Figure 4). Mating populations of Δ*LIA5* accumulated wt levels of scnRNAs (Figure 4A) and acquired methylation on K9 and K27 of histone H3 (Figure 4B, 9hrs). These findings were not unexpected as the peak of Lia5p expression occurs after heterochromatin is targeted to IESs.

Upon establishment of heterochromatin modifications on IES, proteins required for IES excision, including chromodomain-containing proteins Pdd1p and Pdd3p, assemble on the modified chromatin, which is followed by the redistribution of modified chromatin into nuclear foci (Figure 4B, WT-14hrs). The purpose of organizing IESs into these DNA elimination foci is not known, but it has been suggested that their formation facilitates IES excision and/or the degradation of the associated germline-limited DNA. Even though *ΔLIA5* cells establish heterochromatin modifications, the nuclear reorganization of the modified sequences does not occur as both H3K9 and K27 methylation remains dispersed throughout the developing somatic macronuclei (Figure 4B, ΔLIA5-14hrs).

Clearly, partitioning thousands of loci into a countable number of distinct foci necessitates massive nuclear reorganization, which is readily visualized through tracking the dynamic localization of the essential chromodomain protein Pdd1p (Madireddi et al., 1996; Smothers et al., 1997; Yao et al., 2007). Pdd1p, like the methylated IES chromatin

to which it binds, is initially dispersed throughout the developing macronucleus, then assembles into condensed nuclear foci coincident with the onset of IES excision. To further examine how loss of *LIA5* affects this nuclear reorganization, we followed the localization of Pdd1p tagged with yellow fluorescent protein (YFP). Whereas Pdd1-YFP localized to distinct nuclear foci in wt exconjugants (Figure 4C, WT-14hrs), it remained dispersed in the developing macronuclei of Δ*LIA5* cells, even 30 hrs after initiating mating when the fusion protein had disappeared from the fully differentiated macronuclei of wild type cells, presumably as IESs were eliminated (Figure 4C, 30hrs).

These data indicate that Lia5p plays a critical role in the organization of modified IES chromatin into DNA elimination foci. To more closely examine the participation of Lia5p in these events, we asked whether Lia5p functions as a component of these foci. To follow Lia5p localization through the differentiation of new macronuclei, we expressed Pdd1-CFP in strains expressing either an endogenous N-terminally tagged HA-Lia5 (see Figure 1E) or Lia5-YFP (a C-terminally tagged allele expressed from a high copy rDNA vector) and asked whether Lia5 assembles into Pdd1p-containing foci. Lia5p was detected in the developing macronucleus as soon as they emerged (Figure 5A). As macronuclei developed, Lia5p became increasingly concentrated at distinct regions within macronuclei, occupying similar nuclear domains as Pdd1p. Even so, Lia5p did not obviously co-localize with Pdd1p. Whereas Pdd1-CFP formed compact foci, the tagged Lia5p appeared to concentrate in regions surrounding the Pdd1p foci. This is most clearly observed in HA-Lia5p expressing cells (Figure 5B). The Lia5-YFP also localized surrounding Pdd1-CFP, but appeared more dispersed than HA-Lia5 in some mating pairs. It is possible that the ratio of dispersed:localized protein is altered by the large tag

interfering with some Lia5p action as expression of the Lia5-YFP construct in *ΔLIA5* cells did not efficiently rescue the knockout. Nevertheless, as both the HA-Lia5 and Lia5-YFP localize peripherally to Pdd1p foci, our results argue that Lia5p is not a core structural component of these foci. To assess whether Lia5p is recruited to the periphery of DNA elimination foci by Pdd1p, we asked whether these two proteins co-immunoprecipitated from conjugating cells. Immunoprecipitation of HA-Lia5 did not co-precipitate with Pdd1p or Pdd3p, further showing that Lia5p is not present in the core foci (Figure 5C).

LIA5 is required for proper regulation of Pdd1p phosphorylation during conjugation

Like HP1 proteins of other eukaryotes, Pdd1p shows regulated phosphorylation (Madireddi et al., 1996; Smothers et al., 1997), although its function(s) is largely unexplored. Pdd1p phosphorylation has been shown to peak early during macronuclear differentiation and decrease as macronuclei mature (Figure 6A). Thus, Pdd1p dephosphorylation coincides with foci formation and IES excision. To determine whether Pdd1p dephosphorylation might regulate its ability to aggregate into foci, we monitored its phosphorylation state in wt and Δ*LIA5* mating cells. The phospho-isoforms of Pdd1p in whole cell extracts were resolved on 9% SDS polyacrylamide gels and detected with anti-Pdd1p antibodies. For both wt and Δ*LIA5*, Pdd1p migrates as a doublet representing the phosphorylated (upper band) and unphosphorylated (lower band) proteins, which were clearly evident by 9hr. By 12hr when Pdd1p is primarily found in DNA elimination foci, most of the Pdd1p in wt cells appeared to be dephosphorylated as judged by the

collapse of the doublet into a single band (Figure 6A). Furthermore, as macronuclear differentiation proceeds, Pdd1p levels decrease, correlating with the period during which IESs are eliminated (wt 12-15hrs). In contrast, Pdd1p continues to accumulate between 12 and 15 hrs in Δ*LIA5* mating cells, which obscured the resolution of the Pdd1p doublet (Figure 6A). Dilution of 12hr and 15hr protein samples isolated from Δ*LIA5* cells showed that Pdd1p remained phosphorylated (Figure 6B left panel). Alkaline phosphatase treatment of these samples resulted in collapse of the double to a single band, showing that the shift in migration is due to phosphorylation (Figure 6B right panel).

In the above experiments, we could not distinguish between whether failure of Pdd1p dephosphorylation directly blocked foci formation, or whether Δ*LIA5* cells arrested at a stage of development prior to the loss of this modification. To assess whether Pdd1p phosphorylation may inhibit foci formation, we examined the state of Pdd1p in Δ*DCL1* cells. Mutations in components of the RNAi pathway, e.g. Δ*DCL1* or Δ*TWI1*, which lead to failure in scnRNA-directed heterochromatin formation, assemble Pdd1p foci in developing macronuclei as soon as these nuclei appear in cells (Figure 6C). While it is not clear how these foci relate to normal DNA elimination foci as chromatin modifications on IESs are not established, their presence suggest that Pdd1p can assemble into sub-nuclear domains in the absence of small-RNA directed heterochromatin targeting. If phosphorylation prevents foci assembly, Pdd1p should remains unphosphorylated in Δ*DCL* mutants. However, examination of Pdd1p isolated from these mutants show that the Pdd1p phosphorylated isoforms accumulate and, as in *ΔLIA5* cells, remains modified throughout conjugation (Figure 6B). Thus, phosphorylation does not appear to be a physical barrier to the assembly of Pdd1p foci.

Ectopic induced DNA damage is sufficient to rescue Pdd1 protein dephospohrylation and foci formation in *ΔLIA5*

The exact function of DNA elimination foci is not known. Their assembly may bring together components of the DNA elimination machinery to facilitate the excision of nearly 6000 dispersed IESs, or alternatively, they may form after excision to sequester excise IES and/or aid in repair of the programmed DNA double strand breaks. The previous observation that foci do not form upon knockdown of the domesticated transposase, *TPB2*, suggests that excision may be a prerequisite to the formation of these sub-nuclear structures (Cheng et al., 2010). Our analysis of Δ*LIA5* cells also suggests that DNA breaks at IES may lead to foci formation as these mutants fail to eliminate IESs and Pdd1p remains dispersed. To determine whether Δ*LIA5* fail to make programmed DNA breaks, we looked for the presence of γH2AX accumulation when DNA double strand break occurs at IES junctions. This conserved marker of DNA damage has been shown to accumulate in the micronuclei during meiosis, when recombination-associated DNA breakage occurs (Song et al., 2007). Both wt and Δ*LIA5* strains contain γH2AX in the meiotic nuclei (Figure 7A, 3 hrs); however, after developing macronuclei emerge, γH2AX was detected only in wt cells and was largely absent from Δ*LIA5* macronuclei (Figure 7A, 10 hrs). This indicates that Δ*LIA5* cells are deficient in initiating IES associated double strand breaks.

The behavior of Pdd1p and γH2AX in Δ*LIA5* strains suggest that DNA elimination foci form as a response to programmed DNA breaks. If this is indeed the case, we reasoned that we could induce formation of Pdd1p foci by introducing DNA

damage ectopically. Cells expressing Pdd1-CFP were exposed to UV treatment at a stage where Pdd1p is normally dispersed in the developing macronucleus. We observed that Pdd1p assembled into foci in response to this induced DNA damage (Figure 7B). These foci localized to the sites of Ligase IV accumulation, indicating that Pdd1p is recruited to sites of DNA damage.

To further link DNA damage to DNA elimination foci formation, we treated Δ*LIA5* cells with UV. We found that introduction of ectopic DNA lesions rescued the foci formation phenotype in Δ*LIA5* conjugants (Figure 7C). To further examine whether this foci formation mimicked the programmed DNA elimination response, we examined the phosphorylation state of Pdd1p. Whereas phosphorylated Pdd1p accumulated in *LIA5* mutants, induction of ectopic breaks via UV treatment induced Pdd1 dephosphorylation (Figure 7D). Taken together, our data indicate that DNA elimination foci form as a consequence of a programmed DNA damage response and that *LIA5* is somehow involved in triggering such response. Incidentally, these observations are consistent with the fact that Lia5p resembles transposon-derived proteins.

DISCUSSION

Tetrahymena thermophila somatic nuclear differentiation requires genome-wide remodeling to generate the transcribed genome for the next generation. In this study, we found that *LIA5* encodes a protein critical for the chromosome breakage and DNA elimination events that fragment and streamline the genome for efficient gene expression. Lia5p acts after the establishment of RNAi-directed heterochromatin modifications in the zygotic genome (figure 4), but prior to the initiation of ds breaks that results in the

elimination of the marked sequences (IESs). Failure of conjugating *ΔLIA5* cells to excise IESs is apparent as the unrearranged micronuclear form of the M IES accumulates (Figure 3), DNA elimination foci do not form (Figure 4), and the detection of phosphorylated H2AX is severely dimished (Figure 7). Cells lacking *LIA5* display similar developmental phenotypes as in *ΔLIA1* cells and cells with knocked down expression of *TPB2*, which encodes the domesticated *piggyBac* transposase that performs IES excision (Cheng et al. 2010; Rexer et al. 2007). In all three cases, heterochromatin modifications are established, but Pdd1p remains dispersed. These studies make it clear that multiple proteins must cooperate to excise the newly established heterochromatin, marked by the Twi1p-scnRNA machinery, from the somatic genome.

DNA elimination foci form in response to IES excision

The failure of *ΔLIA5* cells to form DNA elimination foci led us to investigate what triggers the assembly of these structures. As Pdd1p, an Hp1-like chromodomain-containing protein, is a major component, it has been suggested that foci represent the mature form of newly established heterochromatin in the developing macronucleus (See Chalker 2008; Madireddi et al. 1996). These foci grow larger in size and fewer in number, appearing to coalesce as macronuclear differentiation proceeds (Yao et al. 2007). In this model, heterochromatin is fully compartmentalized prior to its elimination, and Lia5p acts as an essential chromatin protein that participates in the sub-nuclear partitioning of IESs. We found that Lia5p is not present in the central core of DNA elimination foci, but its localization surrounding Pdd1p structures would be consistent with its involvement in this nuclear reorganization. Nevertheless, the observation that

these structures do not form in *ΔLIA5* cells or in other mutant lines that fail to excise IESs suggested the equally likely possibility that foci develop in response to DNA breaks introduced by IES excision (Cheng et al. 2010; Rexer et al. 2007). Our data support this second hypothesis as we could rescue foci formation in the absence of *LIA5* by introducing ectopic DNA damage (Figure 7).

The redistribution of DNA repair proteins into sub-nuclear foci is a dynamic process induced upon DNA damage. DNA repair foci have not been described in *Tetrahymena* and, before this study, the relationship between DNA elimination structures and repair foci was largely unexplored. We showed here that the NHEJ protein, Ligase IV, is recruited to foci containing Pdd1p in UV-treated mating cells, indicating that Tetrahymena repair proteins are reorganized upon DNA damage. We also found that Pdd1p is dephosphorylated in response to UV-induced DNA damage. These UV-induced events mimic what occurs normally upon IES excision in wild-type cells when DNA elimination foci form and Pdd1p is dephosphorylated, coincident with the introduction of programmed ds breaks. These data strongly suggest that the organization of IES heterochromatin into sub-nuclear structures is triggered by a DNA damage response and implicate Pdd1 as a target of this response. Hp1 is known to respond to DNA damage (reviewed in Ayoub et al. 2009; Ball and Yokomori 2009; Dinant and Luijsterburg 2009). Our findings show that heterochromatin proteins are involved in the repair of DNA damage in ciliates and raise the possibility that their roles are ancient and potentially evolutionarily conserved.

Host DNA repair processes have been shown to participate in ciliate programmed DNA rearrangements. The major NHEJ components Ligase IV and XRCC4 are required

for DNA rearrangement that occur during *Paramecium* macronuclear differentiation (Kapusta et al., 2011), and the *Tetrahymena* Ku80 protein (Tku80p) was recently shown to play an essential role in *Tetrahymena* development as well (Lin et al., 2012). Cells lacking Tku80 (*ΔTKU80*) still excise IESs, but are unable to rejoin the resulting ds breaks. The unrepaired chromosomes are eventually degraded leaving developing macronuclei devoid of detectable DNA. This phenotype was first described for *ΔDIE5* strains (Matsuda et al., 2010). The novel protein Die5p is required to complete DNA rearrangements, but appears to act after IES excision as Pdd1p foci still form. This is in contrast to *ΔTKU80* cells that do not form DNA elimination foci even though IESs are excised. This suggests that the ability to respond to ds breaks is still intact in *ΔDIE5* strains, but not in *ΔTKU80* cells. We propose that Tku80p acts as a DNA damage sensor, without which, the damage response is not conveyed, resulting in the failure to form repair foci. In wild-type cells, Tku80p was not observed to co-localize with Pdd1p, which we believe further indicates that its main roles are to, sense the damage, signal the formation of Pdd1p foci, and protect the free ends until rejoined. Die5p must act downstream of Tku80p assisting in the repair of macronuclear-destined DNA after IES excision.

Even though our data indicated that DNA elimination structures form in response to DNA damage, several observations suggest to us that they may not be analogous to DNA repair foci in other eukaryotes. Tku80p is not observed within DNA elimination structures, but it is required to rejoin the macronucleus-destined DNA flanking IESs after their excision, suggesting that repair of the developing somatic genome appears to occur outside of foci. We believe these data indicate that formation of DNA elimination foci is

more likely the response of heterochromatin to DNA damage. The repeat-rich DNA found in heterochromatin can be challenging to repair pathways. The homologous recombination machinery may have difficulty distinguishing between an undamaged sister chromatid and other nearby homologous sequences when selecting a repair template, which could lead to aberrant genome rearrangements. In Drosophila, it has been shown that homologous recombination is repressed in heterochromatin until the free ends of damaged DNA are moved outside of the heterochromatin domain, where repair can occur without these complications. Similar relocation of ds breaks outside of heterochromatin compartments has been observed in mammalian cells as well. The formation of DNA elimination foci may represent complementary phenomenon leading to the sequestration of the repeat rich IESs as heterochromatin away from repair proteins, promoting accurate joining of the retained genomic sequences.

Transposons and the origin programmed DNA elimination.

LIA5 was first described as encoding a glutamine rich protein containing a zinc finger domain. Through further analysis, Lia5p appears to share structural similarity with IS4 family of transposases. Although lacking the conserved DDD/E catalytic residues, Lia5p contains a domain that belongs to the DDE_Tnp_1_7 family as well as a Tnp_zf-ribbon_2 domain. This protein architecture is shared by many transposon-derived proteins, including the *piggyBac* transposase. Therefore, like the domesticated *piggyBac* transposase gene – *TPB2*, *LIA5* may be the remnant of a transposon that was domesticated during evolution to enforce the silencing of transposon-like IESs. Domesticated transposases have proven important for essential cellular functions across

diverse eukaryotic organisms. Among the most well known examples are the Rag proteins involved in the vertebrate adaptive immune systems, and CenPB proteins, which is essential for centromere function in many eukaryotes.

Although we have not identified the biochemical function of Lia5p, the lack of DNA elimination foci and γH2AX staining in the absence of *LIA5* suggests that the protein it encodes is involved in organizing IES chromatin or recruiting *TPB2* to initiate DNA cleavage. Intriguingly, overlapping the DDE_Tnp_1_7 and the Tnp_zf_ribbon_2 domains are the predicted DNA binding domain of the mismatch repair MutS subunit (MutSD) and the FYVE/PHD Zn finger domains, respectively. While the significance of these domains remain unclear, MutSD links Lia5p to a possible role in DNA repair, and FYVE/PHD Zn finger domain has been implicated in proteins with diverse functions, including chromatin remodeling. Further analysis of both domains to elucidate the exact mechanism of Lia5p function and the origins of *LIA5* gene will provide important insight into the role of transposons in shaping eukaryotic genomes.

METHODS AND MATERIALS

Cell lines and culture. *Tetrahymena thermophila* cells were grown in liquid culture at 30°C according to standard methods (Orias et al., 2000). Wild type strains (B2086, CU427, CU428) and the micronucleus-defective 'star' strains (B*VI, B*VII) were originally obtained from Peter Burns (Cornell University, Ithaca, NY) and are available from the *Tetrahymena* Stock Center (http://tetrahymena.vet.cornell.edu/). These strains transformed with constructs to create knockout strains or cell lines expressing epitope-taged proteins. *ΔDCL1* and Δ*PDD1* strains were described (Malone et al., 2005; Motl and

Chalker, 2011). Cells were made competent to mate by overnight starvation (>6hours) in 10 mM Tris-HCl (pH 7.4), and conjugation was induced by mixing starved cultures of mating compatible strains at equal cell densities (~$2.5x10^5$ cells/ml).

RT-PCR. cDNA was synthesized from 4μg of total RNA isolated at different stages of conjugation with SuperScript II reverse transcriptase (Invitrogen)(Fan et al., 2000; Malone et al., 2005). Oligonucleotide primers designed to flank the 6^{th} intron. (Lia5rtFw 5'-ttctctaggctaagcaccctaaaa-3' Lia5rtRv 5'-tccattgtacccattgttcatt-3') were used to monitor *LIA5* expression by PCR.

Generation of *LIA5* knockout and expression strains.

A *LIA5* knockout construct pLia5KO was generated using a Multisite Gateway Cloning kit (Invitrogen) as previously described (Matsuda et al., 2010; Motl and Chalker, 2011). DNA corresponding to regions upstream and downstream of the *LIA5* coding sequence was amplified from CU427 genomic DNA using the following primer pairs:

LIA5upFw 5'- GGGGACAACTTTGTATAGAAAAGTTggtacctacaaggacaatggcaccaa-3'

LIA5upRv 5' GGGGACTGCTTTTTTGTACAAACTTGtggctaaaatttctgcagtcg-3' and

LIA5downFw 5'- GGGGACAGCTTTCTTGTACAAAGTGgccaatagataaaatggcacct-3'

LIA5downRv 5'- GGGGACAACTTTGTATAATAAAGTTggtacctcatttccgaaaaatatcat-3', respectively (Uppercase letters are att sequences added to facilitate Gateway recombination). The PCR products were used in BP recombination reactions with donor vectors. The resulting clones were combined with the pENTR-Neo3 selection cassette in a multi-plasmid LR clonase reaction into pDEST-R4-R3 to create pLia5KO. This

construct was linearized with Acc65I (site underlined in primers) and introduced by biolistic transformation into 2.5 to 3.5 hr conjugating Cu428xB2086 to obtain micronuclear transformants (Cassidy-Hanley et al., 1997; Bruns and Cassidy-Hanley, 2000). Genomic exclusion crosses of heterozygous germ line transformants with star strains B*VI or B*VII generated homozygous mutants that were subsequently crossed to produce complete *LIA5* knockouts missing all copies of the gene from both the micro- and macronucleus.

A hemagluttinin (HA)-tagged *LIA5* expression construct was created using a two step overlapping PCR strategy (Rexer and Chalker, 2007) to introduce the HA coding sequence immediately after the *LIA5* start codon using primers HALIA5upFw 5'-CACCGGGCCCtagctggcattttcaataaataaa-3' with HALIA5upRv 5'-*taatcaggaacatcataaggat*acattttaaattaattagttttcaaaggggataacttc-3' and HALIA5downFw 5'-*ccttatgatgttcctgattatgct*gaattaggagaagcagatttacatacatcac-3' with HALIA5downRv 5-CTCGAGaaaatgtattagcagctttaaatgtc-3'. The HA coding sequence is italicized and restriction enzyme sites (ApaI, XhoI) are underlined. Primers LIA5dsFw 5'-GGATCCtgatatttttcggaaatgagga-3' and LIA5dsRev 5'-CCGCGGagcaagcaaaggcgaaaata-3' were used to clone the LIA5 downstream genomic sequence (BamHI and SacII sites are underlined). Amplified PCR products were inserted into p4T2 vector containing the histone H4 promoter driven NEO cassette (Gaertig et al., 1994) to create the p4T2-HALia5 knock-in construct, which was lineralized with ApaI and SacII and introduced into the macronucleus of starved cells by biolistic transformation. Phenotypic assortment was achieved by selecting transformants in gradually increased concentration of

paramomycin until complete replacement of the *LIA5* locus with the HALia5 allele was achieved.

Co-Immunoprecipitation and Western Blot analysis. Co-Immunoprecipitation of Lia5p with Pdd1p was performed as previously described (Rexer et al. 2007) using HALia5 strains described above. For werstern blot analysis, Immunoprecipitated samples, or total protein (isolated from ~$1x10^6$ cells at different stages of conjugation with Lysis Buffer) were boiled with 1X Laemmli Sample Buffer. Protein samples were separated with 4% stacking, 9% resolving polyacrylamide gel elecrophoresis, transferred onto nitrocellulose membrane and detected primary antibodies: polyclonal αPdd1 (abcam 1:1000), αPdd3 (abcam 1:1000) and αHA (covance PRB-101P, 1:500), followed by HRP conjugated Goat anti-Rabbit IgG as a secondary. Blots were then overlayed with Pierce supersignal west duro chemiluminesce substrate and imaged using a Fuji imager.

IES excision analysis. Single cells from mating pairs were isolated and lysed for PCR as previously described (Coyne et al., 1999; Mochizuki et al., 2002). The M IES was analyzed with two successive rounds of PCR with following nested primers:

Round I primers:

MIFw 5'-AGCTTAAACAAATGCCATATTGAG-3'

MIRv 5'-AAGGGGGGTGGGGAGGGAGAAGGA-3'

Round II primers:

MIIFw 5'-TACGATAGATCGACTGACGG-3'

MIIRv 5'-GTGGGGAGGGAGAAGGATTCAAC-3'

Northern and Southern blot analyses. Small RNA and Chromosome breakage analysis were performed as previously described (Malone et al., 2005; Matsuda et al., 2010). *LIA5* locus probe was the 850bp fragment isolated from BglII/BamHI digest of the HALia5 construct (see above).

Indirect immunoflourescence. Cells were fixed by adding 10µl of fixative (2 parts satuarated mercuric chloride plus 1 part 95% Ethanol) to 3mls of cells in 10mM Tris. After incubating for 5 min at room temperature, fixative was removed. The cells were washed once with 6mls and then resuspended in 1ml of 100% methanol. For immuno-staining, cells were dropped directly onto the slides, dried and rehydrated with 1xTBS. Rehydrated cells were blocked with 1xTBS containing 1% BSA and 0.01% Tween20. Primary Antibodies: polyclonal αH3K9me2 (upstate 07-441, 1:500), monoclonal αH3K27me3 (Abcam 6002, 1:500), polyclonal αHA (covance PRB-101P, 1:500), polyclonal αH2AvD (Rockland PS137, 1:1000). Secondary Antibodies: αRabbit and αMouse Alexa488 and Alexa594 (Invitrogen 1:500).

Pdd1 phosphorylation analysis. To resolve the different phospho-isoforms of Pdd1, total protein were isolated from ~$1x10^6$ cells at different stages of conjugation by boiling cells with 1X Laemmli Sample Buffer. Protein samples were separated with 4% stacking, 9% resolving polyacrylamide gel elecrophoresis, transferred onto nitrocellulose membrane and blotted with polyclonal αPdd1 antibody (Abcam 5338, 1:1000). Alkaline phosphatase treatment was performed as previously described (Smothers et al., 1997)

UV irradiation. Irradiation was performed with GS Gene Linker™ UV Chamber (Bio-Rad). Conjugating cells in 3-7mls suspension (10mM Tris) were exposed to150mJ of UV-C (254nm) irradiation. Cells were then covered in aluminum foil to prevent photolyase repair and allowed to recover in 30°C for at least 6 hours before harvesting for assays.

REFERENCES

Bruns, P.J., and Cassidy-Hanley, D. (2000). Biolistic transformation of macro- and micronuclei. Methods Cell Biol. *62*, 501–512.

Cassidy-Hanley, D., Bowen, J., Lee, J.H., Cole, E., VerPlank, L.A., Gaertig, J., Gorovsky, M.A., and Bruns, P.J. (1997). Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics *146*, 135–147.

Cheng, C.Y., Vogt, A., Mochizuki, K., and Yao, M.C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Molecular Biology of the Cell *21*, 1753–1762.

Coyne, R.S., Nikiforov, M.A., Smothers, J.F., Allis, C.D., and Yao, M.C. (1999). Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. Molecular Cell *4*, 865–872.

Fan, Q., Sweeney, R., and Yao, M.C. (2000). Creation and use of antisense ribosomes in Tetrahymena thermophila. Methods Cell Biol. *62*, 533–547.

Gaertig, J., Gu, L., Hai, B., and Gorovsky, M.A. (1994). High frequency vector-mediated transformation and gene replacement in Tetrahymena. Nucleic Acids Res. *22*, 5391–5398.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. PLoS Genet. *7*, e1002049.

Lin, I.-T., Chao, J.-L., and Yao, M.-C. (2012). An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. Mol. Biol. Cell *23*, 2213–2225.

Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2007). RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. Genes & Development *21*, 1530–1545.

Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., and Allis, C.D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell *87*, 75–84.

Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Mol. Cell. Biol. *25*, 9151–9164.

Matsuda, A., Shieh, A.W.-Y., Chalker, D.L., and Forney, J.D. (2010). The Conjugation-Specific Die5 Protein Is Required for Development of the Somatic Nucleus in both Paramecium and Tetrahymena. Eukaryot Cell *9*, 1087–1099.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Motl, J.A., and Chalker, D.L. (2011). Zygotic expression of the double-stranded RNA binding motif protein Drb2p is required for DNA elimination in the ciliate Tetrahymena thermophila. Eukaryotic Cell *10*, 1648–1659.

Orias, E., Hamilton, E.P., and Orias, J.D. (2000). Tetrahymena as a laboratory organism: useful strains, cell culture, and cell line maintenance. Methods Cell Biol. *62*, 189–211.

Rexer, C.H., and Chalker, D.L. (2007). Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila. Eukaryotic Cell *6*, 1320–1329.

Shang, Y., Song, X., Bowen, J., Corstanje, R., Gao, Y., Gaertig, J., and Gorovsky, M.A. (2002). A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in Tetrahymena thermophila. Proc. Natl. Acad. Sci. U.S.A. *99*, 3734–3739.

Smothers, J.F., Mizzen, C.A., Tubbert, M.M., Cook, R.G., and Allis, C.D. (1997). Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures. Development *124*, 4537–4545.

Song, X., Gjoneska, E., Ren, Q., Taverna, S.D., Allis, C.D., and Gorovsky, M.A. (2007). Phosphorylation of the SQ H2A.X Motif Is Required for Proper Meiosis and Mitosis in Tetrahymena thermophila. Molecular and Cellular Biology *27*, 2648–2660.

Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in *Tetrahymena*. Cell *110*, 701–711.

Yao, M.-C., Yao, C.-H., Halasz, L.M., Fuller, P., Rexer, C.H., Wang, S.H., Jain, R., Coyne, R.S., and Chalker, D.L. (2007). Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena. J. Cell. Sci. *120*, 1978–1989.

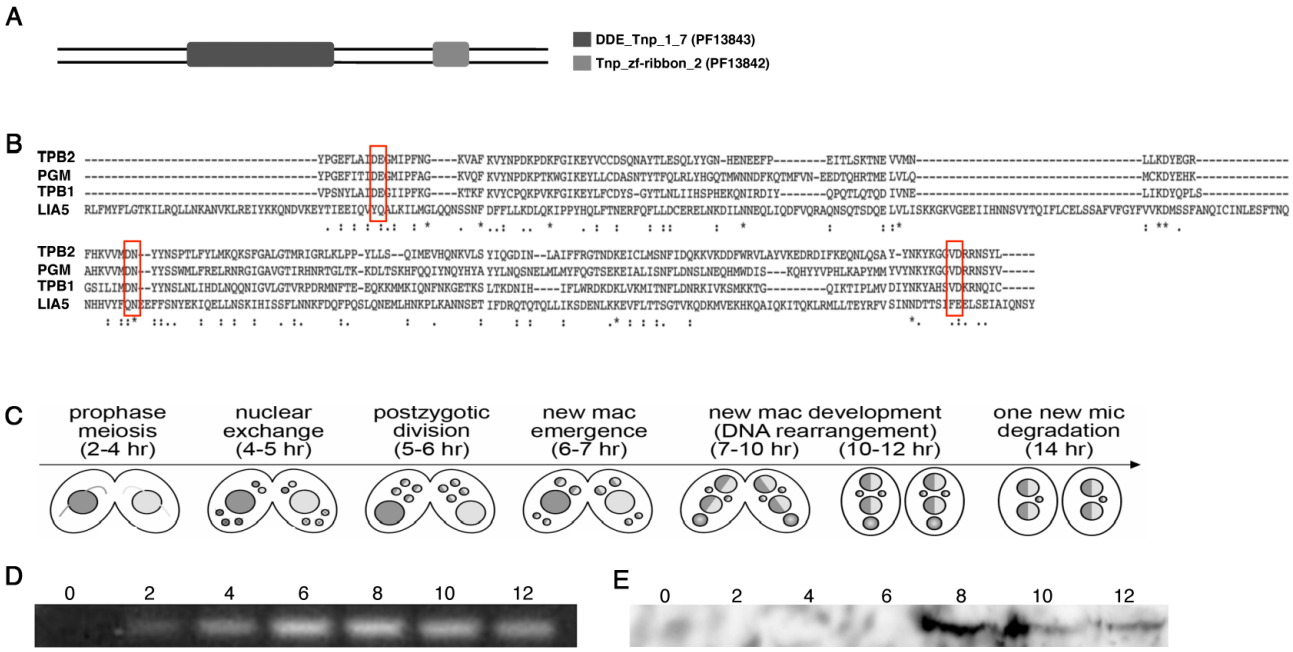

Figure 1 – Lia5 is a developmentally expressed transposon-like protein
(A) Representation of Lia5 protein. Conserved DDE_Tnp_1_7 and Tnp_zfribbon_2 domains are indicated. (B) Protein sequence alignment of Lia5 with ciliate domesticated *piggyBac* transposases including *Pramecium* PGM and *Tetrahymena* TPB2 and TPB1. Conserved DDD domains are indicated with red boxes. (C) Illustration of the major nuclear events during *Tetrahymena* conjugation. The hours associated with each developmental stage is indicated. (D) rt-PCR for Lia5 gene expression at indicated stages of conjugation. (E) Western blot analysis for Lia5 protein expression profile throughout conjugation.

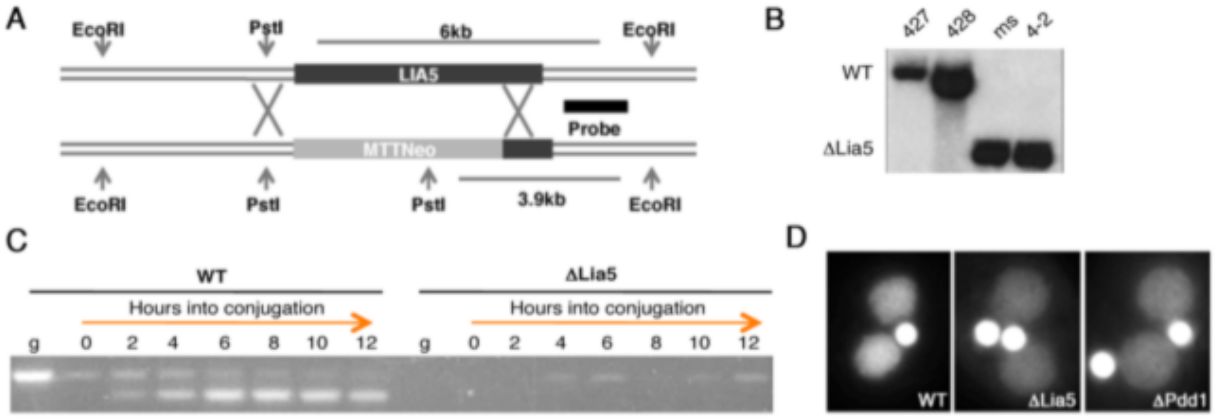

Figure 2 – Lia5 is essential to complete conjugation

(A) Illustration of Lia5 gene replacement strategy by neomycine resistance selectable marker driven by the inducible metallothionein (MTT) promoter. "X" marks the site in which homologous recombination would take place between the knockout construct and the endogenous locus for gene replacement. Arrows denote the restriction enzyme cut sites used for southern analysis to verify gene knockout. The probe, as indicated on the illustration, detects the wild-type locus of Lia5 at 6kb and the knockout copy at 3.9kb. (B) Southern blot hybridization of DNA isolated from wild type (427, 428) and ΔLia5 strains (ms, 4-2) using the probe shown in (A). Expected products from wild type and knockout copies are labeled as 'WT' and 'ΔLia5', respectively. (C) rtPCR for the expression of Lia5 in conjugating wild type (WT) and Lia5 knockout (ΔLia5) at indicated time points. Primers used span an intron. g, genomic DNA used as control for amplification. (D) Fluorescent images of representative DAPI stained WT, ΔLia5 and ΔPDD1 strains post conjugation.

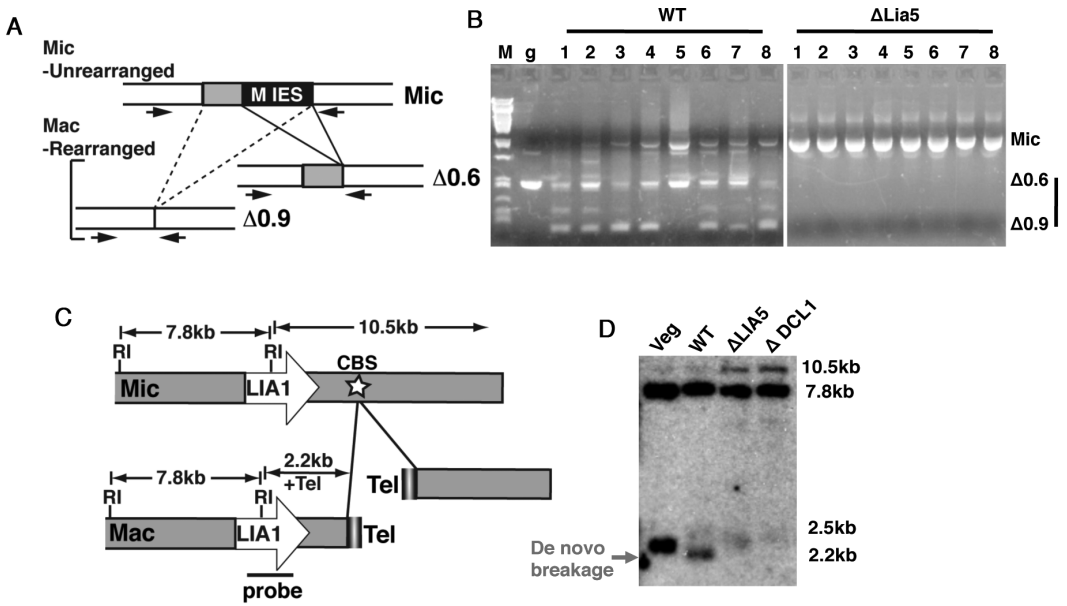

Figure 3 – Lia5 is required for Tetrahymena programmed DNA rearrangement

(A) Schematic of the strategy for PCR based IES excision assay. Arrows denote forward and reverse primers used to amplify across the M IES. Alternative rearrangement products resulting from deletion of 0.6kp (Δ0.6) or 0.9kp (Δ0.9) are shown. (B) M IES excision PCR assay. M, DNA ladder. g, genomic DNA from unmated cells. Each lane (1-8) represents a single mated cell from WT or ΔLia5. Unrearranged (Mic) and the two alternatively rearranged products (Δ0.6 and Δ0.9) are as indicated. (C) Diagram showing the chromosomal scaffold surrounding a chromosomal breakage sequence (CBS) downstream of the Lia1 gene. Relevant EcoRI (RI) restriction sites used for the Southern blot analysis are shown. The probe spans the central EcoRI site and detects the 10.5kb micronucleus-specific fragment (unrearranged) and *de novo* breakage at 2.2kb. The probe also detects a 7.8 kb fragment common to both nuclei as well the 2.5kb to 2.6kb fragments that harbor fully extended telomeric DNA, found in the macronuclei of unmated cells. (D) Southern blot analysis to assess chromosome breakage. Arrow indicates *de novo* breakage. veg, vegetative cells, used to represent unmated cells in the population.

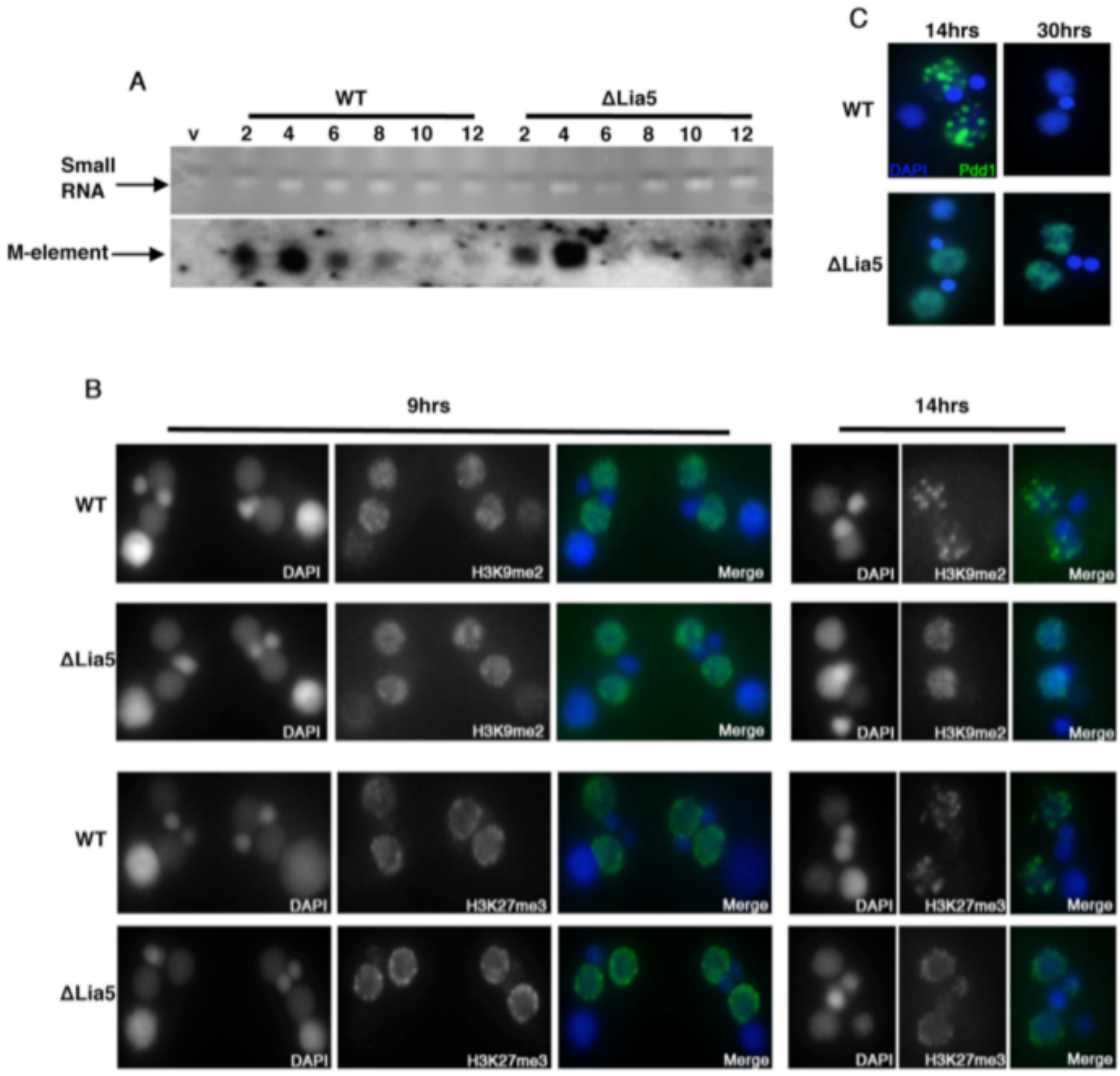

Figure 4 – Lack of Lia5 causes failure in IES excision foci formation while critical events prior to the step remain undisturbed

(A) Upper panel: ethidium bromide staining of total RNA (20 μg per lane) extracted from unmated (v) or conjugating WT and ΔLia5 cells at indicated time points (2-12hrs). Lower panel: Northern blot analysis for M IES. (B) Immunoflouresence staining of H3K27me3 and H3K9me2 histone marks in WT and ΔLia5 at 9hrs and 14hrs into conjugation. (C) Fluorescent images of Pdd1-YFP localization in the developing macronuclei of WT and ΔLia5 conjugating cells, counter stained with DAPI. 30hrs time point represents cells that have finished mating and completed development or reached their final arrest point.

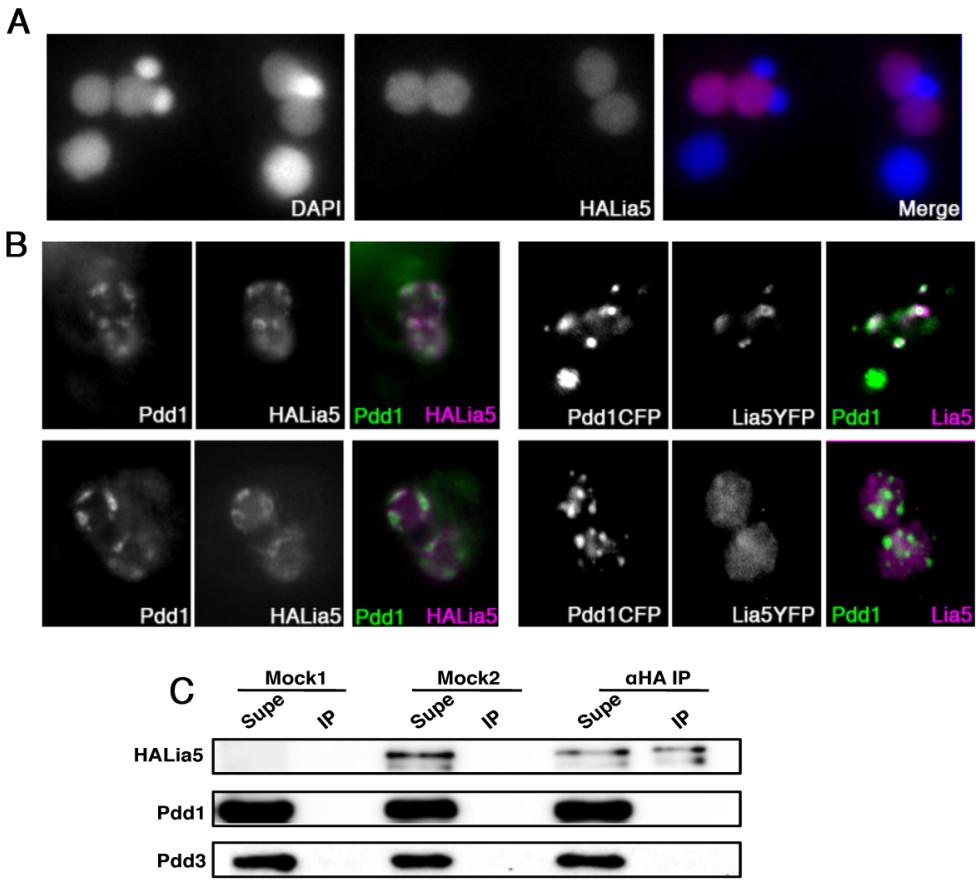

Figure 5 – Lia5 localization at the IES excision foci
(A) Immunostaining of HALia5 at 9hrs into conjugation. (B) Co-localization of Lia5-YFP and Pdd1-CFP or immuno-stained HALia5 and Pdd1-YFP in 14hrs conjugating cells. (C) Immunoprecipitation of HALia5 (αHAIP). Untransformed cells (Mock1) and IgG only pull-down (Mock2) are used as controls. Immunoprecipitated samples (IP) and their respective supernatants (Supe) were blotted for HALia5, Pdd1 and Pdd3 as indicated.

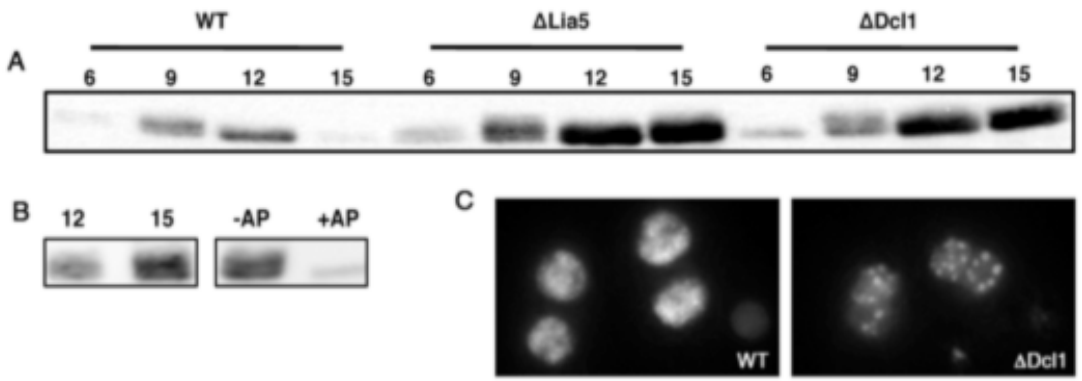

Figure 6 – Pdd1 protein phospohrylation profile in ΔLia5 during conjugation

(A) αPdd1 western blot analysis of total protein isolated from conjugating WT, ΔLia5 and ΔDcl1 cells at the indicated time points. (B) Left panel: αPdd1 western blot analysis of diluted protein from 12 and 15hrs conjugating ΔLia5 cells. Right panel: Alkaline phosphatase treatement (+AP) of ΔLia5 protein at 15hr time point. (C) Fluorescent images of premature Pdd1-YFP foci in ΔDcl1.

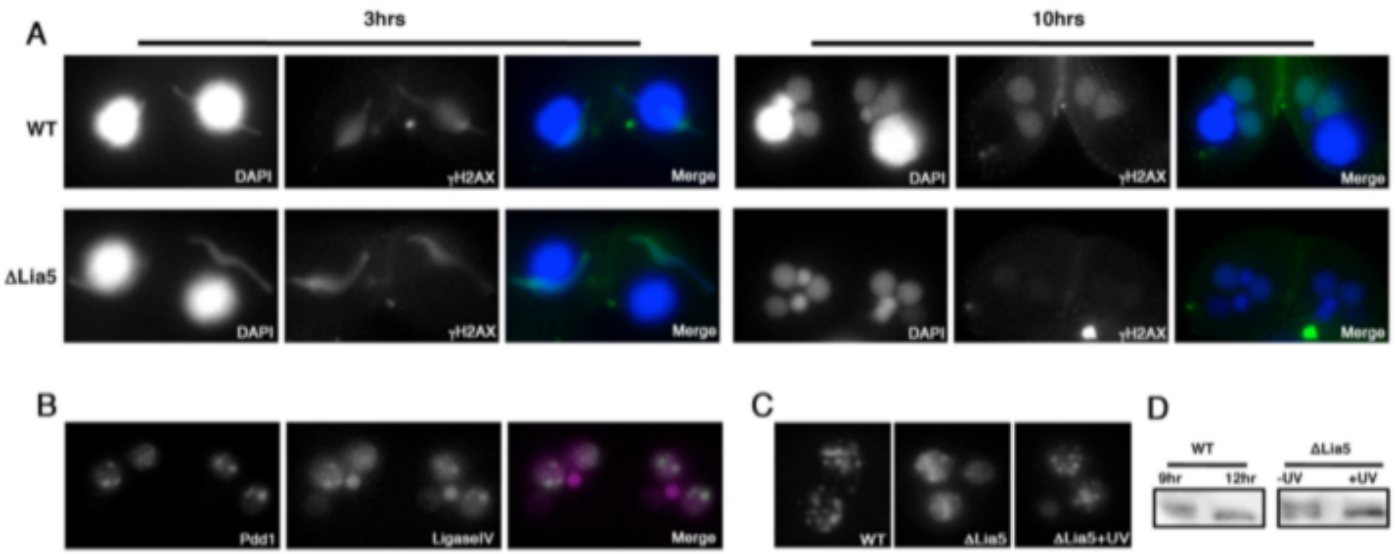

Figure 7 – DNA damage in ΔLia5

(A) Immunostaining of γH2AX in WT and ΔLia5 cells during meiotic stage (3hrs) and macronuclear differentiation (10hrs) stages of conjugation. (B) Fluorescence images of Pdd1-CFP and LiagseIV-YFP co-localization in response to UV. (C-D) UV induced DNA damage is sufficient to rescue Pdd1 localization and phosphorylation phenotype in ΔLia5. (C) Pdd1-YFP localization in 14hrs conjugating WT, ΔLia5 and ΔLia5 cells treated with 150mJ of UV. (D) Left panel: αPdd1 western blot analysis for WT mating at 9hrs and 12hrs. Right panel: Irradiated (+UV) or control (-UV) ΔLia5 mating cells at 12hrs

Chapter 4

General conclusion and future perspective

Genome reprogramming during cellular differentiation involves re-patterning of chromatin through manipulation of chromatin-associated proteins and changes in nuclear architecture to properly silence or activate appropriate genes. During sexual reproduction, *Tetrahymena thermophila* undergoes extensive DNA rearrangement to reprogram the future somatic genome of its progeny. Such process involves the removal of ~50Mbp of the germline-derived DNA from an estimated 5000-6000 loci, which represent >30% of the original germline genome (Yao and Gorovsky, 1974; Fass et al 2011; reviewed in Yao and Chao, 2005; Chalker and Yao, 2011). The sequences eliminated, termed Internal Eliminated Sequences (IESs), range in size from 300bp to >20kbp and are predominantly found in intergenic regions. Efficient and accurate removal of these sequences is crucial for the survival of sexual progeny. This notion is especially prominent in the extreme cases of DNA rearrangement in ciliate subclass *Stichotrichia*, whom eliminates 95-98% of the germline and need to eventually piece together and unscramble their exons into protein coding genes (reviewed in Nowacki et al., 2011). However, the means in which the cells recognize and target the correct sequences for removal remained a mystery for a long time. The reasons being that the removed sequences are vastly diverse and share little to no similarity. The relatively recent discovery that recognition of these sequences is mediated by homologous RNAs has greatly contributed to uncovering the mystery.

Prior to this dissertation, the involvement of small RNAs had been described in *Tetrahymena* programmed DNA rearrangement (Mochizuki et al., 2002; Malone et al., 2005; Mochizuki, 2005). Furthermore, the link to heterochromatin and transcriptional gene silencing had been implicated from the identification of essential PIWI-related and chromodomain proteins (Madireddi et al., 1996; Nikiforov et al., 2000; Mochizuki et al.,

2002). Together with the rapid advances made in other ciliate species and organisms, at the time, the idea that IES removal involves small RNA-directed heterochromatin formation has been well described. The events that follow, including the repositioning of IESs into heterochromatin-like sub-nuclear domains and their subsequent removal, have been largely unexplored. In fact when I joined the lab, only one protein (Lia1p) had been shown to act late enough during conjugation to function downstream of small RNA mediated heterochromatin targeting (Rexer and Chalker, 2007). I had been particularly interested in studying the process of extensive nuclear reorganization that partitions IESs and associated chromatin into distinct sub-nuclear foci since this may provide insights into how activities are compartmentalized in the nucleus. To approach this question, we investigated proteins that function late in conjugation that may play a role in regulating chromatin repositioning. I focused my studies on two promising candidates – Defective IES Excision 5 (*DIE5*) and Localized In macrouclear Anlagen 5 (*LIA5*), both encode novel proteins that are essential for programmed DNA rearrangement in *Tetrahymena*. Interestingly, through studying these two genes, we revealed connections between heterochromatin formation and a programmed DNA damage response that accompanies IES excision during *Tetrahymena* somatic nuclear differentiation.

ΔDIE5 and *ΔLIA5* reveal the importance of programmed DNA damage and repair in IES excision during *Tetrahymena* somatic nuclear differentiation

DIE5 and genome integrity

Both *DIE5* and *LIA5* were identified from some of the first efforts to screen for genes that may play a role in the later stages of ciliate programmed DNA rearrangement

(downstream of small RNA and heterochromatin targeting). The screen conducted by Matsuda and Forney at Purdue University led to the identification of the *Paramecium* protein Die5p, which was found to be conserved in *Tetrahymena*. In both of these evolutionarily distant species, *DIE5* is required for programmed DNA rearrangement and completion of conjugation. Studies done in *Tetrahymena* showed that although Δ*DIE5* failed to excise IESs, Pdd1p nuclear foci were detected. This phenotype was novel since mutations in other known genes that are required for IES excision fail to form proper foci. *DIE5* therefore uncouples the events of nuclear reorganization and IES excision, and may in fact function downstream of chromatin repositioning. This observation placed *DIE5* as the latest known acting gene in the process of *Tetrahymena* IES elimination.

Another phenotype that was unique to Δ*DIE5* strains at the time was that while they arrest in development at a stage common to other IES mutants studied, they eventually lose all detectable DNA content from their developing somatic macronucleus. Recent findings in the field (since the publication of our *DIE5* paper) may provide some explanation for these observations. These advances provide new understanding of the excision events during DNA rearrangement. In both *Paramecium* and *Tetrahymena*, it has been shown that domesticated *piggyBac* transposase is most likely responsible for the actual excision event of the IESs (Baudry et al., 2009; Cheng et al., 2010), consequently, inducing a programmed DNA damage response. Major components of the NHEJ pathway have been implicated in both species to subsequently repair the broken ends (Kapusta et al., 2011; Lin et al., 2012). Particularly, disruption of *Tetrahymena* Ku80 (Tku80) results in a phenotype similar to Δ*DIE5* where the developing nucleus loses its DNA content (Lin et al., 2012). Thus, although the exact biochemical function of Die5p remains to be

determined, we reason that it may be involved in a conserved process (at least between *Tetrahymena* and *Paramecium*) that is essential for maintaining genome stability during programmed DNA damage and repair.

Governing genome integrity is crucial during ciliate DNA rearrangement as it is accompanied by genome-wide induction of DNA lesions. Current research has begun to shed light into the pathways that are involved in repair. Studying Die5p function and the exact role of NHEJ pathway would provide further understanding of how cells deal with such extensive insults to its genome, especially at sites enriched in repetitive elements that are prone to improper recombination between distal homologous sequences. As Die5p does not contain conserved protein domains that identify it as a known component of the NHEJ pathway; it remains unclear whether it plays a direct role in DNA damage repair. To further investigate the role of *DIE5* in ensuring genomic integrity, it would be informative to examine the possible functional and biochemical interactions between *DIE5* and known NHEJ repair pathway components. The results of these experiments would provide mechanistic insights into the final stages of ciliate genome reprogramming

LIA5 – links to transposons, heterochromatin and DNA repair

LIA5 was identified in *Tetrahymena* from the same screen that uncovered *LIA1* (Rexer and Chalker, 2007; Yao et al., 2007). I was especially interested in *LIA5* because it was predicted to encode a protein that contains a PHD/Zn finger domain, which has been implicated in many cellular processes, among them, heterochromatin formation and DNA-protein interaction. We therefore hypothesized that Lia5p would function as a structural scaffold of the nuclear foci to bridge the interaction of essential components.

Further studies and the most current annotation of the protein's conserved domains revealed that *LIA5* function may be more closely related to that of *DIE5* than we previously envisioned. Lia5p exhibits protein architecture that is common to transposon-derived proteins, including the *piggyBac* transposases. As domesticated *piggyBac* transposase Tpb2 has been implicated as the excisase that 'cuts' the IESs out of the genome (Cheng et al., 2010), we suspected that *LIA5* functions in a similar process. Indeed, we showed that *LIA5* knockouts lack detectable γH2AX in the developing nucleus, suggesting that Δ*LIA5* fail to generate DNA damage signals during IES excision. Furthermore, we showed that in Δ*LIA5*, phosphorylation of the essential chromodomain protein – Pdd1p is mis-regulated. In wild type mating, Pdd1p is phosphorylated earlier on. As development progresses, however, coinciding with the event of nuclear foci formation and IES excision, Pdd1p becomes dephosphorylated. We found that phosphorylated Pdd1 persists in Δ*LIA5*. Furthermore, introducing DNA damage by treating cells with UV is sufficient to cause Pdd1 dephosphorylation and foci formation. Importantly, in response to UV treatment, Pdd1 is recruited to nuclear domains where the NHEJ repair protein – LigaseIV aggregates. We therefore infer that Pdd1p acts as a component of DNA damage response (DDR) and that such response is important for triggering the assembly of nuclear foci. It would be informative to further investigate the role of Pdd1p in DNA damage repair and how this might influence the process of nuclear reorganization.

Through studying *LIA5*, we have provided a link between the heterochromatin-like IES excision foci and the DNA repair foci. This realization may influence our future approaches to study these foci. Furthermore, since the formation of these foci (therefore

proper excision of IESs and repair of the broken ends) require *LIA5*, which likely originated from transposons introduced into *Tetrahymena* genome during evolution, further studying the mechanistic function of *LIA5* would provide insights to how transposons helped shape eukaryotic genomes.

Taken together, the results from my thesis projects suggest that *LIA5* and *DIE5* are involved in the generation and protection of program-induced DNA damage during IES excision, respectively. Although the later stages of ciliate conjugation was once an unexplored territory, our findings and the field's progression has generated an increasing list of essential players involved. From these efforts, we are collectively discovering other conserved pathways that are involved in ciliate programmed DNA rearrangement. Further progress would advance our understanding in the relationship between these pathways.

Nuclear reorganization

Studying nuclear reorganization during *Tetrahymena* DNA rearrangement has been difficult, partly due to difficulties in isolating protein complexes that would identify components of the DNA rearrangement foci. The results of my thesis projects suggest that the subnuclear structures formed during the elimination of IESs may be related to DNA repair foci. Although the question of how cells redistribute its nuclear components to compartmentalize activities remain unanswered, our findings provide a handle for studying the identity and possible function of the foci associated with *Tetrahymena* DNA rearrangement.

As more is understood about the stepwise assembly of the DNA repair foci and the proteins involved, their resemblance allows us to better predict the possible components that may be involved in the formation of DNA rearrangement foci. For instance, as Ku80 has been shown to be important, further experiments to test the role of other NHEJ components in foci formation would not only shed light into the process, but also provide understanding to the extent of similarity between NHEJ repair pathway and IES excision repair.

Understanding the link between small RNA-induced heterochromatin formation and DNA damage repair

Perhaps one of the most interesting questions that arises from the studies of *Tetrahymena* DNA rearrangements aims to decipher the interplay between small RNA, heterochromatin and DNA damage repair. Studies of *LIA5* suggested that in *Tetrahymena*, nuclear reorganization and the formation of heterochromatin-like foci is dependent on program- induced DNA damage. Importantly, one of the components of heterochromatin – Pdd1 protein, appears to participate as a sensor for DNA damage. Similarly, a link between heterochromatin components and DNA damage repair has been implicated in other organisms. It has been shown that mammalian Heterochromatin Protein 1 (HP1) responds to chromosome breaks and is required for proper repair (Ayoub et al., 2008, 2009; Luijsterburg et al., 2009). Furthermore, while the role of small RNAs in directing heterochromatin and gene silencing has been extensively discussed, recently, a novel role of small RNA biogensis pathway in DNA damage repair has been revealed in Arabidopsis, mouse, zebrafish and human (Francia et al., 2012; Wei et al., 2012).

These studies showed that when DNA lesions occur, damage-site specific small RNAs are produced, and that RNAi pathway components are required for proper repair. It remains unclear, however, whether small RNA initiated heterochromatin formation is an intermediate to proper repair, and whether the interactions between these pathways are important for repair across different genomic loci (euchromatin vs. heterochromatin).

Small RNAs have been known to associate with ciliate programmed DNA rearrangement. So far, the study of the role of these small RNAs have been limited to the targeting of IESs. However, at least in *Tetrahymena*, the bulk of the small RNAs accumulate beyond the stage where IESs are identified. Likewise, the expression of Dicer-like enzyme (*DCL1*) persists into developmental stages that are much later than the event of IES identification. In fact, *DCL1* expression peaks just prior to the formation of nuclear foci and IES excision – the stage where genome-wide DNA damage occurs. The roles of small RNAs in these late stages of ciliate development have been largely unexplored.

Tetrahyman programmed DNA rearrangement undergoes a genome-wide, step-wise process of small RNA-directed heterochromatin assembly at specific loci, followed by their excision and the repair of surrounding sequences. This provides a unique opportunity to study the functional relationships between small RNA, heterochromatin and DNA repair. A heterochromatin component of the process (Pdd1p) has already been implicated in DNA repair. Further investigating the mechanistic details, in addition to studying the role of the developmental specific small RNAs and *DCL1* in repair, would further our understandings in the functional interconnections between these different pathways.

References

Ayoub, N., Jeyasekharan, A.D., Bernal, J.A., and Venkitaraman, A.R. (2008). HP1-β mobilization promotes chromatin changes that initiate the DNA damage response. Nature *453*, 682–686.

Ayoub, N., Jeyasekharan, A.D., and Venkitaraman, A.R. (2009). Mobilization and recruitment of HP1: a bimodal response to DNA breakage. Cell Cycle *8*, 2945–2950.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., and Betermier, M. (2009). PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes & Development *23*, 2478–2483.

Cheng, C.Y., Vogt, A., Mochizuki, K., and Yao, M.C. (2010). A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Molecular Biology of the Cell *21*, 1753–1762.

Francia, S., Michelini, F., Saxena, A., Tang, D., Hoon, M. de, Anelli, V., Mione, M., Carninci, P., and Fagagna, F. d'Adda di (2012). Site-specific DICER and DROSHA RNA products control the DNA-damage response. Nature.

Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., and Bétermier, M. (2011). Highly precise and developmentally programmed genome assembly in Paramecium requires ligase IV-dependent end joining. PLoS Genet. *7*, e1002049.

Lin, I.-T., Chao, J.-L., and Yao, M.-C. (2012). An essential role for the DNA breakage-repair protein Ku80 in programmed DNA rearrangements in Tetrahymena thermophila. Mol. Biol. Cell *23*, 2213–2225.

Luijsterburg, M.S., Dinant, C., Lans, H., Stap, J., Wiernasz, E., Lagerwerf, S., Warmerdam, D.O., Lindh, M., Brink, M.C., Dobrucki, J.W., et al. (2009). Heterochromatin protein 1 is recruited to various types of DNA damage. The Journal of Cell Biology *185*, 577–586.

Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., and Allis, C.D. (1996). Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. Cell *87*, 75–84.

Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., and Chalker, D.L. (2005). Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of Tetrahymena thermophila. Mol. Cell. Biol. *25*, 9151–9164.

Mochizuki, K. (2005). A Dicer-like protein in Tetrahymena has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes & Development *19*, 77–89.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Nikiforov, M.A., Gorovsky, M.A., and Allis, C.D. (2000). A Novel Chromodomain Protein, Pdd3p, Associates with Internal Eliminated Sequences during Macronuclear Development inTetrahymena thermophila. Molecular and Cellular Biology *20*, 4128–4134.

Nowacki, M., Shetty, K., and Landweber, L.F. (2011). RNA-Mediated Epigenetic Programming of Genome Rearrangements. Annual Review of Genomics and Human Genetics *12*, 367–389.

Rexer, C.H., and Chalker, D.L. (2007). Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in Tetrahymena thermophila. Eukaryotic Cell *6*, 1320–1329.

Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.-G., and Qi, Y. (2012). A Role for Small RNAs in DNA Double-Strand Break Repair. Cell *149*, 101–112.

Yao, M.-C., Yao, C.-H., Halasz, L.M., Fuller, P., Rexer, C.H., Wang, S.H., Jain, R., Coyne, R.S., and Chalker, D.L. (2007). Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in Tetrahymena. J. Cell. Sci. *120*, 1978–1989.

Appendix 1

Epigenetics of Ciliates

Motl, J.A., Shieh, A.W., and Chalker, D.L.

In Encyclopedia of Molecular Cell Biology and Molecular Medicine (2012)

Wiley-VCH Verlag GmbH & Co. KGaA

Contributions to the work

This comprehensive review was written by Jason Motl and myself, with content and editorial input from D.L. Chalker. As a coauthor, I contributed most of the writing of the first two sections ('Ciliate Biology' and 'Epigenetic Phenomena in Ciliates') of this review and provide feedback on the remainder of the article. It was published in Encyclopedia of Molecular Cell Biology and Molecular Medicine

References

Motl, J.A., Shieh, A.W., and Chalker, D.L. (2012) Epigenetics of Ciliates. In Encyclopedia of Molecular Cell Biology and Molecular Medicine, (Wiley-VCH Verlag GmbH & Co. KGaA)

Epigenetics of Ciliates

Jason A. Motl, Annie W. Shieh, and Douglas L. Chalker
Washington University in St Louis, Biology Department, 1 Brookings Drive, St Louis, Missouri, USA

Meyers: Encyclopedia of Molecular Cell Biology and Molecular Medicine: Epigenetic Regulation and Epigenomics, Second Edition. Edited by Robert A. Meyers.

Keywords

Nuclear dimorphism
Containing two different types of nuclei, like ciliates.

Micronucleus (Mic)
The smaller, germline nucleus of nuclear dimorphic ciliates.

Macronucleus (Mac)
The larger, somatic nucleus of nuclear dimorphic ciliates.

Conjugation
Sexual reproduction process of ciliates that involves cross-fertilization and genetic exchange between mating partners to produce progeny.

Autogamy
Sexual reproduction process of ciliates during which one individual self-fertilizes to produce progeny with a completely homozygous genome.

Internal eliminated sequences (IESs)

Sequences ranging from 26 bp to 22 kb, which necessitate removal from introns, exons, and noncoding DNA sequences during sexual reproduction to produce a functional zygotic macronucleus.

DNA elimination

Process of removing repetitive sequences and IESs from the somatic, zygotic macronucleus in ciliates, which occurs during sexual reproduction.

Transposons

DNA elements that can v ''jump'' or transpose around the genome when active.

RNA interference

Process through which ncRNAs and the products of their cleavage, sRNAs, affect transcriptional and post-transcriptional regulation in cells.

Heterochromatin

Chromatin state defined molecularly by histone hypoacetylation and methylation of H3K9 and/or H3K27, which causes condensation of chromatin and gene silencing.

Genetic studies of ciliated protozoa delivered some of the earliest evidence that epigenetic mechanisms play profound roles in determining phenotype. The nuclear dimorphism of these unconventional unicellular organisms has provided a rich context within which to uncover epigenetic mechanisms that regulate genome activities. Comparisons of the chromatin of the transcriptionally active somatic genome and the silent germline have revealed that histone modifications and specialized variants are important regulatory mechanisms, allowing homologous sequences to exist in different states. However, these genomes do not just differ in epigenetic characteristics; they have major structural differences, the result of developmentally programmed DNA rearrangements that occur during nuclear differentiation. These rearrangements eliminate between 15% and 95% of a ciliate's germline-derived DNA to create a streamlined genome that is devoid of most repetitive elements. More recent investigations have revealed that homologous noncoding RNAs (ncRNAs) and RNA interference mechanisms play essential roles in guiding these DNA rearrangements by mediating a comparison of the genome content of the current somatic genome to that in the germline. Continuing research into the process of DNA elimination in ciliates shows promise to provide new insights into the potential of ncRNAs to remodel genomes during development.

1 Ciliate Biology

1.1 Historical Perspective

The concept of ''epigenetics'' was largely formulated by Conrad Waddington to provide a framework to describe the development of multicellular organisms, and to explain how cells with the same genetic composition can differentiate into functionally distinct types. During these early days of genetics research, the chromosome theory of inheritance was viewed to bridge the observations of Mendelian inheritance and microscopic description of chromosome behavior in cells. However, this genetic theory was somewhat inadequate to account for development of different tissues within an individual, where all cells had the same chromosomes. It was difficult to envision how the apparently static chromosomes (genes) could by themselves manifest phenotypic differences – the fundamental basis of cellular differentiation.

Epigenetic theory thus arose from the need to bridge the gap between genotype and observed phenotypes that could not be accounted for by the behavior of chromosomes. The gap was quite apparent in single-celled organisms, most notably the ciliate *Paramecium* in studies by Tracy Sonneborn [1]. Sonneborn and his colleagues described several examples of phenotypic traits – for example, serotype and mating type – which did not follow conventional Mendelian inheritance, but instead appeared to be passed on through cytoplasmic inheritance. Thus, while these traits were encoded by genes, clonal lines with identical genotypes arose with persistently different phenotypes. Through these studies, Sonneborn and others revealed that the cytoplasm was an important supplement to chromosomes in transmitting heritable information.

While studies of ciliate genetics largely started with those of Sonneborn, research using these organisms has continued to provide important understanding of epigenetic phenomena and their underlying mechanisms that help explain unexpected patterns of inheritance. In this chapter, some early examples of non-Mendelian inheritance observed in ciliates are described to provide a historical context, even though the exact mechanisms that account for these phenomena still await discovery. Nonetheless, research efforts aimed at describing the intricate biology of this fascinating group of microbes have provided new ways to consider epigenetics that stretch well beyond ciliates. Fundamental discoveries of the role of chromatin modification in gene regulation, and the role of noncoding RNAs (ncRNAs) in gene silencing, have secured the place of ciliates as pioneering model systems for epigenetic studies. Much of the utility of these organisms for this research stems from their unique biology, with both germline and somatic copies of the genome maintained in a single cell. Below, the germline and somatic dichotomy of ciliates are describe, followed by details of the process of their differentiation, in order to provide the necessary background for describing these epigenetic discoveries.

1.2 Life Cycle and Genetics

The ciliated protozoa belong to the superphylum of Alveolates, which is a lineage that diverged from the ancestors of plants and animals more than a billion years ago [2]. They have evolved into a diverse array of species that have adapted to different environments and strategies for

life. Members of the phylum Ciliophora (i.e., ciliates) are commonly found in fresh water, but can also exist in many water-rich environments as free-living organisms, symbiotes, or even parasites. Ciliates have elaborate cellular architectures, most noticeably the organized arrays of cilia that cover their exteriors. An anterior oral apparatus or ''mouth,'' constant swimming enabled by their cilia, and relatively large size give the ciliates animal-like qualities, despite their being unicellular. Ciliates are capable of both asexual and sexual reproduction:

- *Asexual reproduction* (or vegetative population growth) occurs by binary fission, and is the means through which ciliates amplify their populations clonally (Fig. 1a).
- *Sexual reproduction* occurs upon the conjugation of two cells, and involves the exchange of genetic information between each partner and new somatic genome differentiation, without an increase in cell number (Fig. 1b).

The most important feature of ciliates to consider in regards to inheritance is their nuclear dimorphism. In each single cell, ciliates organize two copies of their genome in nuclei that are structurally and functionally distinct. These two different genomes serve the analogous roles to that of germline and somatic cells in metazoans. The germline copy of the genome is contained in the smaller nuclear compartment, the *micronucleus*. The micronuclei are diploid, but interestingly are transcriptionally silent during vegetative growth, serving only to maintain and transmit the genome to progeny cells upon sexual reproduction. The much larger *macronuclei*, on the other hand, carry the somatic genome and, as such, are responsible for all gene expression necessary for vegetative growth. Macronuclei are polyploid, with different ciliate species having widely different copy numbers in their somatic genomes. For example, *Tetrahymena* retain approximately 50 copies of each macronuclear chromosome, whereas *Paramecium* macronuclei contain several hundred copies. During sexual reproduction, the macronucleus – like the soma of metazoa – is lost when a new one is formed from a zygotic nucleus, which is derived from the germline genomes of the parental cells after meiosis.

While all ciliates exhibit nuclear dimorphism, the actual number of germline micronuclei and somatic macronuclei in each cell differs between species. In many of the figures in the chapter, it has been elected to illustrate a single micronucleus and macronucleus per cell, to simplify the discussion. The key nuclear events that occur throughout the ciliate life cycle are presented in a generalized representation in Fig. 1. Vegetative growth involves clonal amplification of the cell's population, during which the micronucleus is duplicated by closed mitosis (i.e., without the dissociation of nuclear envelope), thus ensuring an accurate maintenance of the germline genome (Fig. 1a). The polyploid macronucleus divides amitotically, splitting its nuclear content into roughly equal halves so as to partition its centromere-less chromosomes into each progeny cell. Exactly how the macronuclei maintain the correct copy number of somatic chromosomes is not well understood, but the results of studies conducted in *Oxytricha* and *Stylonychia* have indicated that the copy number can be regulated epigenetically [3, 4]. Nevertheless, high ploidy and the endoreplication of somatic chromosomes appear to maintain the correct DNA content and prevent lethal gene loss.

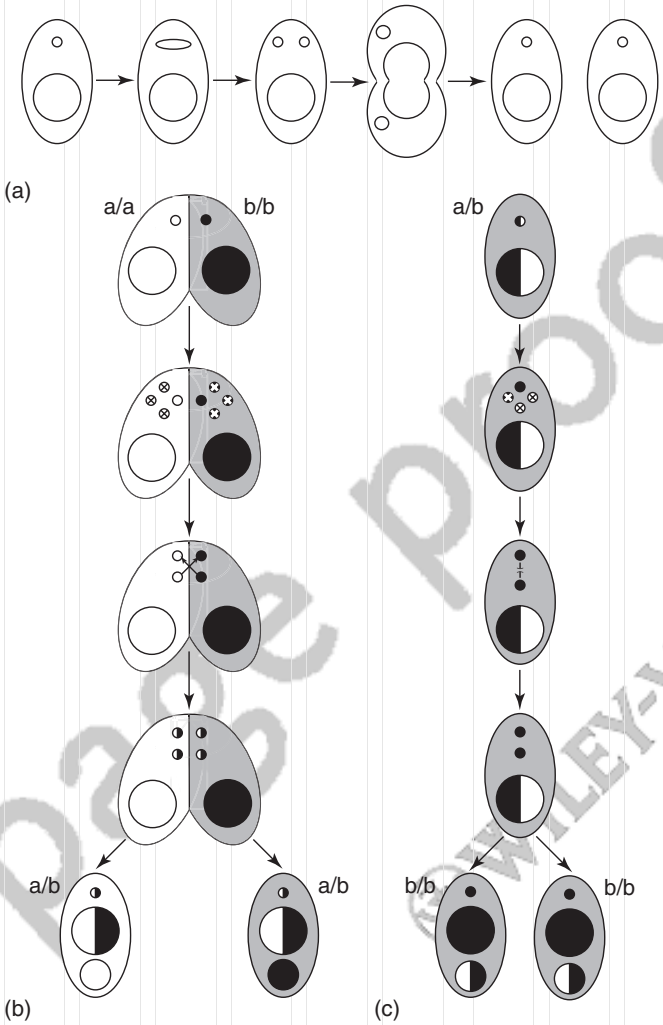

Fig. 1 Ciliates reproduce both vegetatively by binary fission, and sexually via conjugation (or autogamy).

The micronucleus and macronucleus of a ciliate are replaced after each round of sexual reproduction. *Conjugation*, which can be induced in the laboratory by nutrient starvation, begins with the pairing of two mating compatible cells (see Fig. 1b). The micronuclei of the mating partners undergo meiosis, where a single haploid meiotic product in each partner is selected and passed on to their progeny; the nonselected meiotic products are then degraded. The chosen haploid nuclei replicate their chromosomes and then undergo an additional nuclear division to produce two haploid nuclei with identical genomes, one of which is exchanged with the mating partner. The exchanged haploid nucleus then fuses with the partner's stationary haploid nucleus to form the zygotic nuclei of the mating pair.

This nuclear cross-fertilization produces identical heterozygous, diploid genomes in each partner. In the case where mating-compatible partners are unavailable, some species will undergo *autogamy*; this is a form of self-fertilization, where two genetically identical haploid nuclei fuse with each other, producing a homozygous diploid genome (Fig. 1c).

When the haploid "gametic" nuclei have fused (*karyogamy*) to give rise to the zygotic genome, additional rounds of DNA replication and nuclear division produce the precursors of the new micronucleus and macronucleus. As the development proceeds, these progenitors (which often are called *anlagen*) differentiate into the new germline and somatic nuclei. Whereas, the cross-fertilization that occurs during conjugation generates genetically identical progenitor nuclei, the individual progeny cells of a mating pair can differentiate with distinct phenotypes (e.g., different mating types) in non-Mendelian inheritance patterns. In some cases, specific phenotypes can be traced through a particular cytoplasmic lineage. It is important to note that new somatic nuclei differentiate within the cytoplasms of the two parental cells, such that the DNA is replaced while many existing cellular structures are preserved. This feature of ciliate biology is a major contributor to the non-Mendelian inheritance phenomena described in the following sections.

1.3 Differentiation of Somatic and Germline Genomes

Macronuclear differentiation is an extreme example of genome reprogramming, as the cells start with a genome that is transcriptionally silent and remodel it into one that supports regulated gene expression during vegetative growth. In addition to switching the genome from a silent to an active state, this reprogramming involves a transition from mitotic to amitotic division, accompanied by chromosome breakage and extensive DNA rearrangements (Fig. 2). Research efforts to understand the differences between the transcriptional activity of micro- and macronuclei have uncovered regulatory systems that have solidified the ciliates'' place as major models for elucidating epigenetic mechanisms. Before discussing these discoveries further, it is important to briefly touch upon the structural rearrangements

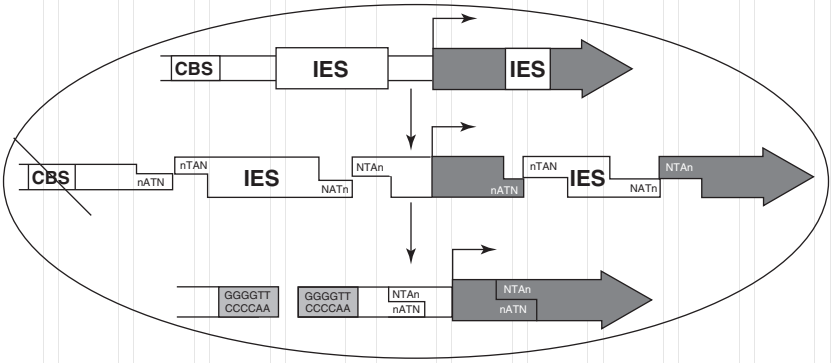

Fig. 2 The somatic genome undergoes extensive DNA rearrangements, including chromosome fragmentation and internal DNA elimination.

that streamline and selectively amplify the genome in differentiating macronuclei. These are important, not only for providing an understanding of some of the historical examples of non-Mendelian inheritance, but also providing – through their study – new avenues by which epigenetic regulation can be further explored.

Ciliates streamline their somatic, macronuclear genome through massive genome rearrangements that fragment the germline-derived chromosomes and eliminate large portions of their genomic complexity (Fig. 3). The fraction of the germline genome removed from the macronucleus ranges from 15% to as much as 95% (for a review, see Ref. [5]). Fragmentation of the developing macronuclear chromosomes is coupled with *de novo* telomere addition, which stabilizes the newly formed termini. The degree of chromosome fragmentation varies widely among the ciliate lineages. For Stichotrichs such as *Oxytricha* and *Euplotes*, this fragmentation is so extensive that the average macronuclear chromosome is only a few kilobase pairs that typically contains a single gene. At the other end of the spectrum, the Oligohymenophora, which include *Tetrahymena* and *Paramecium*, break their developing macronuclear chromosomes at just tens to hundreds of sites to produce chromosomes that are typically several hundred kilobase pairs in size. Following chromosome fragmentation, these small chromosomes are amplified to their final high copy number in the polyploidy macronucleus.

In addition to chromosome fragmentation, ciliates eliminate many DNA segments from internal sites. These germline-limited, internal eliminated sequences (IESs) are numerous in all ciliate genomes that have been studied. They are removed from thousands of loci and, in some species, from up to tens of thousands of loci. In some ciliates, such as *Tetrahymena*, essentially all of the IESs are found within intergenic regions, whereas in most other ciliates studied, they are also common within genes. When IESs are present in coding regions, they are precisely excised during macronuclear differentiation. A common class of IESs found in diverse ciliates species is characterized by flanking 5′-TA-3′ dinucleotides, one copy of which is retained upon excision (for a review, see Ref. [6]). The sequences eliminated from somatic macronuclei represent most of the repetitive sequences residing in the germline genome, including transposable elements [7]. The majority of IESs may actually be the remnants of transposons, or be otherwise derived from the activity of transposable elements ([8–10]; see also a review in Ref. [11]). Intriguingly, recent evidence has suggested that the excision of IESs utilizes domesticated transposases [12–14]. As will be discussed

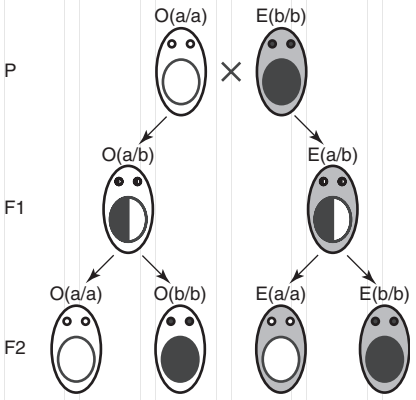

Fig. 3 Cytoplasmic inheritance. The mating type of *Paramecium* is determined by the parental cytoplasm, not the genotype of the progeny. P, Parent; F1/F2, generations.

Tab. 1 Histone modifications found in the nuclei of ciliates.

Nucleus	*Histone composition*	*Histone modifications*[a]
Micronucleus	H2A, H2B, H3, H4, micH1	H3K27me, H3S10ph, micH1ph
Macronucleus	H2A, H2B, H3, H4, hv1, hv2, macH1	H3K4me, H3K9me[b], H3K27me, H2Aac, H2Bac, H3ac, H4ac, H2Aph, macH1ph

The histone composition and modifications of the micro- and macronucleus are listed above. Most of these histones and modifications are found throughout the life cycle of *T. thermophila*, but one (H3K9me) is restricted to developing macronuclei during conjugation.
[a]There is no distinction between mono-, di-, and tri-modifications of each histone in this table.
[b]This modification is only found during sexual reproduction in the developing zygotic macronucleus.

below, the mechanisms that ciliates use to identify IESs are related to RNA interference (RNAi), which is used by many eukaryotes as a surveillance system to limit the activity of transposons in the genome [15]. These mechanisms will be described in detail at this point, as they reveal important insights into the use of homologous, ncRNAs in epigenetic regulation.

1.4 Micro- and Macronuclei: Models for Silent and Active Chromatin

The recognition that the micro- and macronuclei of ciliates have opposite activity states promoted the development of these organisms as models with which to examine cellular mechanisms that differentially regulate identical sequences – the very definition of epigenetics. The most significant contributions in this area have been made by groups investigating the chromatin structure of the different nuclei of *Tetrahymena*. Such efforts began about four decades ago, and helped to establish a number of paradigms of epigenetic control, including the importance of histone variants and the role of histone acetylation in transcriptional regulation. A summary of the histone variants and modifications found in the micro- and macronucleus is listed in Table 1.

1.4.1 Differential Histone Composition of Micro- and Macronuclei

The core histones form the largest fraction of chromatin in both the micro- and macronuclei; however, a comparison of the chromatin proteins found in each type of nucleus led to the characterization of some of the first known histone variants. The histone variants, Hv1 and Hv2, were identified as forms of Histone H2A and H3, respectively, that are localized specifically within the transcriptionally active macronucleus [16, 17]; these proteins represent the equivalent of the widely conserved variants H2A.Z and H3.3, while Hv1 (H2A.Z) is essential in *Tetrahymena* [18]. In addition to its presence in the macronucleus, this variant has been observed in micronuclei during early conjugation, when these nuclei first exhibit transcriptional activity [19–21]. Hv2 (H3.3) has properties consistent with its role as a replacement histone. This variant was shown to be constitutively expressed during the cell cycle, in contrast to core histone H3.1, which is expressed only during early

S-phase [22]. This led to the hypothesis that H3.1 is only deposited into chromatin during DNA replication, whereas Hv2 is deposited outside of S-phase. The exclusive presence of these two histone variants in the macronucleus (or meiotic micronuclei) provided some of the first evidence that specific variants are preferentially associated with transcriptionally active chromatin.

In addition to core histone variants, the micro- and macronuclei also have distinct linker histones. Although neither linker histone is essential [23], when the genes for the micronuclear and macronuclear linker histones were disrupted, the nucleus in which they normally reside was increased in volume. These results were interpreted to mean that, in the absence of the linker histones, the chromosomes exhibited lower degrees of chromatin compaction. In addition, cells lacking the macronuclear linker histone showed altered gene expression profiles, a finding that providing some of the first evidence that linker histones have roles outside of maintaining general chromosome structure [24].

1.4.2 Differential Histone Modifications of Micro- and Macronuclei

The finding that histones in the macronucleus were hyperacetylated relative to those in the micronucleus provided evidence which corroborated Allfrey's observations, namely that acetylated histones were important for transcriptional activity in animals [25]. The ability to make targeted mutations in *Tetrahymena thermophila* allowed Martin Gorovsky and coworkers to test whether acetylation of the H2A.Z tail was critical for transcription, and to further assess whether specific sites needed to be acetylated [26]. In fact, Gorovsky's group found that the mutation of all normally acetylated lysines in the H2A.Z tail to arginines, which were not able to be acetylated, was lethal. However, the mutant phenotype could be rescued by H2A.Z proteins containing a single acetylated lysine. In addition, the Hv1 tail could be substituted for by the core H2A tail, thus demonstrating that the overall histone tail charge density was more important than the modification of particular tail lysine residues [27].

Arguably, one of the landmark discoveries in epigenetics research was the cloning of the first nuclear histone acetyltransferase (HAT). Previously, C. David Allis and coworkers had set out to identify the protein responsible for the hyperacetylation of macronuclear chromatin, by employing an in-gel histone acetylation assay [28]. For this, the histones were first polymerized directly into the denaturing protein gels used to fractionate the *Tetrahymena* extracts. After renaturing the proteins in the polyacrylamide matrix, the gels were incubated with radiolabeled acetyl-CoA. Subsequently, the group identified, and then purified, a 55 kDa protein that shared significant similarity with the yeast GCN5 transcriptional regulator. It was this discovery which established the paradigm that transcriptional regulators act by modifying chromatin [28, 29].

Other histone modifications enriched in either micro- or macronuclei hinted at their biological function. Histone H3 methylated on Lys4 was found exclusively in the macronucleus, thus providing the early evidence that this modification was associated with active chromatin [30]. This modification is absent from micronuclei, but is rapidly established on the bulk of the genome soon after developing macronuclei are formed. In contrast, the methylation of histone H3 on Lys9 is found exclusively during conjugation

on the chromatin of IESs in developing macronuclei [31]. This modification is lost from macronuclei as the IESs are removed from the genome. While the methylation of histone H3 on Lys9 was already known to be associated with silent heterochromatin in *Schizosaccharomyces pombe* and other eukaryotes, its linkage to IES excision – which was found concurrently to be controlled by a RNAi-related mechanism – provided one of the first examples (along with studies conducted in *S. pombe*) that RNAi-directed transcriptional gene silencing targeted the chromatin modifications to specific genomic regions [31–33].

While most chromatin modifications are enriched in macronuclei, the phosphorylation of histone H3 on Ser10 was found to be highly enriched in micronuclei undergoing mitosis or meiosis, indicating that this modification may be involved in chromosome condensation [34]. The mutation of Ser10 to alanine resulted in chromosome segregation defects, which further supported the importance of phosphorylation of this position on histone H3 in chromatin compaction during nuclear division [35]. These structural and functional differences between the micro- and macronuclei provided a rich biological context by which to start unraveling the role of chromatin proteins and their post-translational modifications for controlling epigenetic phenomena in ciliates.

2
Epigenetic Phenomena in Ciliates

Ciliates had been firmly established as genetic models for uncovering epigenetic phenomena long before many research groups began to use the differentiation of micro- and macronuclei as a means of resolving the molecular basis of epigenetic control. The many classical examples of non-Mendelian inheritance and other epigenetic phenomena that are described in the following sections have been included on the basis that it is useful to revisit these early observations in light of more recent molecular studies. Such examples of structural and cytoplasmic inheritance have a common feature, notably that the pre-existing phenotypic state of the parent cells is able somehow to ''template'' the phenotype that emerges in the next generation. These phenomena challenge many of the preconceived ideas of simple genetic inheritance, and beg for further investigation to decipher their underlying mysteries.

2.1
Structural Inheritance

In addition to nuclear dualism, ciliates are characterized by the extraordinary complexity and asymmetry of their cellular structures. The ciliate cortex is comprised of a matrix of cytoskeletal and membranous components, while organized within the cortex are organelles with specialized functions, such as the anteriorly positioned oral apparatus (a mouth-like phagocytic structure) and a posterior cytoproct. The elaborate ciliate body plan is faithfully reconstructed after each round of binary fission. The anterior daughter cell must reform the posterior structures, and the posterior daughter must generate a new mouth and other anterior components. Both, genetic and physical manipulations of the cortex have revealed that the cellular structure of ciliates is largely organized by the pre-existing structures, thereby demonstrating that a cell's phenotype is not determined solely by genotype.

These cells' numerous cilia, which are used primarily for locomotion and feeding,

project from arrays of cortical units, aligned into rows that are organized along the anteroposterior axis. Each cortical unit assumes a distinct anterior–posterior and left–right orientation that is crucial for the correct function of the cilia. During each cell cycle, the units are duplicated to ensure that each daughter cell inherits a complete set of structures that assumes the correct orientations. An early scientific question was whether this cortical organization was determined by the action of genes; subsequently, it was revealed that the structural organization of daughter cells is not established purely by the cells genotype, but rather is templated by the geometry of the pre-existing units (i.e., it is inherited through a non-genic mechanism). One of the earliest studies on cortical inheritance was performed using doublet cells. The ''doublet'' phenotype arises from a failure of pair separation at the end of conjugation, which leads to a fusion of the progeny. This phenotype is fairly stable, and can be propagated such that the vegetative progeny inherits a duplicated set of cortical structures. Genetic crosses demonstrated that the heredity of the doublet phenotype was not determined by genes or the cytoplasm, but rather was communicated through the architecture of the cortex itself [36].

Cortical inversion, a condition in which the cells have one or more ciliary rows rotated 180° in the plane of the cell surface, further illustrates the phenomenon of structural inheritance. In this case, an inverted patch of cilia is produced that results in the cells exhibiting an abnormal ''twisting'' swimming phenotype. As with the ''doublet'' phenotype, the progeny of cells with inverted patches inherits the inverted orientation of cilia, as the new cortical organization is templated by the parental cortical organization [37]. What this and other experiments show, in the case of ciliary orientation, is that whilst the genes supply the building blocks, the assembly into a functional organelle is determined by the structure of the pre-existing cortex. The ciliate cortex thus provides an example of structural memory, and reveals that genes are not the only cellular component that can pass on heritable information to the next generation.

2.2 Cytoplasmic Inheritance

The inheritance of pre-existing cellular structures is a specialized example of epigenetic influence on the phenotype. A more general non-nuclear medium for transmission of heritable information is the cytoplasm, the role of which as a director for epigenetic information is well documented in ciliates, notably in sexually reproducing *Paramecium aurelia* and related species. One reason for this is that, unlike some ciliates (e.g., *T. thermophila*), the conjugation of *P. aurelia* involves almost no cytoplasmic exchange between the mating pairs. Therefore, while cross-fertilization produces identical zygotic nuclei, these identical genomes develop in the different cytoplasmic environments of their respective parental cells. The interesting observation here is that these progeny – which are genetic twins – commonly express different phenotypes as determined by the cytoplasm in which their macronuclei develop.

Cytoplasmic inheritance in ciliates is most easily illustrated by determination of the mating type trait (Fig. 3) [38, 39]. *Paramecium* exist as two mating types: Even (E) and Odd (O). When two cells of opposite type mate, the progeny that arise from the E parent almost always assume

the E mating type, whereas those from the O parent almost always assume the O mating type, despite each having received identical genotypes. This observation suggests that something other than genes is directing the determination of the mating phenotype. A comparison of progeny mating types from crosses that do, and do not, exchange cytoplasm during conjugation further implicated cytoplasm as a key component in mating type determination [40–42]. If cytoplasmic exchange occurred between the mating pairs during conjugation, then the progeny of the O cell would often be switched to the E mating type. Furthermore, an injection of cytoplasm from the E mating partner into the O partner was found to transform the progeny's mating type from O to E. No effect was observed upon the transfer of O cytoplasm into E cells, which suggested that the cytoplasmic factor(s) must exist in the E cell to determine the E mating type, and that the E mating type is dominant over O [43].

Similar to mating type, the serotype of the *Paramecium* progeny can be strongly influenced by the cytoplasm in which a new somatic genome differentiates. Serotype is determined by the specific surface antigen protein that is expressed and displayed on the cell surface. Although several genes encode the different antigen proteins, only one gene is expressed in any given cell. Upon conjugation, the sexual progeny typically express the parental serotype. For instance, when cells of serotype A are crossed with serotype B, the progeny of both types will emerge expressing the serotype of the parent in which their nuclei developed [44]. The inheritance of mating type and serotype is, therefore, specified by the cytoplasmic environment rather than purely as genetic traits.

2.3 Epigenetic Control of Traits Converge with the Regulation of DNA Rearrangements

As noted above, the differentiation of a developing somatic macronucleus from its zygotic precursor involves an extensive streamlining of its germline-derived genome by removing extraneous "junk" DNA (see Fig. 2). Thus, the process of genome rearrangement directs major changes to the overall DNA sequence in the somatic macronucleus relative to the input from the germline. As the DNA removed is primarily noncoding, the suggestion that this DNA reorganization may or may not affect gene expression has not been extensively studied. For many ciliates, which have IESs imbedded within their coding regions, DNA elimination must occur to generate an expressible protein-coding region. It has been postulated – and supported by several experimental observations – that the epigenetic control of these DNA rearrangements may underlie at least some of the examples of non-Mendelian inheritance that have been discovered. The proposal that ciliates may differentially eliminate DNA sequences as a mechanism to alter the phenotype expressed by their progeny, is discussed in the following sections.

A genetic screen that initially was aimed at elucidating the molecular basis for mating-type expression eventually uncovered an intriguing link between this trait and the control of DNA rearrangement. A genetic mutation, mtF^E, was isolated in a cell line that produces only mating type E [45]. As noted above, *Paramecium* sexual progeny almost always assume the mating type of the parent (i.e., O parent, O progeny; E parent, E progeny). Hence, when an E individual that carries the mtF^E mutation (mtF^E/mtF^E) is

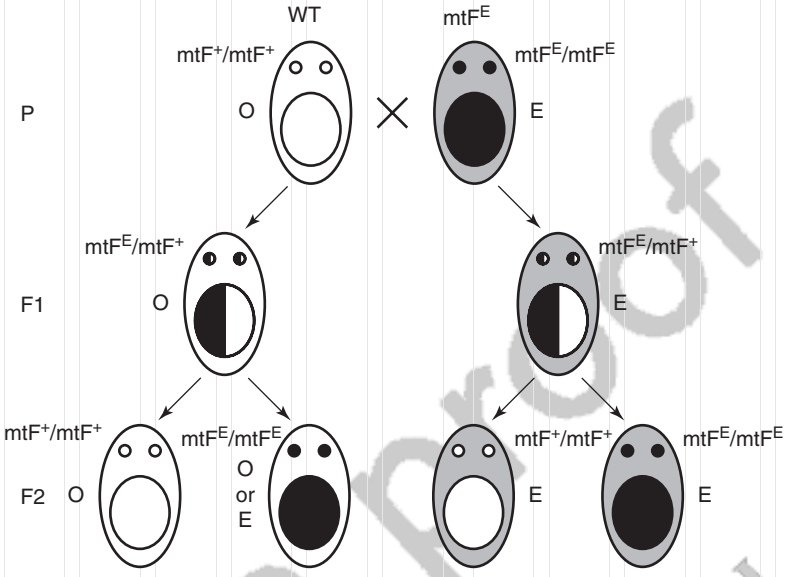

Fig. 4 The mtFE mutation: a genetic lesion that alters the cytoplasmic inheritance of mating type in *Paramecium*. P, Parent; F1/F2, generations; WT, wild-type.

crossed with a wild-type O individual (mtF^+/mtF^+), the mating type of the F1 progeny still follows the cytoplasmic inheritance typical of a wild-type mating (Fig. 4). However, homozygous mtF^E F2 progeny produced from autogamy of F1 O individuals (mtF^+/mtF^E) frequently switch to mating type E (compare Figs 3 and 4). Although the gene mutated in mtF^E strains has not been identified, a detailed study of the mtF^E/mtF^E strains showed that this mutation also led to a failure to eliminate an IES located in the G surface antigen gene. In turn, this observation led to the hypothesis that the gene mutated in mtF^E strains was involved in DNA rearrangement and by extension, that DNA rearrangement may be involved with mating type determination [46].

However, the mtF^E mutation studies provided more than just a link between DNA rearrangement and mating type; rather, they uncovered a means by which the epigenetic regulation of DNA rearrangements could alter the expression of specific traits. Further studies – not of mating type, but of G gene expression – revealed that the IES+ state of the G gene (apparently caused by the mtF^E mutation) became the heritable state of the G gene that was propagated through subsequent generations, even after reintroduction of the wild-type mtF^+ allele. Given the observed cytoplasmic inheritance patterns of both mating type and serotype traits in *Paramecium*, this finding offered an intriguing connection between alternative rearrangements and altered phenotypes.

The propagation of the IES+ state in the mtF^+ progeny showed that it was not a genetic lesion or other alteration to the germline genome that limited expression of the G gene. It was, in fact, the IES+ state itself that was present in the parental macronucleus and which elicited the transmittable influence of the

"cytoplasm" during development. This was demonstrated more conclusively by directly injecting the IES+ version of the G coding sequence into the maternal macronucleus, and showing that this alone was able to block the elimination of the homologous IES from the newly developed macronucleus after autogamy [47]. It is important to note that the injected DNA is destroyed along with the maternal macronucleus, so the IES+ state must be communicated to the developing macronucleus through the cytoplasm. The injection of plasmid DNA containing just the one IES, without any flanking G gene coding sequence, was found to be sufficient to block the elimination of this IES, while the remaining IESs within the G gene were excised efficiently. Thus, particular IES sequences present in the maternal macronucleus are able to communicate their presence to the zygotic macronucleus, and alter the normally efficient removal of the homologous sequence.

However, not all IESs were found to be subject to this form of homology-dependent regulation. When ten different IESs were microinjected into parental macronuclei to test their ability to block the excision of the homologous sequence, only four were able to inhibit DNA rearrangement. Whilst it was difficult to see why only some IESs in the zygotic macronucleus could sense the presence of homologous copies in the parental macronuclei, the clear implication here was that many characteristics could be reproducibly inherited in a non-Mendelian fashion, every time a new macronucleus is formed.

The serotype genes of *Paramecium* have proven to be fertile ground for uncovering epigenetic phenomena relating to genome rearrangements. One early and particularly interesting example was revealed by studies of a mutant strain called *d48*, that lacked the ability to express the surface antigen A gene [48]. Subsequent carefully conducted genetic studies showed that the d48 micronucleus contained a wild-type copy of the A gene; but that the macronucleus was missing the A gene-coding region [49]. The remarkable discovery was that the progeny of d48 strains reproducibly eliminated the A gene from their developing macronuclei during conjugation, making these progeny unable to express the A serotype.

The results of a series of microinjection and nuclear transplantation experiments confirmed that the presence of the A gene in the parental macronucleus was necessary for it to be retained in the progeny. Subsequently, microinjection of the A gene into the macronucleus of strains lacking the A gene in both the micro- and the macronuclei was sufficient to restore A gene expression during vegetative growth; however, this expression was lost during sexual reproduction when the microinjected parental macronuclei were fragmented and destroyed [50]. On the other hand, in the d48 strain – which lacks the A gene only in the macronucleus – microinjection of the A gene was sufficient to rescue A gene expression during vegetative growth, both in the parental strain and also in progeny cells following sexual reproduction [51–53]. Strains missing the surface antigen B gene have also been observed and rescued in a similar fashion [54, 55].

The rescue of A gene expression in the *Paramecium* d48 strain was found to be sequence-specific. Microinjection of the A gene or an allele of the A gene that has 97% identity resulted in A gene retention in the newly formed macronuclei of progeny. In contrast, introduction of the G surface antigen gene – which shares approximately 80% similarity with the A gene – failed to

rescue the A gene deficiency in the progeny [56, 57]. Thus, the DNA sequence of the parental macronucleus was again shown to have the ability to dramatically influence the types of sequence retained during the development of new macronuclei.

The observations made with d48 strains share intriguing parallels with both the inheritance of the IES+ state in the mtF progeny, and with the examples of cytoplasmic inheritance described above. In each case, the trait (or sequence) propagated is that which was expressed from the parental macronucleus. Thus, for ciliates the regulation of DNA rearrangements allows for somatic states of gene expression to be transmitted to the next generation. Recent studies of the mechanisms that guide DNA rearrangements have shown that homologous RNAs and chromatin-based regulatory schemes are key components. Studies of ciliate DNA elimination during macronuclear development have revealed that ncRNAs may also be the molecules responsible for many of the cytoplasmic and homology-dependent inheritance phenomena observed previously. These mechanisms will be described in more detail in the following subsections, as they offer many unique insights into how ncRNAs can pattern the genome and influence chromatin structure.

3 RNA-Mediated Epigenetic Mechanisms

3.1 Homology-Dependent Gene Silencing

Homology-dependent epigenetic phenomena have been observed widely, with the introduction of transgenes into plant cells often leading to a silencing of the endogenous copy. One of the most-often cited such examples resulted from an effort to create petunias that had darker flower petals, by adding exogenous copies of the chalcone synthase gene that generates the purple pigment [58]. However, instead of producing the expected increase in petal pigmentation, the transgenic petunias showed a decrease in coloration, in conjunction with an overall reduction in the mRNA level of chalcone synthase; this phenomenon was termed *co-suppression*. Similarly, the introduction of transgenes into the fungi *Neurospora crassa* induced a phenomenon known as *quelling*, which involved a silencing of the homologous endogenous gene [59]. Co-suppression has also been observed in the ciliate, *Paramecium tetraurelia*, upon high-copy microinjection of transgenes that lack 5′ and 3′ regulatory regions (i.e., lacking either promoters or transcription terminators), which resulted in a silencing of the endogenous homologous genes [60, 61].

The mysterious mechanism underlying these phenomena was discovered to be RNAi. A mechanistic insight into homology-dependent phenomena in ciliates has likewise been provided via connections to RNAi. In general, RNAi refers to a diverse collection of cellular mechanisms that employ RNA molecules to regulate the expression of genes (for reviews, see Refs [62–64]). In this case, the triggering molecule is typically double-stranded RNA (dsRNA) that is recognized by a ribonuclease known as *Dicer*, which cleaves dsRNA into fragments of approximately 20–30 nt. These so-called small RNA (sRNA) species serve as the specificity factors that guide an associated protein complex to a target mRNA or gene, where these effector RNA–protein complexes can promote silencing, either transcriptionally or post-transcriptionally.

RNAi appears to be an integral part of a variety of processes in ciliates. An examination of the bulk sRNA species in either *Paramecium* or *Tetrahymena* revealed distinct size classes, thus suggesting the existence of at least two different RNAi pathways [33, 65, 66]. The larger species (ca. 25 nt in *Paramecium* and 27–30 nt in *Tetrahymena*) were shown to be produced exclusively during conjugation, and to guide the extensive DNA rearrangements that occur in the differentiating somatic macronucleus (this RNA-guided genome reorganization is discussed in detail in Sect. 4).

A second class of ca. 23 nt RNAs is produced in growing cells, as well as during conjugation in *Paramecium* and *Tetrahymena*. This size class mediates post-transcriptional gene silencing (PTGS), and also the transgene co-suppression introduced above. In addition, the introduction of dsRNA aimed to experimentally induce gene silencing, either through feeding or direct injection into *Paramecium* cells, or by hairpin RNA expression in *Tetrahymena*, resulted in the production of these ca. 23 nt RNAs [61, 66–70]. Thus, these sRNAs are similar in function to the small interfering RNAs (siRNAs) discovered initially in plants by Baulcombe and colleagues, in that they carry out PTGS [71]. In *Tetrahymena*, these sRNAs are produced by Dcr2p from presumed pseudogenes or defective endogenous genes, which triggered the production of dsRNA precursors necessary for siRNA production [65, 72]. They are anti-sense to these predicted open-reading frames (ORFs), and depend on the activity of RNA-dependent RNA polymerase (RdRP), Rdr1p, which is found in a common complex with Dcr2p [72]. In *Paramecium*, a subclass of these smaller sRNAs is only anti-sense to mRNA transcripts, and is produced by a secondary amplification that involves the RdRPs, Rdr1p and Rdr2p [66, 73]. Although RNAi is clearly an important mechanism during the vegetative life of ciliates, its critical role has yet to be carefully examined. On the other hand, the function of RNAi pathways during development of the zygotic macronucleus has promoted new considerations regarding epigenetic programming of the genome.

3.2 RNA-Guided Genome Reorganization

Both, ncRNAs and RNAi-related mechanisms provide much more than a gene-silencing role in ciliates, as these organisms employ RNAs as guides to extensively remodel their genomes during sexual differentiation. Investigations aimed at elucidating the molecular mechanisms associated with the reorganization of the somatic genome of several ciliates have uncovered the involvement of ncRNAs [69, 74–77]. Indeed, the mechanisms identified have been shown to vary substantially among the different ciliate species studied, such that the data relating to *Paramecium*, *Tetrahymena*, and *Oxytricha* will be described separately in the following sections. Nevertheless, a common theme has emerged, in that these RNAs can serve as potent mediators capable of transmitting sequence-specific information between generations. The examples of homology-dependent regulation of phenotypes (particularly those described earlier in *Paramecium*; see Sect. 2.3) hinted that the mechanism(s) guiding genome rearrangements utilized some form of nucleic acid to transmit sequence-specific information between the somatic macronucleus of one generation and the developing macronucleus

of the next. These phenomena require that the state of the DNA in the parental macronucleus serves as a ''template'' for the traits expressed from the genome of the progeny.

Studies of *Paramecium* and *Tetrahymena* DNA rearrangements have identified two types of sequence-specific mediator RNAs – one which is produced from the germline genome, and a second produced from the parental somatic genome [69, 74, 75]. The germline-specific RNAs are in the form of sRNAs (known as scan RNAs; scnRNAs), that are produced during meiosis and act to identify the IESs as germline-limited sequences to be eliminated from the developing somatic genome [33, 66]. The second type of mediator RNA consists of longer transcripts produced from the parental macronucleus, and which appear to antagonize the action of the scnRNAs [69, 75]. It is these macronuclear transcripts that are the key epigenetic regulators that may explain the non-Mendelian inheritance of specific traits. In *Oxytricha*, analogous transcripts created from the parental somatic genome are postulated to serve as templates to directly guide the rearrangements, while a role for sRNAs is, as yet, unknown [76].

Genome scanning is a term used to describe the mechanism by which RNAs from the germline and somatic genomes can communicate the existing genomic content of the parental nucleus to the next generation [33]. Scanning occurs by a comparison of the germline-derived scnRNAs, with long ncRNA transcripts produced by the parental macronucleus [69, 75]. Such scanning assures that those scnRNAs made to regions of the genome which are not IESs, are removed from the pool of scnRNAs that target specific sequences for elimination. Scanning not only allows a ''proofreading'' of the sRNA pool to prevent any inadvertent elimination of sequences that should be retained, but also permits the retention of IESs that were maintained in the macronucleus of the previous generation and which offered some advantage or specified an alternative phenotype. The mechanisms of RNA-guided genome reorganization and genome scanning are described in the following sections, as these studies reveal the power of homologous RNAs to direct the programming of the somatic genome.

4 Small RNA-Mediated DNA Rearrangements

4.1 RNAi-Dependent DNA Elimination in *Paramecium*

The germline genome of *Paramecium tetraurelia* contains approximately 60 000 IESs that range in size from 28 to 886 bp [8, 78]. Many of these are found within coding sequences, and must be identified and excised with precision from the developing macronuclear chromosomes. Furthermore, during this genome maturation in *P. tetraurelia*, the more than 50 micronuclear chromosomes are fragmented into an unknown number of mini-chromosomes, amplified to 800*n* [79, 80]. The elimination of IESs occurs during both self-mating and sexual reproduction, at which time the parental macronucleus is destroyed and a new zygotic macronucleus is generated.

The results of studies performed over the past decade have revealed that the IESs are identified through the actions of homologous RNAs via an RNAi-related mechanism in *Paramecium*, and support the model shown in Fig. 5 [69]. A class of sRNAs each of ca. 25 nt, produced only

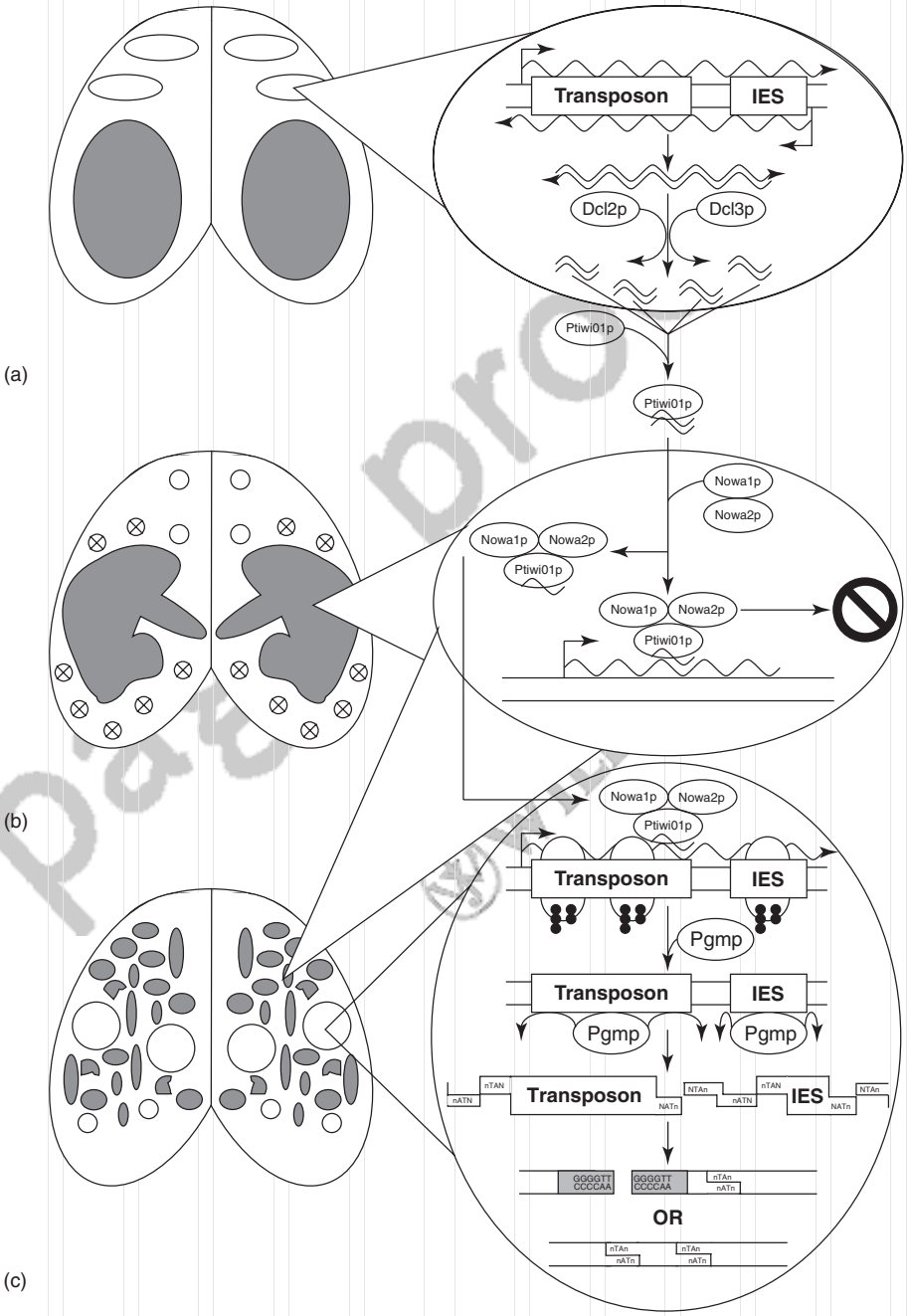

Fig. 5 *P. tetraurelia* uses RNAi and a domesticated transposase, Pgmp, for programmed DNA elimination.

during meiosis, has been shown to be necessary and sufficient to trigger the DNA elimination of IES sequences [66]. These were found to be homologous to a variety of DNA sequences throughout the genome, and likely function in similar manner to the *T. thermophila* scnRNAs described below [33, 66, 69]. These *Paramecium* scnRNAs have 2 bp 3′ overhangs, consistent with cleavage by an RNase III homolog [66, 81–83] (also see review in Ref. [84]). Seven RNase III homologs are present in *P. tetraurelia*, together with three Dicer (DCR) and four Dicer-like (DCL) homologs [66]. Although a single knockdown of the DCL genes has no effect on scnRNA production, double knockdowns of DCL2 and DCL3 will cause it to be abolished. The localization of Dcl2p in the crescent micronucleus early in meiosis indicates that the production of scnRNAs only takes place there at this early time point of conjugation. Double knockdowns of DCL2 and DCL3 also caused a failure of DNA elimination and produced non-viable progeny, further supporting the conclusion that the scnRNAs which they produce target the IESs for excision.

The scnRNAs produced by Dcl2p and Dcl3p cleavage in the crescent micronucleus are transported by the Piwi homologs, Ptiwi01p and Ptiwi09p, into the parental macronucleus to carry out genome scanning [85]. The scnRNAs that match the parental macronuclear genome are removed from the population that will be transported to the developing macronucleus later in development, to participate in genome restructuring. This scanning occurs by comparison of these germline-derived scnRNAs with a second type of regulatory RNA (long ncRNA transcripts produced in the maternal macronucleus), and ensures that scnRNAs made to regions of the genome that are not IESs are not inadvertently excised [69, 75].

Only a few proteins are known to play a role in the genome-scanning process in *P. tetraurelia*. Two glycine-tryptophan (GW) repeat proteins, Nowa1p and Nowa2p [86], have been identified as playing a role in this process; these were found initially to localize within the parental macronucleus during pre-zygotic development, and then to move to the developing macronucleus after its formation. A deletion analysis of Nowa1p showed that the N-terminal portion of the protein has nucleic acid-binding capabilities, particularly for RNA/DNA duplexes. The dimerization of Nowa1p, either with itself or perhaps with Nowa2p, appears to be essential for the nucleic acid-binding function. The double knockdown of NOWA1 and NOWA2 caused a failure of the DNA elimination of a specific class of IESs in *P. tetraurelia*; this was referred to as a maternally controlled internal eliminated sequences (mcIESs) [78, 86]. The failure of DNA elimination was complete in some cases, but incomplete in others [86]. A double knockdown of NOWA1 and NOWA2 also produced non-viable progeny, which indicated an essential function for the completion of autogamy or conjugation.

The question then was, ''How might the NOWA proteins contribute to the epigenetic control of IES excision?'', and ''What RNAs might they interact with?'' Previously, long ncRNA has been shown to have a role in several epigenetic phenomena in higher eukaryotes, including dosage compensation and genomic imprinting [87–93]. Data derived from *P. tetraurelia* have provided strong support for an interaction between the maternal transcription of long ncRNA and meiotic scnRNAs, and revealed exactly why this interaction is likely to be fundamental to genome

programming [69]. Reverse-transcription polymerase chain reaction (RT-PCR) studies of RNA isolated early in autogamy demonstrated the production of ncRNA without IESs, which were thought to be transcribed from the parental macronucleus. When a strain of *P. tetraurelia* containing a mcIES in the parental macronucleus was fed bacteria producing dsRNA prior to autogamy, or were directly injected with 23 nt siRNAs or 25 nt scnRNAs early during autogamy against this mcIES, the latter was removed from the developing macronucleus later in autogamy. These results indicated that genome scanning could be affected by degrading the long ncRNA in the parental macronucleus through bacterial feeding to produce 23 nt siRNAs, or by the direct injection of 23 nt siRNAs, as well as directly injecting the biologically active 25 nt scnRNAs to allow removal of an mcIES that normally would be retained on the completion of autogamy.

Long ncRNA also plays a role in the developing macronucleus by directing the remaining scnRNAs to sequences of DNA that are to be eliminated. In *P. tetraurelia*, the transport of these remaining scnRNA complexes to the developing macronucleus is mediated by the Piwi homologs, Ptiwi01 and Ptiwi09, where the production of long, ncRNA containing IESs has been detected using RT-PCR [69, 85]. Injection of the 25 nt scnRNAs in the same *P. tetraurelia* strain containing a mcIES in the parental macronucleus later during autogamy also causes removal of the mcIES, but the simultaneous injection of 23 nt siRNAs failed to cause DNA elimination [69]. In this case, it seemed likely that the 23 nt siRNAs actually promoted a failure of DNA elimination by targeting the long ncRNA needed for DNA elimination for degradation, while the 25 nt scnRNAs were able to recruit the necessary proteins for the DNA elimination of this mcIES.

4.2 The Role of a Domesticated PiggyBac Transposase in DNA Elimination and Chromosome Breakage in the Developing Somatic Nucleus of *Paramecium*

Each of the different varieties of RNA that are seen only during autogamy or conjugation in *P. tetraurelia* are all directed to one goal, namely the elimination of IESs and repetitive sequences. The removal of any of these types of RNA during the reproductive process causes nonviability [66, 69]. In order to eliminate IESs and repetitive sequences from the genome, these scnRNAs must recruit an excisase, a role for which recent data have implicated the domesticated piggyBac transposase, Pgmp [12]. In order to understand the role of Pgmp in DNA elimination, a brief description of IESs is called for. In *P. tetraurelia*, each IES is flanked by terminal inverted repeats, the consensus sequence of which is 5′-tggTAYAGYNR-3′ [8, 94]. Subsequently, cleavage occurs between the two guanosines in the consensus sequence, to produce a 5′ 4 bp overhang centered around the TA dinucleotide [95]. Mutations in either the T, A, or G in the third, fourth, and eighth position, respectively, of the above consensus sequence are then sufficient to block cleavage [96–99]. Cleavage of the consensus sequence, 5′-TTAA-3′, by piggyBac transposases to produce a 5′ 4 bp overhang is somewhat similar to the *P. tetraurelia* consensus IES sequence and cleavage product [100, 101]. An analysis of the *P. tetraurelia* genome identified a piggyBac homolog, called piggyMac (PGM) [12]. Localization of the green fluorescent protein (GFP)–Pgmp was found only

in the developing macronucleus late in autogamy. The knockdown of PGM late in conjugation resulted in a failure to produce any viable progeny, a failure of IES excision and chromosome breakage, and an overexpression of IES-containing ncRNA from the developing macronucleus. These knockdown phenotypes implicated Pgmp as having an essential role in the completion of DNA elimination and chromosome breakage in *P. tetraurelia*, most likely through Pgmp-mediated dsDNA breakage to remove IESs and other repetitive sequences. The repair of these dsDNA breaks is mediated by the DNA ligase IV homologs, LIG4a and LIG4b [102].

The removal of these IESs and other repetitive sequences in *P. tetraurelia* and other ciliates is the ultimate epigenetic action. Unlike most other eukaryotes, which heterochromatize their repetitive and noncoding sequences, the ciliates excise and degrade these sequences from their somatic macronucleus, and then amplify the remaining sequences so as to create a streamlined genome that allows a greater cell size than most other eukaryotes and a growth rate comparable to that of yeast. As discussed earlier in brief, a removal of IESs and other repetitive elements occurs in completion of sexual reproduction (the actual removal of these two types of sequence may differ slightly, and even impact on the final state of the genome after sexual reproduction). Two different classes of IESs have been identified – namely mcIESs and non-mcIESs – which are small, are found throughout the genome, and eliminated in a precise fashion [47, 78, 95]. The mcIESs are capable of having their excision blocked by the insertion of a copy of the mcIES into the parental macronucleus [47, 78]. The mcIESs tend to be larger in general, and it has been hypothesized that their elimination is dependent on chromatin modifications directed by genome scanning [103]. In contrast, non-mcIESs are smaller, with most being shorter than the amount of DNA wrapped around a nucleosome, which would necessitate a different targeting method for DNA elimination. It seems possible that their elimination could take place through a directed binding of Pgmp, or through guidance of Pgmp via a nucleotide modification to their cleavage sequences. Repetitive sequences are removed with much less precise methods, and this results in either variable cleavage or fragmentation of the chromosome [104]. Both types of DNA elimination depend on the action of Pgmp [12].

Despite all that has been learned regarding the epigenetic phenomenon of RNAi-directed DNA elimination in *P. tetraurelia*, many questions remain to be answered:

- "How are these ncRNAs produced in any of the nuclei?"
- "What is the difference between mcIESs and non-mcIESs, and how does that affect their DNA elimination?"
- "How does DNA elimination, RNAi, and heterochromatin function in related ciliates, and in general how is this biological process related to other epigenetic processes in other eukaryotes?"

Investigations into the RNAi-directed DNA elimination process in a related ciliate, *T. thermophila*, have provided additional insights into many of these questions.

4.3 RNAi-Dependent DNA Elimination in *Tetrahymena*

Like *P. tetraurelia*, the ciliate *T. thermophila* also undergoes massive DNA elimination

and chromosome breakage during sexual reproduction or conjugation. During conjugation in *T. thermophila*, the developing zygotic macronucleus is fragmented into approximately 200 minichromosomes from five chromosomes, while 15% of the overall DNA content is removed and the remaining DNA content is amplified to 50*n* [7, 105–110] (for a review, see Ref. [5]). Similar to *P. tetraurelia*, the mechanism of this process was poorly understood until the discovery of conjugation-specific, long ncRNAs and a class of sRNA (termed *scnRNAs*) that are derived from the ncRNAs, and which has led to the model shown in Fig. 6 [33, 74, 75].

The scnRNA model of RNAi-dependent DNA elimination in *T. thermophila* can effectively be broken into two parts: (i) production and selection of the scnRNAs by conventional RNAi-associated proteins; and (ii) transduction of the scnRNA signal into heterochromatin formation, which subsequently triggers a DNA elimination of the heterochromatic DNA in the developing zygotic macronucleus. For each of these parts, the experimental data supporting the model, how that data can be used to further elucidate the mechanism of RNAi-dependent DNA elimination, and how the results obtained relate to epigenetics in ciliates and other eukaryotes, are discussed in the following subsections.

4.4 RNAi Apparatus and Genome Scanning in DNA Elimination

The role of RNA during the development of many eukaryotes has been well documented [33, 74, 87–92, 111–116]. For example, *T. thermophila*, like *P. tetraurelia*, has been shown to possess two classes of sRNAs that range from 23 to 24 nt and from 28 to 30 nt in size [33, 65, 117, 118], where the larger class – the scnRNAs – is restricted to conjugation [33]. These appear to be functionally similar to piRNAs that have been described in a variety of organisms, and which are known to act to protect the germline genome in the micronucleus against possible deleterious effects that active transposons can inflict, such as gene inactivation, chromosome translocation, and chromosome breakage [118–125]. Unlike piRNAs, which are Dicer-independent, scnRNA production in both *P. tetraurelia* and *T. thermophila* is totally dependent on a group of DCL proteins [66, 117, 118, 120, 123]. If the DCL genes are either knocked out or knocked down, the scnRNAs are not produced during conjugation, and this triggers a developmental arrest [66, 117, 118]. Whilst it is intriguing that these scnRNAs in *P. tetraurelia* and *T. thermophila* exhibit properties of both piRNAs and siRNAs, further studies of the scnRNA pathway may contribute to a fundamental understanding of how both the piRNA and siRNA pathways arose in higher eukaryotes.

4.5 Bidirectional Transcription of Long dsRNAs

The production of scnRNAs depends on the synthesis of long dsRNA precursors [117, 118]. At an early stage during conjugation, the micronucleus detaches from a groove in the parental macronucleus and elongates to form a crescent that is approximately the length of two cells [127, 128]. During vegetative growth in *T. thermophila*, the micronucleus is transcriptionally silent, although some decades ago it had been observed that early during conjugation (starting after micronuclear detachment from the parental macronucleus and peaking just prior to full crescent elongation) there was copious

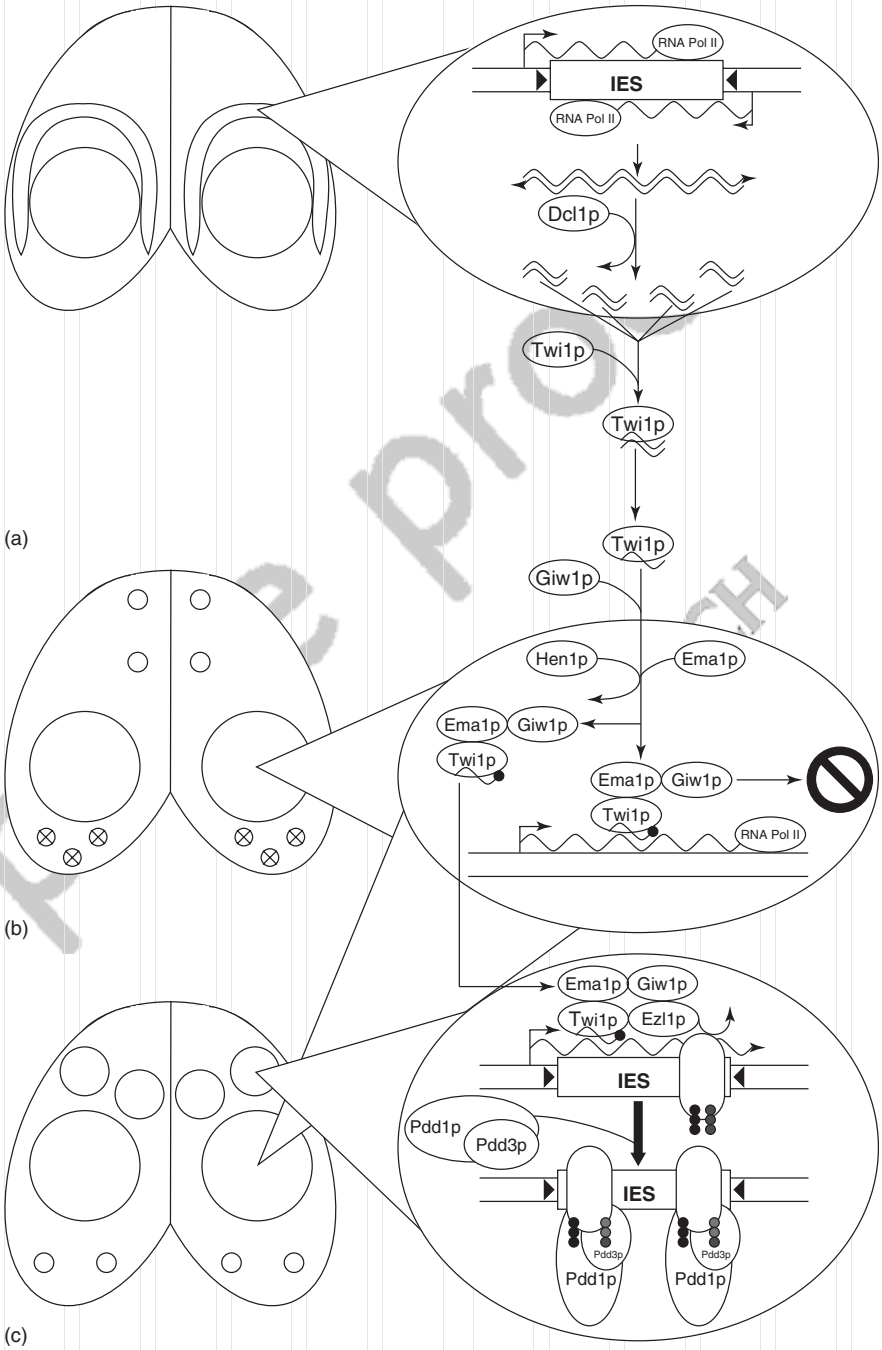

Fig. 6 Meiotic germline transcription as a RNAi pathway direct DNA elimination in *T. thermophila*.

transcription from the micronucleus [19, 20]. The results of later studies conducted in *T. thermophila* showed that, just prior to this period of micronuclear transcription, the *T. thermophila* H2A.Z homolog was deposited in the micronucleus, despite normally being found only in the macronucleus [21, 129]. Other studies also showed that RNA polymerase components, including a putative TATA-binding protein, TBP1, and a RNA polymerase II subunit, RPB3, were localized to the micronucleus during this burst of transcription [130, 131]. This implied that the RNA polymerase responsible for this early micronuclear transcription was RNA polymerase II. Investigations performed on an IES, the M element, showed that the transcription of both strands was markedly increased early in conjugation during the same time period that general micronuclear transcription was increased [74]. These transcripts produced early in conjugation were also heterogeneous at the 5′ and 3′ ends and, unlike RNA polymerase II mRNA transcripts, lacked 3′ polyadenylation. Further studies of the transcription of other known IESs indicated that this is a general characteristic of RNAs produced during this time point in conjugation in *T. thermophila*, which meant that the burst of transcription seen in the micronucleus produced the long, IES-specific dsRNA precursors required for scnRNA production.

4.5.1 Processing of Long dsRNAs into scnRNAs, and Their Subsequent Nuclear Localization

The long, IES-specific dsRNA transcripts are scnRNA precursors, which are processed by Dicer proteins [74, 117, 118]. An analysis of the sequence of the *T. thermophila* macronuclear genome indicated the presence of three putative Dicer proteins [117, 118], two of which were expressed throughout the *T. thermophila* life cycle, while the third Dicer protein, Dicer-like protein 1 (Dcl1p), was expressed exclusively during conjugation. Although the DCL proteins, such as DCL1 in *T. thermophila* and DCL2 and DCL3 in *P. tetraurelia*, lack the conserved RNA helicase domain, they have been shown to play an important role in epigenetic phenomena in other organisms besides ciliates, including *Arabidopsis thaliana* [132]. Knockouts of DCL1 caused a massive increase in these long, IES-specific dsRNA transcripts, yet at the same time they caused the abrogation of scnRNAs [33, 117, 118]; this verified that the long, IES-specific dsRNA transcripts produced early in conjugation are precursors for scnRNAs [117, 118]. Knockouts of DCL1 also failed to complete conjugation and, more importantly, failed to undergo DNA elimination similar to the DCL2/DCL3 double knockdown in *P. tetraurelia* [66, 117, 118]. The localization of Dcl1p, like Dcl2p in *P. tetraurelia*, showed that it was exclusively a micronuclear protein, which meant that the long dsRNAs produced in the micronucleus were processed into scnRNAs in the micronucleus itself, and not exported for cleavage.

Studies of the scnRNA structure itself showed that they were phosphorylated at the 5′ end, and also contained a 3′ hydroxyl group, which was consistent with cleavage by the ribonuclease III family member Dcl1p [33, 117, 118, 133–136]. Hybridization of these scnRNAs to micronuclear and macronuclear genomic DNA preparations from early to late in conjugation (2–10 h) showed a gradual increase in the ratio of scnRNAs hybridizing to micronuclear DNA when compared to macronuclear DNA, thus indicating the existence of a scnRNA

sorting mechanism [33, 137]. At 2 h, the ratio of micronuclear DNA to macronuclear DNA binding was approximately threefold [137], but as conjugation proceeded this ratio gradually increased to a maximum of approximately 30-fold at 10 h [33]. Further analysis of some of these scnRNAs showed that they were homologous to the M and long terminal repeat (LTR) IES sequences, consistent with their production from long, IES-specific dsRNAs [75, 117].

Argonaute proteins have been shown to be essential effector proteins in sRNA pathways [138]. The same is true for *T. thermophila* as an Argonaute homolog, TWI1, was shown to bind scnRNAs [33]. A phylogenetic analysis of Twi1p indicated that it was homologous to the *Drosophila melanogaster* Piwi protein, and belonged to the Piwi subfamily of Argonaute proteins. TWI1 was predicted to contain functional PAZ and PIWI domains, which facilitate nucleic acid binding and "Slicer" or ribonuclease activity, respectively. The immunoprecipitation of Twi1p shortly after the production of scnRNAs at 5 h into conjugation demonstrated Twi1p/scnRNAs interaction [137]. The localization of Twi1p showed that the protein was predominantly macronuclear with some cytoplasmic localization, but was excluded completely from the crescent micronucleus; this indicated that the scnRNAs would have to undergo active or passive transport into the cytoplasm to interact with Twi1p [33]. Mutation of the DDH motif in the PIWI domain of TWI1 abolishes ribonuclease activity in Twi1p, and prevents removal of the passenger strand in Twi1p/scnRNA complexes found in the cytoplasm [139]. Mutation of the DDH motif also blocks the import of the Twi1p/scnRNA complexes into the parental macronucleus, which leads to scnRNA instability and degradation over a similar time course when compared to TWI1 knockouts [137, 139].

Like Argonaute proteins in other organisms, Twi1p does not act alone during RNAi-dependent DNA elimination in *T. thermophila*. In order for import into the parental macronucleus of the Twi1p/scnRNA complexes to occur, Twi1p must also interact with an accessory protein called *Giw1p* [139]. Although GIW1 shows no homology to any known domains of any gene, Giwi1p coimmunoprecipitates with full-length Twi1p, interacting with the PAZ and PIWI domains of Twi1p along several discrete protein sequences. Mutation of the DDH motif in Twi1p, which blocks cleavage of the double-stranded scnRNA and also prevents binding of Giw1p to Twi1p, ensures Twi1p/scnRNA complex activation prior to parental macronuclear import. Localization of Giw1p is seen generally in the parental macronucleus and the cytoplasm early in conjugation, where it is capable of participation with Twi1p/scnRNA complexes before importing them into the parental macronucleus. Giw1p also localizes to the developing zygotic macronucleus later in conjugation, although its function there at that time is not known. Knockouts of GIW1 cause failure of Twi1p/scnRNA complex import into the parental macronucleus, but do not affect scnRNA cleavage or unwinding of the scnRNA passenger strand which, along with the Twi1p/scnRNA complex binding data, indicates the activation of Twi1p/scnRNA complexes before Giw1p-dependent import. Like the DCL1 knockout, knockouts of TWI1 and GIW1, as well as the TWI1 PIWI domain mutation, fail to complete conjugation and block the DNA elimination of IESs [33, 137, 139].

4.5.2 Genome Scanning via Comparison of scnRNA Complexes to the Parental Genome

Localization of the Twi1p/scnRNA complexes into the parental macronucleus sets the stage for one of the unique aspects of DNA elimination in *T. thermophila*. As noted above, there is an increase in hybridization levels of scnRNAs to micronuclear genomic DNA when compared to macronuclear genomic DNA as conjugation proceeds, indicating the presence of a sorting mechanism [33, 137]. The sorting process through which micronuclear-specific scnRNA enrichment occurs is referred to as *genome scanning* [33]; this is similar to the situation in *P. tetraurelia*, and involves comparing each Twi1p/scnRNA complex to ncRNA transcribed from the parental macronucleus. Those Twi1p/scnRNA complexes which bind to the parental macronuclear ncRNA are removed from the biologically active Twi1p/scnRNA complex pool through unknown means, although a handful of proteins have been identified that play a role in this genome-scanning process.

Emphasizing the connection of scnRNAs with piRNAs, a homolog of HEN1 (the protein which is known to stabilize piRNAs through methylation) has also been found to have the same role in *T. thermophila* with scnRNAs [140]. The homolog in *T. thermophila*, which is also called *HEN1*, is a RNA methyltransferase that adds a methyl group to the terminal 2′ hydroxyl group of scnRNAs and has homologs in *A. thaliana*, *D. melanogaster*, and *Mus musculus* [140–144]. Hen1p colocalizes with Twi1p in the parental macronucleus early in conjugation during meiosis of the micronucleus; indeed, *in vitro* experiments with recombinant Hen1p and Twi1p have shown that Hen1p also coimmunoprecipitates with Twi1p during this period of development [140]. Knocking out HEN1 causes a loss of 2′-O-methylation in scnRNAs, and decreases scnRNA stability in a similar fashion to the TWI1 knockout and TWI1 PIWI domain mutant [137, 139, 140]. However unlike the TWI1, GIW1, and DCL1 knockouts, knockouts of HEN1 do not show a complete failure of conjugation and blockage of DNA elimination [117, 118, 137, 139, 140]. HEN1 knockouts are able to produce only 3% of possible progeny, but are able to undergo a complete rearrangement of the IESs tested on 67.8% (38/56) of occasions [140]. It is possible that, since scnRNA destabilization is not as extreme as in a TWI1 knockout or PIWI domain mutant, the sheer number of scnRNAs remaining is able to facilitate DNA elimination of IESs and the completion of conjugation.

Several Argonaute proteins that associate with piRNAs in other organisms have also been found to associate with RNA helicases [126, 145–147] (for a review, see Ref. [148]). An RNA helicase in *T. thermophila*, Ema1p, interacts with Twi1p/scnRNA complexes and plays a pivotal role in genome scanning by facilitating the Twi1p/scnRNA/ncRNA interaction [75]. Ema1p colocalizes with Twi1p in the parental macronucleus early in conjugation and later in the developing zygotic macronucleus, where the proteins have also been found to interact through coimmunoprecipitation [33, 75]. Ema1p localization is unaffected in TWI1 or GIW1 knockouts, which indicates that it is imported into the parental macronucleus either by itself, or by the same group of proteins that imports Giw1p/Twi1p/scnRNA complexes [75, 139]. Knockouts of EMA1 logically do not inhibit scnRNA cleavage or import

of Twi1p/scnRNA complexes into the parental macronucleus, since it is never seen to accumulate in the cytoplasm where these processes occur. However, chromatin-spreading experiments and RNA immunoprecipitation followed by RT-PCR, have shown that in EMA1 knockouts the Twi1p/scnRNA complexes are no longer able to interact with chromatin and ncRNA when compared to wild-type. This was especially significant since chromatin was through to be the site of ncRNA production, and that Twi1p/scnRNA/ncRNA interaction was required for genome scanning. As conjugation proceeds, the EMA1 knockouts also displayed an increase in macronuclear-specific scnRNAs compared to wild-type matings. These data implied that Ema1p would facilitate genome scanning by coupling Twi1p/scnRNA complexes with the ncRNA produced in the parental macronucleus, and also through an unknown mechanism which negatively selected against those Twi1p/scnRNA complexes capable of binding successfully to the ncRNA. Finally, EMA1 knockouts failed to complete conjugation yet, curiously, only showed a failure of DNA rearrangement in a select set of IESs. This may point towards the existence of different classes of IESs in *T. thermophila* (as occurs in *P. tetraurelia*) that do not undergo this selection process [47, 75, 78].

Although relatively few proteins are known to play a role in the RNAi-dependent DNA elimination process, there exist a few situations in this process where homologs in one ciliate are found to play the same or similar role in another ciliate [12, 14, 33, 66, 85, 86, 117, 118, 149]. One of these sets of homologs is the GW repeat proteins Nowa1p and Nowa2p in *P. tetraurelia*, and Wag1p and CnjBp in *T. thermophila* [86, 149]. The GW repeat proteins have been found to interact with Argonaute family proteins in *A. thaliana*, *D. melanogaster*, and *Homo sapiens*, and to play a role in sRNA effector function [150–152]. Although, Nowa1p and Nowa2p appear to have RNA-binding capabilities, the function of their homologs, Wag1p and CnjBp, in *T. thermophila*, is unclear [86, 149]. Subsequent colocalization and coimmunoprecipitation experiments with Wag1p and CnjBp demonstrated a protein–protein interaction with Twi1p [75, 149]. CnjBp was also shown to localize to the crescent micronucleus during meiosis (unlike Twi1p and Wag1p), although its role there is currently unknown [149]. Double knockouts of WAG1 and CNJB caused the retention of macronuclear-specific scnRNAs compared to wild-type matings, as conjugation proceeded in a similar fashion to the EMA1 knockout [75, 149]. Unlike the EMA1 knockout, the double WAG1/CNJB knockout also showed a slight increase in the retention of micronuclear-specific scnRNAs. This may entail a more general function of these two GW repeat proteins in the genome-scanning process, for the Twi1p/scnRNA complexes that need to be sequestered in the parental macronucleus, and for those complexes that need eventually to be transported to the developing zygotic macronucleus [148]. Although double knockouts of WAG1/CNJB show an increased retention of scnRNAs, the Twi1p/scnRNA complexes are able to interact with ncRNA through Ema1p normally, indicating that their biological function lies downstream of the initial binding of Twi1p/scnRNA complexes with ncRNA. Like many of the proteins involved in RNAi-directed DNA elimination, the double knockouts of WAG1/CNJB failed to complete

conjugation, but failed DNA elimination in a specific set of IESs only (much like EMA1 knockouts) [75, 149]. Curiously, this set of IESs was slightly different from those in EMA1 knockouts [149], and although GW repeat proteins have been shown to affect Argonaute function, the actual mechanism remains a mystery [149–152]. In fact, even among ciliates there is no clear mode of action for these GW repeat proteins [86, 149]. Nonetheless, as more information becomes available regarding the RNAi-dependent DNA elimination pathway in both *P. tetraurelia* and *T. thermophila*, it will be interesting to see whether Nowa1p and Nowa2p in *P. tetraurelia* function similarly to Wag1p and CnjBp in *T. thermophila*, through sorting Argonaute/scnRNA complexes. Likewise, the proof of RNA binding by Wag1p and CnjBp (which has already been demonstrated in Nowa1p and Nowa2p) could help to define a common mode of action for GW repeat proteins in ciliates, and possibly in other eukaryotes in general.

Long ncRNA has been shown to play a vital role in a variety of epigenetic phenomena, as noted above [74, 75, 87–93]. In both *P. tetraurelia* and *T. thermophila* there appear to be three sources of long ncRNA during sexual reproduction: the crescent micronucleus; the parental macronucleus; and the developing zygotic macronucleus [74, 75]. The ncRNA produced in the parental macronucleus is vital to the genome scanning process, and was initially detected in *T. thermophila* alongside the bidirectional transcribed long, IES-specific dsRNA scnRNA precursors, and the ncRNA produced in the developing macronucleus [74]. PCR-based assays devised to further examine ncRNA transcription during conjugation showed that the long, IES-specific dsRNA scnRNA precursor transcription peaked at 3 h, ncRNA transcription from the parental macronucleus necessary for genome scanning peaked at 6 h, and ncRNA transcription from the developing zygotic macronucleus for IES targeting peaked at 10 h [75]. Blocking the transcription of parental macronuclear ncRNA by treatment with actinomycin D during the peak hours of genome scanning (4–6 h into conjugation) caused a significant increase in the failure of IES excision and DNA elimination [74]. Besides using actinomycin D, it is also possible to block the excision of individual IESs by inserting the IES sequence into the parental macronucleus prior to conjugation, similar to the blockade of mcIES excision in *P. tetraurelia* [47, 78, 153, 154]. For example, in *T. thermophila*, an insertion of the M element IES into the parental macronucleus causes a massive increase in M element long dsRNAs, but with no change in the level of scnRNAs [153]. This indicates that the excess long dsRNAs were not being processed into scnRNA, but were most likely acting as ncRNAs in the parental macronucleus, thereby removing M element scnRNA/Twi1p complexes from the biological active pool of Twi1p/scnRNA complexes.

4.6 DNA Elimination of DNA Sequences from the Developing Somatic Nucleus

When initially discovered, the phenomenon of DNA elimination in ciliates appeared to be an aberration in the world of biology, that was focusing increasingly on genetic processes. However, the rise of epigenetics has facilitated a clearer view of how DNA elimination relates to other biological processes. Whilst the link between scnRNAs and piRNAs was discussed in Sect. 4.5, this is not

the only biologically relevant link that DNA elimination in ciliates has to other organisms. Just as RNAi was shown to direct heterochromatin formation in *A. thaliana* and *S. pombe*, it was also shown that a correct heterochromatin formation in the developing zygotic macronucleus through H3K9 and H3K27 methylation would depend on the normal function of RNAi components in *T. thermophila* [112, 117, 118, 155–157]. Thus, DNA elimination depends on an establishment of heterochromatin to control the glut of repetitive elements in its genome [31, 155, 158]. As with other eukaryotes, the initial methylation of histones associated with repetitive elements precipitates heterochromatin formation and the compaction of these sequences. Typically, *T. thermophila* and other ciliates take the additional step of removing these heterochromatic sequences out of their somatic genome, in order to create a streamlined genome (not unlike many simple eukaryotes) to optimize their fitness. This streamlining process begins when the Twi1p/scnRNA complexes have been transported to the developing zygotic macronucleus to target the H3K9 and H3K27 methylation of IESs [31, 137, 158]. These methylated histones then act to recruit chromodomain and other accessory proteins, which ultimately promote IES excision and DNA elimination by the domesticated piggyBac transposase, Tpb2p [14, 159–163]. The link between RNAi and heterochromatin, IES-specific chromatin modifications, heterochromatin readers, and the nature of IESs and DNA elimination in *T. thermophila*, will be described in the following section, together with details of relevant experiments to determine each of these steps.

4.6.1 Targeting of scnRNA Complexes and Modification of Chromatin of DNA Sequences to be Eliminated

Like RNAi-directed heterochromatin formation in *A. thaliana* and *S. pombe*, RNAi-dependent DNA elimination in *T. thermophila* requires the production of ncRNA [74, 75, 112, 156, 164]. This ncRNA (which is created in the developing zygotic macronucleus) is necessary for targeting IESs, and interacts with the remaining Ema1p/Twi1p/scnRNA complexes, which are transported there once the developing macronucleus has moved to the anterior of the cell and has begun to enlarge [33, 75, 137, 139]. The Twi1p accessory proteins involved in genome scanning, Ema1p, Wag1p, and CnjBp, are also transported to the developing macronucleus, although it is unclear whether this occurs in a greater complex with Twi1p, or independently [75, 149]. The Ema1p/Twi1p/scnRNA/ncRNA complex interaction facilitates the binding of this complex with another group of proteins referred to as the *Ezl1p complex* (S.D. Taverna *et al.*, unpublished data) [75].

In the RNAi-directed heterochromatin formation pathways in *A. thaliana* and *S. pombe*, heterochromatin formation is directed by H3K9me2, which is catalyzed by the Su(var) three to nine homologs, Kryptonite (KYP), and Clr4, respectively [165–167]. RNAi-dependent DNA elimination in *T. thermophila* is dependent instead on Ezl1p, an E(z) homolog, and other associated proteins (S.D. Taverna *et al.*, unpublished data) [158]. The Ezl1p complex, which consists of Ezl1p, Esc1p, Rnf1p, Rnf2p, and Nud1p, contains homologs from two protein complexes, PRC1 and PRC2, as found in higher eukaryotes. These complexes are known to play a fundamental role in the developmental regulation of heterochromatin through

histone methylation and gene silencing in many organisms, which the Ezl1p complex has subsumed in *T. thermophila* (S.D. Taverna *et al.*, unpublished data) [158, 168–170]. Immunoprecipitations of Ezl1p, Nud1p and Rnf1p are able to pull-down Ema1p, thus demonstrating a protein–protein interaction between the Ema1p/Twi1p/scnRNA complex and the Ezl1p complex (S.D. Taverna *et al.*, unpublished data). Nud1p, Rnf1p, Rnf2, and Esc1p of the Ezl1p complex appear to have no catalytic function themselves, unlike other homologs found in PRC1 and PRC2 complexes, but instead act to enhance targeting of Ezl1p to IESs and Ezl1p methylase activity at the IESs (S.D. Taverna, unpublished data) [171–175] (see review in Ref. [176]). Ezl1p, which is the effector component of the Ezl1p complex, is an E(z) homolog and contains the SET domain, which is capable of trimethylation of H3K9 and H3K27 (S.D. Taverna *et al.*, unpublished data) [158, 177–180]. The coimmunoprecipitation of Ezl1p is able to pull-down the other members of the Ezl1p complex, Nud1p, Rnf1p, Rnf2, and Esc1p (S.D. Taverna *et al.*, unpublished data); reciprocal pulldowns using tagged-Nud1p and -Rnf1p are also able to immunoprecipitate Ezl1p. The colocalization of H3K9me3 and H3K27me3 with Rnf1p of the Ezl1p complex shows that it is capable of histone methylation during conjugation. A knockout of any of the Ezl1p complex components causes disassociation of the complex and loss of H3K9 methylation along with aberrant H3K27 methylation, which implicates the Ezl1p complex in both H3K9me3 and H3K27me3 during conjugation (S.D. Taverna *et al.*, unpublished data) [158]. Knockouts of the EZL1 complex also result in an increased accumulation of scnRNAs and ncRNAs produced in the developing macronucleus from the M IES, which indicates the existence of a feedback mechanism controlling both scnRNA and ncRNA production throughout the cell during conjugation (S.D. Taverna *et al.*, unpublished data). DCL1, TWI1, and EZL1 complex knockouts also form aberrant DNA elimination bodies, which contain a number of proteins including the chromodomain proteins, Pdd1p, and Pdd3p (S.D. Taverna *et al.*, unpublished data) [158]. Like other components of RNAi-directed DNA elimination, knockouts of the Ezl1p complex caused failure of DNA elimination (S.D. Taverna *et al.*, unpublished data). In the case of EZL1 knockouts, a failure to complete conjugation has also been observed [158].

Methylation of H3K9 and H3K27 by the Ezl1p complex is an integral part of the RNAi-dependent DNA elimination process [31, 155, 158]. Indeed, the inhibition of this methylation by the Ezl1p complex through knockout of any component of RNAi-directed DNA elimination upstream or mutation of histone 3 itself is sufficient to block binding of the chromodomain proteins, Pdd1p and Pdd3p, and its association with other proteins to form DNA elimination bodies necessary for DNA elimination [75, 117, 118, 155, 158]. Mutation of H3K9Q directly blocks the site from methylation, while mutations of H3S10E and H3S28E created an artificially phospho-switch, which naturally prevents methylation of the lysine directly downstream. All of these histone 3 mutations prevent Pdd1p and Pdd3p association with IESs [155, 158].

4.6.2 Protein Binding of Modified Chromatin, Protein Aggregate Formation, and DNA Elimination

The role of chromodomain proteins in RNAi-directed heterochromatin

formation and heterochromatin formation in general in eukaryotes is well documented [111, 181–183]. Once H3K9me3 and H3K27me3 modification occurs on histones associated with IESs, the aforementioned chromodomain proteins, Pdd1p and Pdd3p, are able to bind the IES chromatin which, along with other associated proteins, condenses the approximately 6000 IES loci into a handful of cellular foci referred to as *DNA elimination bodies* [14, 31, 158, 159, 161–163, 184–186]. In these DNA elimination bodies a domesticated piggyBac transposase, Tbp2p, directs the endonucleolytic cleavage of IESs at the IES boundaries, excising the IES [14]. Although these double-strand breaks are thought to be repaired through one of the dsDNA break repair pathways, it is currently unknown which pathway is responsible for this repair in *T. thermophila*.

Chromodomain proteins are pivotal heterochromatin histone readers. Knockouts of chromodomain proteins cause derepression of heterochromatin [183, 187]; likewise, knockouts of PDD1 also see a decrease in heterochromatin formation [31, 158]. This implies that the establishment of H3K9me3 and H3K27me3, and the binding of the two chromodomain proteins (Pdd1p and Pdd3p) to these marks, are interconnected in DNA elimination body formation and DNA elimination (see Fig. 7a,b) [31, 158]. Pdd1p and Pdd3p, along with Pdd2p, were discovered by the isolation of proteins enriched in developing zygotic macronuclei late during conjugation, and were the first identified proteins shown to play a role in DNA elimination [159, 161, 184, 185]. Pdd1p contains two chromodomains, and is capable of binding either H3K9me3 or H3K27me3 peptides *in vitro*, and to colocalize with H3K9me3, H3K27me3-modified chromatin and IESs late in conjugation [31, 158, 159, 188].

Pdd1p may play multiple roles during development, as it has been shown to localize within crescent micronuclei early during meiosis, within parental macronuclei and developing zygotic macronuclei, as well as in a cytoplasmic body known as the *conjusome* [159, 184, 189, 190]. The biological roles of Pdd1p in the crescent micronucleus and parental macronucleus are unknown, although a loss of expression during the early developmental stages is sufficient to block DNA elimination, thereby indicating that such Pdd1p localization is biologically relevant [190]. The localization of Pdd1p in the conjusome is thought to reflect the conjusome's role as a distribution center for the parental and developing macronuclei, or as a staging ground for Pdd1p transition from the parental macronuclei into the developing zygotic macronuclei later in conjugation [189]. Other proteins that are known to localize to the developing zygotic macronucleus later in conjugation, such as Lia1p, Lia3p, and Lia5p, also appear in the conjusome [162, 163]. In order to signal a transition from the parental macronucleus to the conjusome and the developing zygotic macronucleus, Pdd1p is phosphorylated up to four times [159]; this phosphorylation is lost as the conjugation proceeds, however, which may trigger DNA elimination body formation. The colocalization of Pdd1p with H3K9me3, H3K27me3, and IESs occurs in the developing zygotic macronucleus [31, 158, 159, 188]. Initially, the localization of Pdd1p is diffuse throughout the entire nucleus, but as the developing zygotic macronucleus matures the Pdd1p is concentrated into approximately 10 foci of average size 1 μm, termed *DNA elimination bodies* [184]. These Pdd1p-containing DNA elimination

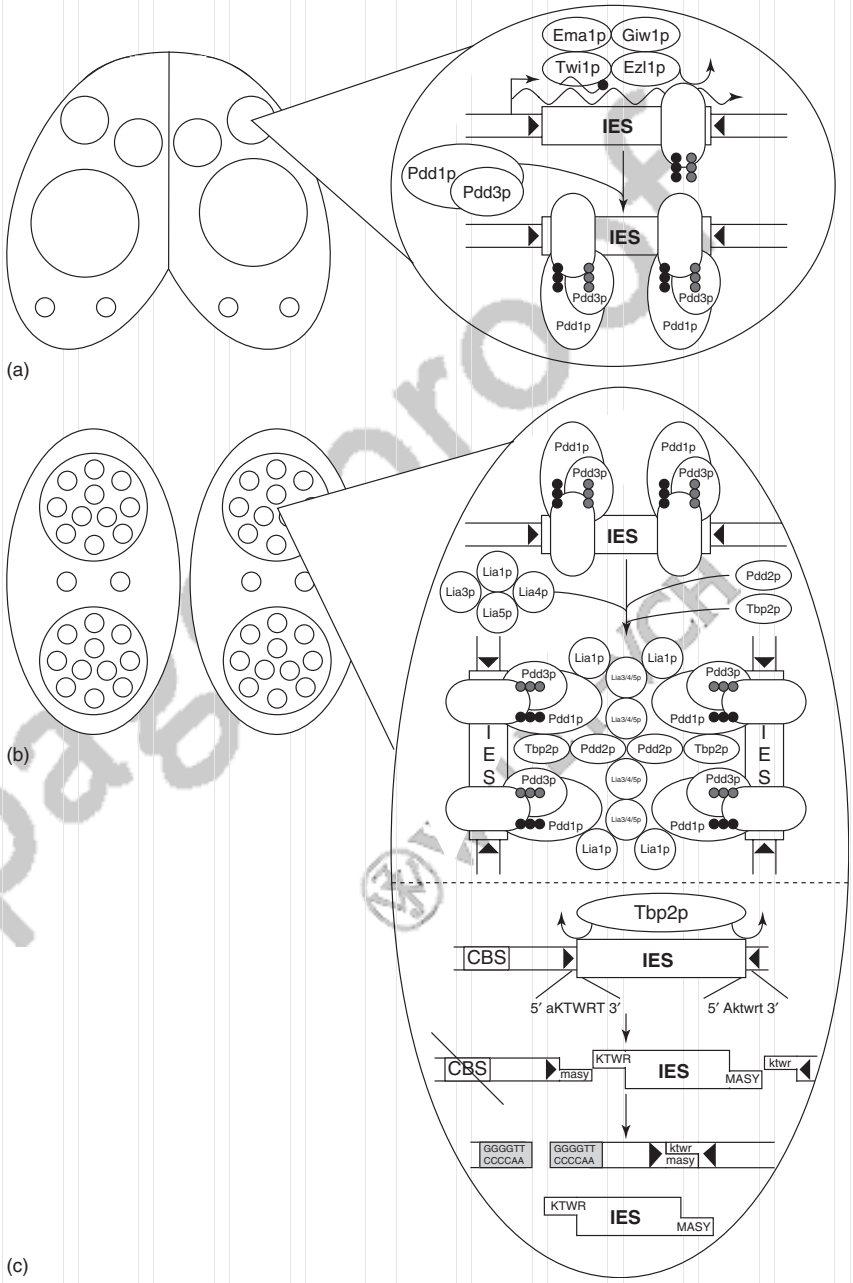

Fig. 7 RNAi-directed histone methylation on internal eliminated sequences (IESs) leads to their assembly into DNA elimination bodies and excision by a domesticated transposase, Tpb2p.

bodies also contain a number of other proteins including Pdd2p, Pdd3p, Lia1p, Lia3p, Lia4p, Lia5p, Tbp2p, and the Ezl1p complex (see Fig. 7b) (S.D. Taverna *et al.*, unpublished data) [14, 158, 161–163, 188]. Double knockouts of WAG1 and CNJB, LIA1 knockouts, EZL1 complex knockouts, PDD1 somatic knockouts, and PDD2 somatic knockouts each disrupt DNA elimination body formation (S.D. Taverna *et al.*, unpublished data) [149, 190, 191]. The tethering of Pdd1 to an artificial IES with no native histone methylation is also sufficient to direct DNA elimination, indicating that Pdd1p itself is sufficient to recruit its accessory proteins such as Tbp2p and to trigger DNA elimination [31].

The third chromodomain protein, Pdd3p, has been shown to bind strongly to H3K9me3, but not to H3K27me3, *in vitro* [31, 158]. Pdd3p localization is limited to the developing zygotic macronucleus where, like Pdd1p, it is initially diffuse but later condenses into the DNA elimination bodies [161]. The second programmed DNA degradation protein, Pdd2p, has no known homology [185] but demonstrates a localization that differs slightly from that of Pdd1p, by localizing only to the parental and developing macronuclei [185, 191]. Like Pdd1p and Pdd3p, the localization of Pdd2p in the developing zygotic macronucleus is initially diffuse until DNA elimination body formation. In a similar manner to Pdd1p, Pdd2p is phosphorylated once during transition from the parental macronucleus to the developing zygotic macronucleus [188]; again, this phosphorylation is removed immediately prior to DNA elimination body formation. PDD2 somatic knockouts are sufficient to cause the failure of cells to undergo DNA elimination and to complete conjugation which, like PDD1 somatic knockouts, may indicate a vital role for early localization in the parental macronucleus [191].

Other proteins have been found to influence DNA elimination body formation. For example, a diverse group of proteins that participated in this process were identified by their localization specifically to differentiating macronuclei, and thus were named localization in macronuclear anlagen (Lia) proteins [162, 163]. Lia1p, Lia4p, and Lia5p each play a role in DNA elimination body formation; typically, Lia5p contains a plant homeodomain (PHD) Zn Finger, while Lia4p contains a putative chromo shadow domain; otherwise, these proteins show no obvious homology to other known proteins. Of the Lia proteins, Lia1p is the best characterized, and localizes to both the conjusome and developing zygotic macronucleus [162]. Late in conjugation Lia1p is found in association with Pdd1p and IESs in DNA elimination bodies. Knockouts of LIA1 fail to eliminate IESs and complete conjugation, much like many other proteins in RNAi-directed DNA elimination. The preliminary characterization of Lia3p, Lia4p, and Lia5p has shown a diffuse localization early in the developing zygotic macronucleus, and later localization in DNA elimination bodies [163]. LIA3, LIA4, and LIA5 knockouts also fail to undergo DNA elimination and complete conjugation (A.W.-Y. Shieh *et al.*, unpublished data). While the role of these non-chromodomain proteins in RNAi-directed DNA elimination is not clear, it is possible that these proteins form a scaffold through which Pdd1p and Pdd3p, by interacting with specific classes of IESs, can be brought together to form the foci necessary for DNA elimination by the domesticated piggyBac transposase, Tbp2p.

Domesticated transposases have been shown to play an important role in a variety

of eukaryotic organisms, for example, RAG1/RAG2 recombinase in VDJ (variable, diverse, and joining) recombination in the human immune system [192, 193]. Ciliates appear to have domesticated transposases in order to facilitate the removal of repetitive sequences and IESs during conjugation [12–14]. In *T. thermophila*, Tbp2p – a piggyBac transposase homolog – is essential for removing IESs during conjugation (see Fig. 7c) [14]. An analysis of the TBP2 ORF shows homology with, and preservation of, the catalytic DDD motif in the domesticated piggyBac transposase in *P. tetraurelia*, PGM, and other piggyBac transposases in *H. sapiens*, *Xenopus* spp., and the moth, *Trichoplusia ni*. Tbp2p colocalizes with H3K9m3, H3K27me3, and Pdd1p in the developing zygotic macronucleus, before and after DNA elimination body formation. The knockdown of TBP2 using RNA hairpins does not inhibit Pdd1p association with H3K9me3 and H3K27me3 [14, 70]; however, TBP2 knockdown does inhibit DNA elimination body formation, IES removal, and completion of conjugation, thus implying an essential function downstream of Pdd1p and Pdd3p binding [14]. An *in vitro* analysis of the catalytic DDD motif of Tbp2p has shown that it is capable of cutting the consensus piggyBac cleavage sequence, 5′-TTAA-3′, as well as a variety of divergent sequences (see Fig. 7c) [14, 194, 195]. As noted above, Tbp2p cleavage produces a 4 bp 5′ overhang, which is not observed in mutants of the Tbp2p DDD catalytic motif [14, 195].

4.7 Chromosome Breakage in the Developing Somatic Nucleus

The epigenetic RNAi-directed DNA elimination process in *T. thermophila* is only a part of the global genome rearrangement that occurs in the developing zygotic macronucleus during conjugation. Chromosome breakage and differential chromosome amplification must also take place for this process to be complete [105–110, 196–199] (for a review, see Ref. [5]). This epigenomic process differs between *P. tetraurelia* and *T. thermophila*; in the former species the process seems to depend on RNAi-dependent DNA elimination machinery, whereas in *T. thermophila* chromosome breakage during conjugation is prompted by a conserved DNA sequence called the chromosome breakage sequence (CBS) [12, 104, 200]. Chromosome breakage and differential chromosome amplification have been shown to be essential for completion of conjugation, and are linked to RNAi-directed DNA elimination [33, 117, 149, 158, 191]. The conserved 15 bp CBS sequence is sufficient and necessary for chromosome breakage and telomere addition, which is blocked in CBS mutants (see Fig. 7c) [200–203]. Genomic analysis of the *T. thermophila* genome has shown that, with little variation, the CBS is present at all sites of chromosome breakage [107, 108]. Like IES excision, chromosome breakage appears to be dependent on the piggyBac transposase, Tbp2p [14].

5 Chromosome Fragmentation and Elimination of DNA during Conjugation in *Oxytricha*

The studies of DNA elimination in *P. tetraurelia* and *T. thermophila*, as described above, have revealed the role of sRNAs and long ncRNAs in remodeling genomes during development. They have also hinted to the possible mechanisms that allow

phenotypic traits to be propagated to the next generation. Whilst DNA elimination and chromosome fragmentation occur throughout the entire ciliate clade [5], it remains unclear whether RNAs play a similar role in more distantly related ciliates. However, recent investigations on these processes in a subgroup of ciliates known as *stichotrichs* has provided a definitive answer to this question [76]. Whilst the stichotrichs – which include the genera *Oxytricha* and *Stylonichia* – undergo DNA elimination and chromosome breakage, these processes are much more extreme and result in the elimination of more than 95% of the genome and of gene-sized mini-chromosomes of approximately 2 kb in size [204–207] (for a review, see Ref. [5]). A further complication in the understanding of these processes in *Oxytricha* and *Stylonichia* was the discovery of scrambled genes in the micronucleus [208–214]. Recent data acquired from *Oxytricha trifallax* have indicated that parental macronuclear ncRNA is able to direct the unscrambling of genes, DNA elimination, and chromosome breakage [76].

5.1 Gene Unscrambling and Domesticated Transposases in DNA Elimination and Chromosome Breakage

As in other ciliates, it seems likely that in stichotrichs DNA elimination – and, by extension, gene scrambling – in the micronucleus represent ways to prevent active transposons from appearing in the somatic macronuclear genome [5]. Yet, by scrambling the macronuclear-destined sequences (MDSs) of genes in the germline micronucleus, the stichotrichs ensure that DNA elimination must occur during sexual reproduction, in order to generate intact coding regions if progeny are to be viable. In this case, gene scrambling takes several forms, with some MDSs having undergone permutation in linear order, while others are even inverted with respect to the other MDSs to complicate the unscrambling process further (see Fig. 8b) [208–214].

To date, the scrambled genes discovered have included actin I, α telomere-binding protein (αTBP) and DNA polymerase α, with many more likely waiting to be discovered. Similar to *P. tetraurelia*, the MDSs of *O. trifallax* are bordered by short repeats (termed *pointers*) that may help direct gene unscrambling and DNA elimination [215] although, unfortunately, these repeats are too short to unambiguously accomplish this task. The discovery of parental macronuclear ncRNA during conjugation, and its role in gene unscrambling and DNA elimination, illuminates how these processes occur in *O. trifallax* and possibly in stichotrichs in general [76]. Subsequent RT-PCR analyses of RNA isolated from conjugating *O. trifallax* early and late in conjugation detected the presence of both sense and anti-sense ncRNAs. These ncRNAs, which are longer than mRNAs and contain telomeres, imply that the general transcription of all mini-chromosomes is initiated at the telomere sequence early during conjugation. RNAi against these ncRNAs during conjugation was sufficient to block the rearrangement of the target genes in the developing macronucleus. In order to validate the role of the parental macronucleus in producing these ncRNAs, Landweber and coworkers injected (into either the macronucleus or the cytoplasm) artificial DNA and RNA transcripts to a known gene (telomere-end-binding protein-β; TEBPβ), which contained different permutations of the MDSs. Upon the completion of conjugation, some TEBPβ genes containing

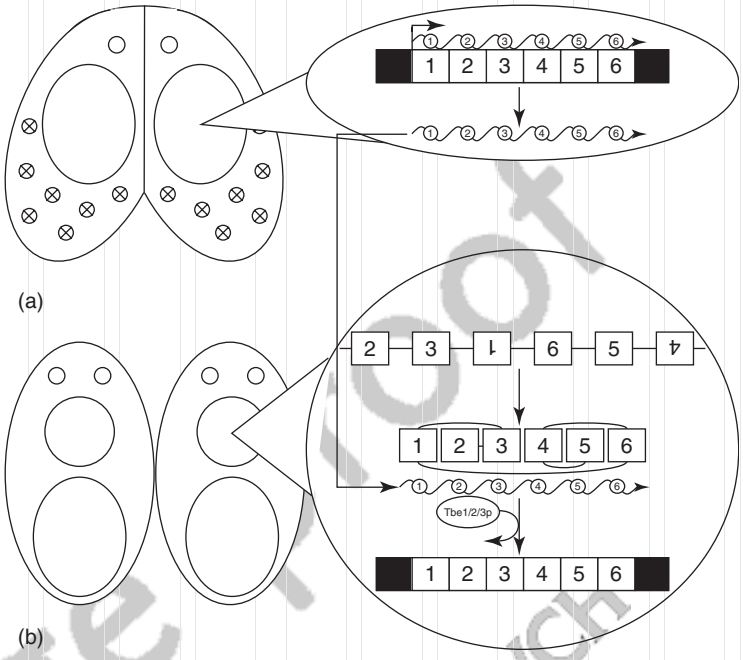

Fig. 8 Unscrambling of genes that are jumbled in the germ line genome of *O. trifallax* is guided by maternally produced template RNAs.

the alternative MDS order were found in the developing macronucleus, thus verifying the ability of artificial DNA in the parental macronucleus to produce ncRNA transcripts and to alter DNA elimination in the developing macronucleus.

Similar to *P. tetraurelia* and *T. thermophila*, a family of domesticated transposases has been found to play a role in gene unscrambling and DNA elimination in *O. trifallax* [12–14]. In this case, the transposases, termed telomere-bearing element 1 (TBE1), TBE2, and TBE3, belong to the TBE family of transposons and are not retained in the macronucleus after DNA elimination and chromosome breakage [13, 216, 217]. The triple knockdown of these transposases is sufficient to cause aberrant gene unscrambling and DNA elimination [13].

Taken together, these data have led to the proposal of a model (see Fig. 8) for gene unscrambling, DNA elimination, and chromosome breakage in *O. trifallax* [76]. At an early stage in conjugation, the bidirectional transcription of all mini-chromosomes in the parental macronucleus produces ncRNA. Following its appearance, the latter is transported to the developing macronucleus later during conjugation, where it directs gene unscrambling (if necessary) and the DNA elimination of IESs via a family of domesticated transposases (TBE1, TBE2, and TBE3) to produce a functional minichromosome in the developing macronucleus [13, 76]. Although the presence of sRNAs, a Piwi homolog, and heterochromatin marks have each been found in the stichotrich, *Stylonichia*, it remains to be seen whether any of these play a role in gene

unscrambling and DNA elimination in *Oxytricha*, *Stylonichia*, and in other stichotrichs, similar to *P. tetraurelia* and *T. thermophila* [33, 66, 85, 155, 158, 218]. It should be noted here that one point remains consistent in DNA elimination in all ciliates, namely the presence of ncRNA.

6 Perspective

Ciliates have long held the fascination of scientists, as they were among some of the first microorganisms to be studied in detail. Indeed, it was while developing *Paramecium* as a genetic model that Sonneborn first realized that many traits did not follow simple Mendelian rules of inheritance, and instead proposed that the cytoplasm might play a role in regulating the development of stable phenotypes. Although, today, molecular explanations for many of Sonneborn's observations have still not been provided, ciliates have nevertheless emerged as an important study system when investigating epigenetic mechanisms. Notably, their nuclear dimorphism has provided an informative biological context within which to uncover the mechanisms responsible for the differential regulation of homologous sequences. Ultimately, many of the mechanisms identified were shown to be common regulatory schemes used widely among eukaryotes. As an example, studies conducted in *Tetrahymena* provided the key data to show that transcriptional regulators acted by modifying chromatin [28, 29].

More recently, studies with ciliates have helped to reveal important roles for both long and short ncRNAs in mediating epigenetic regulation [33, 65, 66, 69, 76, 117, 118]. The majority of these new insights have resulted from studies aimed at elucidating the mechanisms that these organisms employ to remodel their somatic genomes during nuclear differentiation. An important paradigm that is now emerging from investigations of somatic nuclear differentiation of *Paramecium* and *Tetrahymena*, is that DNA rearrangement provides a means of genome surveillance, serving to remove the repetitive DNA from the transcriptionally active somatic nucleus, so that any potentially deleterious elements (e.g., transposons) which are silent in the germline cannot be spread. The ciliates identify this ''junk'' DNA by making an RNA copy of their germline genome during meiosis, thus processing bidirectional transcripts into an abundant class of sRNAs (scnRNAs) that can be used as the specificity factors to recognize germline-limited sequences [33, 66, 69, 74, 117, 118]. DNA rearrangement can be considered an innovative endpoint in the ciliate version of the piRNA pathway. In metazoans, the piRNA pathway serves to protect the germline from transposable elements via RNAi-directed silencing [120, 123, 125, 126, 219, 220] (see also Refs [63, 221]). In ciliates, the silencing of these sequences is permanent in the somatic genome, as they are eliminated during differentiation. It is clear, therefore, that a piRNA-mediated genome defense can serve as an evolutionary ancient mechanism.

The mechanistic connection between epigenetic silencing and DNA elimination is quite direct, as evidenced in *Tetrahymena*, where the germline-derived scnRNAs guide DNA rearrangements by directing heterochromatic modifications to the IESs. As noted in Sect. 4.3, both histone H3K9 and H3K27 methylation are established on IES chromatin at the

start of differentiation of the somatic macronucleus [31, 158], a discovery that was made immediately after RNAi was found to direct heterochromatin modification to silent genomic domains in *S. pombe* [112]. Taken together, the results of investigations in these unicellular models confirmed that RNAi-directed heterochromatin formation could provide a common means of instituting transcriptional gene silencing at homologous loci. Although the exact details of how sRNAs can direct chromatin modifications to specific sequences remain rather unclear, future studies in *Tetrahymena* and in other model systems will surely provide more detailed insights into these fundamental mechanisms.

What has become increasingly apparent is that ciliates have ways to communicate homologous sequence information between the germline and somatic genomes, from one generation to the next. As first revealed in studies of d48 *Paramecium* strains, the simple absence or presence of a DNA sequence in the parental somatic nucleus can "template" the same genome structure after DNA rearrangement of the new copy in the zygotic somatic genome [51, 49]. Evidence acquired from both *Paramecium* and *Tetrahymena* has indicated that this comparison of genome content is mediated by an interaction between scnRNAs and longer ncRNAs (see Figs 5 and 6), produced from the different nuclei [69, 75]. The syntheses and sites of action of these different ncRNAs exhibit both temporal and spatial separations, which allows the ncRNAs created in the parental somatic nucleus to block the action of scnRNAs, whereas those in the developing zygotic macronucleus will help to guide DNA elimination by interacting with the remaining scnRNA pool. Indeed, it is quite likely that these RNA-mediated genome comparisons that occur during development are responsible for some of the enigmatic examples of non-Mendelian inheritance, as originally described by Sonneborn.

The control of gene unscrambling in *Oxytricha*, via ncRNAs produced from the parental somatic genome, is perhaps the most intriguing phenomenon yet discovered [76]. As illustrated in Fig. 8, these ncRNAs are proposed to interact directly at the scrambled loci derived from the germline genome, and to guide the correct ordering of the mixed-up and inverted gene segments to ensure the assembly of a functional ORF. Whilst it is rather remarkable to consider that RNA could dramatically restructure the DNA of an organism, recent data acquired from this group of ciliates has further revealed that the copy number of the putative ncRNA templates can epigenetically regulate the copy number of the homologous chromosomes in the next generation [3, 4]. While the detailed mechanisms underlying these phenomena remain to be elucidated, these observations reveal nonetheless that homologous RNAs have a much-underappreciated capacity to influence gene expression and genome organization. Today, with much biology still awaiting illumination, the ciliated protozoa are clearly an important group of eukaryotes that are capable of revealing surprising modes of epigenetic regulation.

Cross References to Other Volumes – Topics and Subtopics

References

1 Sonneborn, T.M. (1937) Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA*, **23** (7), 378–385.

2 Philippe, H., Germot, A., Moreira, D. (2000) The new phylogeny of eukaryotes. *Curr. Opin. Genet. Dev.*, **10** (6), 596–601.

3 Nowacki, M., Haye, J.E., Fang, W., Vijayan, V., Landweber, L.F. (2010) RNA-mediated epigenetic regulation of DNA copy number. *Proc. Natl Acad. Sci. USA*, **107** (51), 22140–22144.

4 Heyse, G., Jonsson, F., Chang, W.J., Lipps, H.J. (2010) RNA-dependent control of gene amplification. *Proc. Natl Acad. Sci. USA*, **107** (51), 22134–22139.

5 Prescott, D.M. (1994) The DNA of ciliated protozoa. *Microbiol. Rev.*, **58** (2), 233–267.

6 Betermier, M. (2004) Large-scale genome remodelling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.*, **155** (5), 399–408.

7 Yao, M.C., Gorovsky, M.A. (1974) Comparison of the sequences of macro- and micronuclear DNA of *Tetrahymena pyriformis*. *Chromosoma*, **48** (1), 1–18.

8 Klobutcher, L.A., Herrick, G. (1995) Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic Acids Res.*, **23** (11), 2006–2013.

9 Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A., Williams, K. (1985) Mobile elements bounded by C4A4 telomeric repeats in *Oxytricha fallax*. *Cell*, **43**, 759–768.

10 Ribas-Aparicio, R.M., Sparkowski, J.J., Proulx, A.E., Mitchell, J.D., Klobutcher, L.A. (1987) Nucleic acid splicing events occur frequently during macronuclear development in the protozoan *Oxytricha nova* and involve the elimination of unique DNA. *Genes Dev.*, **1**, 323–336.

11 Jahn, C.L., Klobutcher, L.A. (2002) Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.*, **56**, 489–520.

12 Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., Betermier, M. (2009) PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.*, **23** (21), 2478–2483.

13 Nowacki, M., Higgins, B.P., Maquilan, G.M., Swart, E.C., Doak, T.G., Landweber, L.F. (2009) A functional role for transposases in a large eukaryotic genome. *Science*, **324** (5929), 935–938.

14 Cheng, C.Y., Vogt, A., Mochizuki, K., Yao, M.C. (2010) A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Mol. Biol. Cell*, **21** (10), 1753–1762.

15 Hannon, G.J. (2002) RNA interference. *Nature*, **418** (6894), 244–251.

16 Allis, C.D., Glover, C.V., Bowen, J.K., Gorovsky, M.A. (1980) Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, *Tetrahymena thermophila*. *Cell*, **20** (3), 609–617.

17 Hayashi, T., Hayashi, H., Fusauchi, Y., Iwai, K. (1984) *Tetrahymena* histone H3. Purification and two variant sequences. *J. Biochem. (Tokyo)*, **95** (6), 1741–1749.

18 Liu, X., Li, B., Gorovsky, M.A. (1996) Essential and nonessential histone H2A variants in *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **16** (8), 4305–4311.

19 Sugai, T., Hiwatashi, K. (1974) Cytological and autoradiographic studies of the micronucleus at meiotic prophase in *Tetrahymena pyriformis*. *J. Protozool.*, **21**, 542–548.

20 Martindale, D.W., Allis, C.D., Bruns, P.J. (1985) RNA and protein synthesis during meiotic prophase in *Tetrahymena thermophila*. *J. Protozool.*, **32** (4), 644–649.

21 Stargell, L.A., Bowen, J., Dadd, C.A., Dedon, P.C., Davis, M., Cook, R.G., Allis, C.D., Gorovsky, M.A. (1993) Temporal and spatial association of histone H2A variant hv1 with transcriptionally competent chromatin during nuclear development in *Tetrahymena thermophila*. *Genes Dev.*, **7**, 2641–2651.

22 Yu, L., Gorovsky, M.A. (1997) Constitutive expression, not a particular primary sequence, is the important feature of the H3 replacement variant hv2 in *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **17** (11), 6303–6310.

23 Shen, X., Yu, L., Weir, J.W., Gorovsky, M.A. (1995) Linker histones are not essential and affect chromatin condensation in vivo. *Cell*, **82** (1), 47–56.

24 Shen, X., Gorovsky, M.A. (1996) Linker histone H1 regulates specific gene expression

but not global transcription in vivo. *Cell*, **86** (3), 475–483.

25 Allfrey, V.G., Faulkner, R., Mirsky, A.E. (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl Acad. Sci. USA*, **51**, 786–794.

26 Ren, Q., Gorovsky, M.A. (2001) Histone H2A.Z acetylation modulates an essential charge patch. *Mol. Cell*, **7** (6), 1329–1335.

27 Ren, Q., Gorovsky, M.A. (2003) The nonessential H2A N-terminal tail can function as an essential charge patch on the H2A.Z variant N-terminal tail. *Mol. Cell. Biol.*, **23** (8), 2778–2789.

28 Brownell, J.E., Allis, C.D. (1995) An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in *Tetrahymena* macronuclei. *Proc. Natl Acad. Sci. USA*, **92** (14), 6364–6368.

29 Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., Allis, C.D. (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell*, **84** (6), 843–851.

30 Strahl, B.D., Ohba, R., Cook, R.G., Allis, C.D. (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl Acad. Sci. USA*, **96** (26), 14967–14972.

31 Taverna, S.D., Coyne, R.S., Allis, C.D. (2002) Methylation of histone h3 at lysine 9 targets programmed DNA elimination in *Tetrahymena*. *Cell*, **110** (6), 701–711.

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

33 Mochizuki, K., Fine, N.A., Fujisawa, T., Gorovsky, M.A. (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *Tetrahymena*. *Cell*, **110** (6), 689–699.

34 Wei, Y., Mizzen, C.A., Cook, R.G., Gorovsky, M.A., Allis, C.D. (1998) Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in *Tetrahymena*. *Proc. Natl Acad. Sci. USA*, **95** (13), 7480–7484.

35 Wei, Y., Yu, L., Bowen, J., Gorovsky, M.A., Allis, C.D. (1999) Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell*, **97** (1), 99–109.

36 Sonneborn, T.M. (1963) Does Pre-Formed Cell Structure Play an Essential Role in Cell Heredity? in: Allen, J.M. (Ed.) *The Nature of Biological Diversity*, McGraw-Hill, New York, pp. 165–221.

37 Beisson, J., Sonneborn, T.M. (1965) Cytoplasmic inheritance of the organization of the cell cortex in *Paramecium aurelia*. *Proc. Natl Acad. Sci. USA*, **53**, 275–282.

38 Sonneborn, T.M. (1977) Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells. *Annu. Rev. Genet.*, **11**, 349–367.

39 Sonneborn, T.M. (1974) Paramecium aurelia, in: King, R.C. (Ed.) *Handbook of Genetics: Plants, Plant Viruses and Protists*, Plenum Press, New York, pp. 469–594.

40 Sonneborn, T.M. (1947) Recent advances in the genetics of *Paramecium* and *Euplotes*. *Adv. Genet.*, **1**, 263–358.

41 Nanney, D.L. (1957) Mating-type inheritance at conjugation in variety 4 of *Paramecium aurelia*. *J. Protozool.*, **4**, 89–95.

42 Butzel, H.M. (1973) Abnormalities in nuclear behavior and mating type determination in cytoplasmically bridged exconjugants of doublet *Paramecium aurelia*. *J. Eukaryot. Microbiol.*, **20** (1), 140–142.

43 Koizumi, S. (1971) The cytoplasmic factor that fixes macronuclear mating type determination in *Paramecium aurelia*, syngen 4. *Genetics*, **68** (Suppl.), 34.

44 Sonneborn, T.M. (1948) The determination of hereditary antigenic differences in genically identical *Paramecium* cells. *Proc. Natl Acad. Sci. USA*, **34** (8), 413–418.

45 Brygoo, Y., Keller, A.-M. (1981) Genetic analysis of mating type differentiation in *Paramecium tetraurelia*. III. A mutation restricted to mating type E and affecting the determination of mating type. *Dev. Genet.*, **2**, 13–22.

46 Meyer, E., Keller, A.-M. (1996) A Mendelian mutation affecting mating-type determination also affects developmental genomic rearrangements in *Paramecium tetraurelia*. *Genetics*, **143**, 191–202.

47 Duharcourt, S., Butler, A., Meyer, E. (1995) Epigenetic self-regulation of developmental

excision of an internal eliminated sequence in *Paramecium tetraurelia*. *Genes Dev.*, **9**, 2065–2077.

48 Epstein, L.M., Forney, J.D. (1984) Mendelian and non-Mendelian mutations affecting surface antigen expression in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **4**, 1583–1590.

49 Preer, L.B., Hamilton, G., Preer, J.R. (1992) Micronuclear DNA from *Paramecium tetraurelia*: serotype 51A gene has internally eliminated sequences. *J. Protozool.*, **39**, 678–682.

50 Godiska, R., Aufderheide, K.J., Gilley, D., Hendrie, P., Fitzwater, T., Preer, L.B., Polisky, B., Preer, J.R.J. (1987) Transformation of *Paramecium* by microinjection of a cloned serotype gene. *Proc. Natl Acad. Sci. USA*, **84** (21), 7590–7594.

51 Koizumi, S., Kobayashi, S. (1989) Microinjection of plasmid DNA encoding the A surface antigen of *Paramecium tetraurelia* restores the ability to regenerate a wild-type macronucleus. *Mol. Cell. Biol.*, **9** (10), 4398–4401.

52 You, Y., Aufderheide, K., Morand, J., Rodkey, K., Forney, J. (1991) Macronuclear transformation with specific DNA fragments controls the content of the new macronuclear genome in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **11** (2), 1133–1137.

53 Jessop-Murray, H., Martin, L.D., Gilley, D., Preer, J.R., Jr, Polisky, B. (1991) Permanent rescue of a non-Mendelian mutation of *Paramecium* by microinjection of specific DNA sequences. *Genetics*, **129** (3), 727–734.

54 Scott, J.M., Leeck, C.L., Forney, J.D. (1994) Analysis of the micronuclear B type surface protein gene in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **22** (23), 5079–5084.

55 Scott, J.M., Mikami, K., Leeck, C.L., Forney, J.D. (1994) Non-Mendelian inheritance of macronuclear mutations is gene specific in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **14** (4), 2479–2484.

56 Kim, C.S., Preer, J.R. Jr, Polisky, B. (1994) Identification of DNA segments capable of rescuing a non-Mendelian mutant in *Paramecium*. *Genetics*, **136** (4), 1325–1328.

57 Forney, J.D., Yantiri, F., Mikami, K. (1996) Developmentally controlled rearrangement of surface protein genes in *Paramecium tetraurelia*. *J. Eukaryot. Microbiol.*, **43** (6), 462–467.

58 Napoli, C., Lemieux, C., Jorgensen, R. (1990) Introduction of a chimeric chalcone synthase gene into *Petunia* results in reversible co-suppression of homologous genes in trans. *Plant Cell*, **2** (4), 279–289.

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

60 Ruiz, F., Vayssié, L., Klotz, C., Sperling, L., Madeddu, L. (1998) Homology-dependent gene silencing in *Paramecium*. *Mol. Biol. Cell*, **9** (4), 931–943.

61 Galvani, A., Sperling, L. (2001) Transgene-mediated post-transcriptional gene silencing is inhibited by 3' non-coding sequences in *Paramecium*. *Nucleic Acids Res.*, **29** (21), 4387–4394.

62 Lejeune, E., Allshire, R.C. (2011) Common ground: small RNA programming and chromatin modifications. *Curr. Opin. Cell Biol.*, **23** (3), 258–265.

63 Malone, C.D., Hannon, G.J. (2009) Small RNAs as guardians of the genome. *Cell*, **136** (4), 656–668.

64 Carthew, R.W., Sontheimer, E.J. (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell*, **136** (4), 642–655.

65 Lee, S.R., Collins, K. (2006) Two classes of endogenous small RNAs in *Tetrahymena thermophila*. *Genes Dev.*, **20** (1), 28–33.

66 Lepere, G., Nowacki, M., Serrano, V., Gout, J.F., Guglielmi, G., Duharcourt, S., Meyer, E. (2009) Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **37** (3), 903–915.

67 Galvani, A., Sperling, L. (2002) RNA interference by feeding in *Paramecium*. *Trends Genet.*, **18** (1), 11–12.

68 Garnier, O., Serrano, V., Duharcourt, S., Meyer, E. (2004) RNA-mediated programming of developmental genome rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **24** (17), 7370–7379.

69 Lepere, G., Betermier, M., Meyer, E., Duharcourt, S. (2008) Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.*, **22** (11), 1501–1512.

70 Howard-Till, R.A., Yao, M.C. (2006) Induction of gene silencing by hairpin RNA expression in *Tetrahymena thermophila* reveals a second small RNA pathway. *Mol. Cell. Biol.*, **26** (23), 8731–8742.

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

72 Lee, S.R., Collins, K. (2007) Physical and functional coupling of RNA-dependent RNA polymerase and Dicer in the biogenesis of endogenous siRNAs. *Nat. Struct. Mol. Biol.*, **14** (7), 604–610.

73 Marker, S., Le Mouel, A., Meyer, E., Simon, M. (2010) Distinct RNA-dependent RNA polymerases are required for RNAi triggered by double-stranded RNA versus truncated transgenes in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **38** (12), 4092–4107.

74 Chalker, D.L., Yao, M.C. (2001) Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev.*, **15** (10), 1287–1298.

75 Aronica, L., Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W., Loidl, J., Pearlman, R.E., Gorovsky, M.A., Mochizuki, K. (2008) Study of an RNA helicase implicates small RNA-noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.*, **22** (16), 2228–2241.

76 Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., Landweber, L.F. (2008) RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature*, **451** (7175), 153–158.

77 Yao, M.C., Fuller, P., Xi, X. (2003) Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, **300** (5625), 1581–1584.

78 Duharcourt, S., Keller, A., Meyer, E. (1998) Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell Biol.*, **18** (12), 7075–7085.

79 Aury, J.M., Jaillon, O., Duret, L., Noel, B., Jubin, C.,, Porcel, B.M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J., Wincker, P. (2006) Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444** (7116), 171–178.

80 Dippell, R.V. (1954) A preliminary report on the chromosomal constitution of certain variety 4 races of *Paramecium aurelia*. *Caryologia*, **6**, 1109–1111.

81 Blaszczyk, J., Tropea, J.E., Bubunenko, M., Routzahn, K.M., Waugh, D.S., Court, D.L., Ji, X. (2001) Crystallographic and modeling studies of RNase III suggest a mechanism for double-stranded RNA cleavage. *Structure*, **9** (12), 1225–1236.

82 Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., Kim, V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18** (24), 3016–3027.

83 Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E., Filipowicz, W. (2004) Single processing center models for human Dicer and bacterial RNase III. *Cell*, **118** (1), 57–68.

84 Kim, V.N., Han, J., Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10** (2), 126–139.

85 Bouhouche, K., Gout, J.-F., Kapusta, A., Bétermier, M., Meyer, E. (2011) Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.*, **39** (10), 4249–4264.

86 Nowacki, M., Zagorski-Ostoja, W., Meyer, E. (2005) Nowa1p and Nowa2p: novel putative RNA binding proteins involved in trans-nuclear crosstalk in *Paramecium tetraurelia*. *Curr. Biol.*, **15** (18), 1616–1628.

87 Meller, V.H., Wu, K.H., Roman, G., Kuroda, M.I., Davis, R.L. (1997) roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell*, **88**, 445–457.

88 Amrein, H., Axel, R. (1997) Genes expressed in neurons of adult male *Drosophila*. *Cell*, **88**, 459–469.

89 Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., Willard, H.F. (1992) The human XIST

gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, **71**, 527–542.

90 Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., Rastan, S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**, 515–526.

91 Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., Willard, H.F. (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349** (6304), 38–44.

92 Bartolomei, M.S., Zemel, S., Tilghman, S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature*, **351** (6322), 153–155.

93 Ponting, C.P., Oliver, P.L., Reik, W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136** (4), 629–641.

94 Klobutcher, L.A., Herrick, G. (1997) Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.*, **56**, 1–62.

95 Gratias, A., Betermier, M. (2003) Processing of double-strand breaks is involved in the precise excision of *Paramecium* internal eliminated sequences. *Mol. Cell Biol.*, **23** (20), 7152–7162.

96 Mayer, K.M., Mikami, K., Forney, J.D. (1998) A mutation in *Paramecium tetraurelia* reveals functional and structural features of developmentally excised DNA elements. *Genetics*, **148**, 139–149.

97 Mayer, K., Forney, J. (1999) A mutation in the flanking 5′-TA-3′ dinucleotide prevents excision of an internal eliminated sequence from the *Paramecium tetraurelia* genome. *Genetics*, **151** (2), 597–604.

98 Gratias, A., Lepere, G., Garnier, O., Rosa, S., Duharcourt, S., Malinsky, S., Meyer, E., Betermier, M. (2008) Developmentally programmed DNA splicing in *Paramecium* reveals short-distance crosstalk between DNA cleavage sites. *Nucleic Acids Res.*, **36** (10), 3244–3251.

99 Ruiz, F., Krzywicka, A., Klotz, C., Keller, A., Cohen, J., Koll, F., Balavoine, G., Beisson, J. (2000) The SM19 gene, required for duplication of basal bodies in *Paramecium*, encodes a novel tubulin, eta-tubulin. *Curr. Biol.*, **10** (22), 1451–1454.

100 Elick, T.A., Bauser, C.A., Fraser, M.J. (1996) Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase. *Genetica*, **98** (1), 33–41.

101 Mitra, R., Fain-Thornton, J., Craig, N.L. (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.*, **27** (7), 1097–1109.

102 Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., Betermier, M. (2011) Highly precise and developmentally programmed genome assembly in *Paramecium* requires ligase IV-dependent end joining. *PLoS Genet.*, **7** (4), e1002049.

103 Duharcourt, S., Lepere, G., Meyer, E. (2009) Developmental genome rearrangements in ciliates: a natural genomic subtraction mediated by non-coding transcripts. *Trends Genet.*, **25** (8), 344–350.

104 Le Mouel, A., Butler, A., Caron, F., Meyer, E. (2003) Developmentally regulated chromosome fragmentation linked to imprecise elimination of repeated sequences in paramecia. *Eukaryot. Cell*, **2** (5), 1076–1090.

105 Altschuler, M.I., Yao, M.C. (1985) Macronuclear DNA of *Tetrahymena thermophila* exists as defined subchromosomal-sized molecules. *Nucleic Acids Res.*, **13** (16), 5817–5831.

106 Woodard, J., Kaneshiro, E., Gorovsky, M.A. (1972) Cytochemical studies on the problem of macronuclear subnuclei in *Tetrahymena*. *Genetics*, **70** (2), 251–260.

107 Cassidy-Hanley, D., Bisharyan, Y., Fridman, V., Gerber, J., Lin, C., Orias, E., Orias, J.D., Ryder, H., Vong, L., Hamilton, E.P. (2005) Genome-wide characterization of *Tetrahymena thermophila* chromosome breakage sites. II. Physical and genetic mapping. *Genetics*, **170** (4), 1623–1631.

108 Hamilton, E., Bruns, P., Lin, C., Merriam, V., Orias, E., Vong, L., Cassidy-Hanley, D. (2005) Genome-wide characterization of *Tetrahymena thermophila* chromosome breakage sites. I. Cloning and identification of functional sites. *Genetics*, **170** (4), 1611–1621.

109 Doerder, F.P., Deak, J.C., Lief, J.H. (1992) Rate of phenotypic assortment in *Tetrahymena thermophila*. *Dev. Genet.*, **13** (2), 126–132.

110 Conover, R.K., Brunk, C.F. (1986) Macronuclear DNA molecules of *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **6** (3), 900–905.

111 Allis, C.D., Jenuwein, T., Reinberg, D., Caparros, M.-L.A.E. (Eds) (2007) *Epigenetics*, Cold Spring Harbor Press, Cold Spring Harbor.

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

113 Reinhart, B.J., Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science*, **297** (5588), 1831.

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

115 Brown, D.D., Dawid, I.B. (1968) Specific gene amplification in oocytes. *Science*, **160**, 272–280.

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

117 Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., Chalker, D.L. (2005) Germ line transcripts are processed by a Dicer-like protein that is essential for developmentally programmed genome rearrangements of *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **25** (20), 9151–9164.

118 Mochizuki, K., Gorovsky, M.A. (2005) A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. *Genes Dev.*, **19** (1), 77–89.

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

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

121 Carmell, M.A., Girard, A., van de Kant, H.J., Bourc'his, D., Bestor, T.H., de Rooij, D.G., Hannon, G.J. (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev. Cell*, **12** (4), 503–514.

122 Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., Matthews, N., Berezikov, E., Ketting, R.F., Tavare, S., Miska, E.A. (2008) Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol. Cell*, **31** (1), 79–90.

123 Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., Siomi, M.C. (2007) A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in *Drosophila*. *Science*, **315** (5818), 1587–1590.

124 Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., Asada, N., Kojima, K., Yamaguchi, Y., Ijiri, T.W., Hata, K., Li, E., Matsuda, Y., Kimura, T., Okabe, M., Sakaki, Y., Sasaki, H., Nakano, T. (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.*, **22** (7), 908–917.

125 Saito, K., Nishida, K.M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., Siomi, M.C. (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes Dev.*, **20** (16), 2214–2222.

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

127 Martindale, D.W., Allis, C.D., Bruns, P. (1982) Conjugation in *Tetrahymena thermophila*: a temporal analysis of cytological stages. *Exp. Cell Res.*, **140**, 227–236.

128 Ray, C. (1956) Meiosis and nuclear behavior in *Tetrahymena pyriformis*. *J. Eukaryot. Microbiol.*, **3** (2), 88–96.

129 Wenkert, D., Allis, C.D. (1984) Timing of the appearance of macronuclear-specific histone variant hv1 and gene expression in developing new macronuclei of *Tetrahymena thermophila*. *J. Cell Biol.*, **98** (6), 2107–2117.

130 Stargell, L.A., Gorovsky, M.A. (1994) TATA-binding protein and nuclear differentiation in *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **14** (1), 723–734.

131 Mochizuki, K., Gorovsky, M.A. (2004) RNA polymerase II localizes in *Tetrahymena thermophila* meiotic micronuclei when micronuclear transcription associated with genome rearrangement occurs. *Eukaryot. Cell*, **3** (5), 1233–1240.

132 Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., Jacobsen, S.E. (2006) Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.*, **38** (6), 721–725.

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

134 Robertson, H.D., Dunn, J.J. (1975) Ribonucleic acid processing activity of *Escherichia coli* ribonuclease III. *J. Biol. Chem.*, **250** (8), 3050–3056.

135 Schweitz, H., Ebel, J.P. (1971) A study of the mechanism of action of *E. coli* ribonuclease 3. *Biochimie*, **53** (5), 585–593.

136 Crouch, R.J. (1974) Ribonuclease 3 does not degrade deoxyribonucleic acid-ribonucleic acid hybrids. *J. Biol. Chem.*, **249** (4), 1314–1316.

137 Mochizuki, K., Gorovsky, M.A. (2004) Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.*, **18** (17), 2068–2073.

138 Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, **404** (6775), 293–296.

139 Noto, T., Kurth, H.M., Kataoka, K., Aronica, L., DeSouza, L.V., Siu, K.W., Pearlman, R.E., Gorovsky, M.A., Mochizuki, K. (2010) The *Tetrahymena* argonaute-binding protein Giw1p directs a mature argonaute-siRNA complex to the nucleus. *Cell*, **140** (5), 692–703.

140 Kurth, H.M., Mochizuki, K. (2009) 2′-O-methylation stabilizes Piwi-associated small RNAs and ensures DNA elimination in *Tetrahymena*. *RNA*, **15** (4), 675–685.

141 Horwich, M.D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., Zamore, P.D. (2007) The *Drosophila* RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr. Biol.*, **17** (14), 1265–1272.

142 Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., Siomi, M.C. (2007) Pimet, the *Drosophila* homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. *Genes Dev.*, **21** (13), 1603–1608.

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

144 Kirino, Y., Mourelatos, Z. (2007) The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *RNA*, **13** (9), 1397–1401.

145 Kirino, Y., Vourekas, A., Kim, N., de Lima Alves, F., Rappsilber, J., Klein, P.S., Jongens, T.A., Mourelatos, Z. (2010) Arginine methylation of vasa protein is conserved across phyla. *J. Biol. Chem.*, **285** (11), 8148–8154.

146 Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T.W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H., Matsuda, Y., Nakano, T. (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development*, **131** (4), 839–849.

147 Thomson, T., Liu, N., Arkov, A., Lehmann, R., Lasko, P. (2008) Isolation of new polar granule components in *Drosophila* reveals P body and ER associated proteins. *Mech. Dev.*, **125** (9-10), 865–873.

148 Arkov, A.L., Ramos, A. (2010) Building RNA-protein granules: insight from the germline. *Trends Cell Biol.*, **20** (8), 482–490.

149 Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W., Pearlman, R.E., Mochizuki, K., Gorovsky, M.A. (2009) Two GW repeat proteins interact with *Tetrahymena thermophila* argonaute and promote genome rearrangement. *Mol. Cell. Biol.*, **29** (18), 5020–5030.

150 He, X.J., Hsu, Y.F., Zhu, S., Wierzbicki, A.T., Pontes, O., Pikaard, C.S., Liu, H.L., Wang, C.S., Jin, H., 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.

151 Chekulaeva, M., Parker, R., Filipowicz, W. (2010) The GW/WG repeats of *Drosophila* GW182 function as effector motifs for miRNA-mediated repression. *Nucleic Acids Res.*, **38** (19), 6673–6683.

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

153 Chalker, D.L., Fuller, P., Yao, M.C. (2005) Communication between parental and developing genomes during *Tetrahymena* nuclear differentiation is likely mediated by homologous RNAs. *Genetics*, **169** (1), 149–160.

154 Chalker, D.L., Yao, M.C. (1996) Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol. Cell. Biol.*, **16** (7), 3658–3667.

155 Liu, Y., Mochizuki, K., Gorovsky, M.A. (2004) Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in *Tetrahymena*. *Proc. Natl Acad. Sci. USA*, **101** (6), 1679–1684.

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

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

158 Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., Allis, C.D. (2007) RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in *Tetrahymena*. *Genes Dev.*, **21** (12), 1530–1545.

159 Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., Allis, C.D. (1996) Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in *Tetrahymena*. *Cell*, **87** (1), 75–84.

160 Coyne, R., Nikiforov, M.A., Smothers, J.F., Allis, C.D., Yao, M.C. (1999) Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. *Mol. Cell*, **4** (5), 865–872.

161 Nikiforov, M.A., Gorovsky, M.A., Allis, C.D. (2000) A novel chromodomain protein, pdd3p, associates with internal eliminated sequences during macronuclear development in *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **20** (11), 4128–4134.

162 Rexer, C.H., Chalker, D.L. (2007) Lia1p, a novel protein required during nuclear differentiation for genome-wide DNA rearrangements in *Tetrahymena thermophila*. *Eukaryot. Cell*, **6** (8), 1320–1329.

163 Yao, M.C., Yao, C.H., Halasz, L.M., Fuller, P., Rexer, C.H., Wang, S.H., Jain, R., Coyne, R.S., Chalker, D.L. (2007) Identification of novel chromatin-associated proteins involved in programmed genome rearrangements in *Tetrahymena*. *J. Cell Sci.*, **120** (Pt 12), 1978–1989.

164 Cam, H.P., Sugiyama, T., Chen, E.S., Chen, X., FitzGerald, P.C., Grewal, S.I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.*, **37** (8), 809–819.

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

166 Ivanova, A.V., Bonaduce, M.J., Ivanov, S.V., Klar, A.J.S. (1998) The chromo and SET domains of the Clr4 protein are essential for silencing in fission yeast. *Nat. Genet.*, **19** (2), 192–195.

167 Jackson, J.P., Lindroth, A.M., Cao, X., Jacobsen, S.E. (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature*, **416** (6880), 556–560.

168 Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E.M., Coupland, G. (1997) A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature*, **386** (6620), 44–51.

169 Fong, Y., Bender, L., Wang, W., Strome, S. (2002) Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. *Science*, **296** (5576), 2235–2238.

170 van der Lugt, N.M., Domen, J., Linders, K., van Roon, M., Robanus-Maandag, E., te Riele, H., van der Valk, M., Deschamps, J., Sofroniew, M., van Lohuizen, M., Berns, A. (1994) Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev.*, **8** (7), 757–769.

171 Cao, R., Tsukada, Y., Zhang, Y. (2005) Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell*, **20** (6), 845–854.

172 Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., Zhang, Y. (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, **431** (7010), 873–878.

173 Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A., Verrijzer, C.P. (2008) dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.*, **22** (20), 2799–2810.

174 Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., Bardwell, V.J. (2006) Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol. Cell. Biol.*, **26** (18), 6880–6889.

175 Sanchez, C., Sanchez, I., Demmers, J.A., Rodriguez, P., Strouboulis, J., Vidal, M. (2007) Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. *Mol. Cell. Proteomics*, **6** (5), 820–834.

176 Simon, J.A., Kingston, R.E. (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.*, **10** (10), 697–708.

177 Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298** (5595), 1039–1043.

178 Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., Pirrotta, V. (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, **111** (2), 185–196.

179 Muller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., Simon, J.A. (2002) Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell*, **111** (2), 197–208.

180 Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., Reinberg, D. (2002) Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.*, **16** (22), 2893–2905.

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

182 Ekwall, K., Javerzat, J.P., Lorentz, A., Schmidt, H., Cranston, G., Allshire, R. (1995) The chromodomain protein Swi6: a key component at fission yeast centromeres. *Science*, **269** (5229), 1429–1431.

183 Lorentz, A., Ostermann, K., Fleck, O., Schmidt, H. (1994) Switching gene swi6, involved in repression of silent mating-type loci in fission yeast, encodes a homologue of chromatin-associated proteins from *Drosophila* and mammals. *Gene*, **143** (1), 139–143.

184 Madireddi, M.T., Davis, M.C., Allis, C.D. (1994) Identification of a novel polypeptide involved in the formation of DNA-containing vesicles during macronuclear development in *Tetrahymena*. *Dev. Biol.*, **165** (2), 418–431.

185 Smothers, J.F., Mizzen, C.A., Tubbert, M.M., Cook, R.G., Allis, C.D. (1997) Pdd1p associates with germline-restricted chromatin and a second novel anlagen-enriched protein in developmentally programmed DNA elimination structures. *Development*, **124** (22), 4537–4545.

186 Yao, M.C., Choi, J., Yokoyama, S., Austerberry, C.F., Yao, C.H. (1984) DNA elimination in *Tetrahymena*: a developmental process involving extensive breakage and rejoining of DNA at defined sites. *Cell*, **36** (2), 433–440.

187 Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., Elgin, S.C. (1990) Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA*, **87** (24), 9923–9927.

188 Smothers, J.F., Madireddi, M.T., Warner, F.D., Allis, C.D. (1997) Programmed DNA degradation and nucleolar biogenesis occur in distinct organelles during macronuclear development in *Tetrahymena*. *J. Euk. Microbiol.*, **44**, 79–88.

189 Janetopoulos, C., Cole, E., Smothers, J.F., Allis, C.D., Aufderheide, K.J. (1999) The conjusome: a novel structure in *Tetrahymena* found only during sexual reorganization. *J. Cell Sci.*, **112** (Pt 7), 1003–1011.

190 Coyne, R.S., Nikiforov, M.A., Smothers, J.F., Allis, C.D., Yao, M.C. (1999) Parental expression of the chromodomain protein Pdd1p is required for completion of programmed DNA elimination and nuclear differentiation. *Mol. Cell*, **4** (5), 865–872.

191 Nikiforov, M., Smothers, J., Gorovsky, M., Allis, C. (1999) Excision of micronuclear-specific DNA requires parental expression of Pdd2p and occurs independently from DNA replication in *Tetrahymena thermophila*. *Genes Dev.*, **13** (21), 2852–2862.

192 Agrawal, A., Eastman, Q.M., Schatz, D.G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*, **394** (6695), 744–751.

193 Hiom, K., Melek, M., Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell*, **94** (4), 463–470.

194 Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E., Fraser, M.J. (1989) Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, **172** (1), 156–169.

195 Saveliev, S.V., Cox, M.M. (1996) Developmentally programmed DNA deletion in *Tetrahymena thermophila* by a transposition-like reaction pathway. *EMBO J.*, **15** (11), 2858–2869.

196 Doerder, F.P., Debault, L.E. (1975) Cytofluorimetric analysis of nuclear DNA during meiosis, fertilization and macronuclear development in the ciliate *Tetrahymena pyriformis*, syngen 1. *J. Cell Sci.*, **17** (3), 471–493.

197 Yao, M.C. (1981) Ribosomal RNA gene amplification in *Tetrahymena* may be associated with chromosome breakage and DNA elimination. *Cell*, **24** (3), 765–774.

198 Yao, M.C., Yao, C.H. (1981) Repeated hexanucleotide C-C-C-C-A-A is present near free ends of macronuclear DNA of *Tetrahymena*. *Proc. Natl Acad. Sci. USA*, **78** (12), 7436–7439.

199 Yao, M.C., Kimmel, A.R., Gorovsky, M.A. (1974) A small number of cistrons for ribosomal RNA in the germinal nucleus of a eukaryote, *Tetrahymena pyriformis*. *Proc. Natl Acad. Sci. USA*, **71** (8), 3082–3086.

200 Yao, M.C., Zheng, K., Yao, C.H. (1987) A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in *Tetrahymena*. *Cell*, **48** (5), 779–788.

201 Fan, Q., Yao, M.-C. (1996) New telomere formation coupled with site-specific chromosome breakage in *Tetrahymena thermophila*. *Mol. Cell. Biol.*, **16** (3), 1267–1274.

202 Kapler, G.M., Blackburn, E.H. (1994) A weak germ-line excision mutation blocks developmentally controlled amplification of the rDNA minichromosome of *Tetrahymena thermophila*. *Genes Dev.*, **8** (1), 84–95.

203 Fan, Q., Yao, M.C. (2000) A long stringent sequence signal for programmed chromosome breakage in *Tetrahymena thermophila*. *Nucleic Acids Res.*, **28** (4), 895–900.

204 Swanton, M.T., Heumann, J.M., Prescott, D.M. (1980) Gene-sized DNA molecules of the macronuclei in three species of hypotrichs: size distributions and absence of nicks. DNA of ciliated protozoa. VIII. *Chromosoma*, **77** (2), 217–227.

205 Swanton, M.T., Greslin, A.F., Prescott, D.M. (1980) Arrangement of coding and non-coding sequences in the DNA molecules coding for rRNAs in *Oxytricha*

sp. DNA of ciliated protozoa. VII. *Chromosoma*, **77** (2), 203–215.

206 Wunning, I.U., Lipps, H.J. (1983) A transformation system for the hypotrichous ciliate *Stylonychia mytilus*. *EMBO J.*, **2** (10), 1753–1757.

207 Steinbruck, G. (1983) Over-amplification of genes in macronuclei of hypotrichous ciliates. *Chromosoma*, **88**, 156–163.

208 Greslin, A.F., Loukin, S.H., Oka, Y., Prescott, D.M. (1988) An analysis of the macronuclear actin genes of *Oxytricha*. *DNA*, **7** (8), 529–536.

209 Greslin, A.F., Prescott, D.M., Oka, Y., Loukin, S.H., Chappell, J.C. (1989) Reordering of nine exons is necessary to form a functional actin gene in *Oxytricha nova*. *Proc. Natl Acad. Sci. USA*, **86** (16), 6264–6268.

210 Mitcham, J.L., Lynn, A.J., Prescott, D.M. (1992) Analysis of a scrambled gene: the gene encoding alpha-telomere-binding protein in *Oxytricha nova*. *Genes Dev.*, **6** (5), 788–800.

211 Prescott, D.M., Greslin, A.F. (1992) Scrambled actin I gene in the micronucleus of *Oxytricha nova*. *Dev. Genet.*, **13** (1), 66–74.

212 DuBois, M., Prescott, D.M. (1995) Scrambling of the actin I gene in two *Oxytricha* species. *Proc. Natl Acad. Sci. USA*, **92** (9), 3888–3892.

213 Hoffman, D.C., Prescott, D.M. (1997) Evolution of internal eliminated segments and scrambling in the micronuclear gene encoding DNA polymerase alpha in two *Oxytricha* species. *Nucleic Acids Res.*, **25** (10), 1883–1889.

214 Prescott, J.D., DuBois, M.L., Prescott, D.M. (1998) Evolution of the scrambled germline gene encoding alpha-telomere binding protein in three hypotrichous ciliates. *Chromosoma*, **107** (5), 293–303.

215 Landweber, L.F., Kuo, T.-C., Curtis, E.A. (2000) Evolution and assembly of an extremely scrambled gene. *Proc. Natl Acad. Sci. USA*, **97** (7), 3298–3303.

216 Doak, T.G., Doerder, F.P., Jahn, C.L., Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. *Proc. Natl Acad. Sci. USA*, **91** (3), 942–946.

217 Hunter, D.J., Williams, K., Cartinhour, S., Herrick, G. (1989) Precise excision of telomere-bearing transposons during *Oxytricha fallax* macronuclear development. *Genes Dev.*, **3** (12B), 2101–2112.

218 Juranek, S.A., Rupprecht, S., Postberg, J., Lipps, H.J. (2005) snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates. *Eukaryot. Cell*, **4** (11), 1934–1941.

219 Bourc'his, D., Bestor, T.H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, **431** (7004), 96–99.

220 Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M., Sasaki, H. (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.*, **16** (19), 2272–2280.

221 Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A. (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.*, **12** (4), 246–258.