Programmed Genome Rearrangements in Tetrahymena
MENG-CHAO YAO, JU-LAN CHAO, and CHAO-YIN CHENG
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

ABSTRACT Ciliates are champions in programmed genome rearrangements. They carry out extensive restructuring during differentiation to drastically alter the complexity, relative copy number, and arrangement of sequences in the somatic genome. This chapter focuses on the model ciliate Tetrahymena, perhaps the simplest and best-understood ciliate studied. It summarizes past studies on various genome rearrangement processes and describes in detail the remarkable progress made in the past decade on the understanding of DNA deletion and other processes. The process occurs at thousands of specific sites to remove defined DNA segments that comprise roughly one-third of the genome including all transposons. Interestingly, this DNA rearranging process is a special form of RNA interference. It involves the production of double-stranded RNA and small RNA that guides the formation of heterochromatin. A domesticated piggyBac transposase is believed to cut off the marked chromatin, and the retained sequences are joined together through nonhomologous end-joining processes. Many of the proteins and DNA players involved have been analyzed and are described. This link provides possible explanations for the evolution, mechanism, and functional roles of the process. The article also discusses the interactions between parental and progeny somatic nuclei that affect the selection of sequences for deletion, and how the specific deletion boundaries are determined after heterochromatin marking.

INTRODUCTION
This chapter reviews recent studies on the remarkable phenomenon of programmed DNA rearrangements in ciliated protozoa, focusing primarily on the species Tetrahymena thermophila. The phenomenon occurs widely among ciliates, a diverse group of single-celled eukaryotes. It varies significantly in mechanistic detail among the species that have been described, chiefly Tetrahymena, Paramecium, Euplotes, Stylonychia, and Oxytricha. Readers are referred to other chapters in this volume for studies in Paramecium and Oxytricha. Tetrahymena displays perhaps the simplest version of these DNA rearrangements and is the easiest to grow and manipulate in the laboratory, hence offering excellent opportunities for in-depth understanding. Since the publication of Mobile DNA II (1), significant progress has brought about fundamental changes in our understanding. Among other things, clear links have now been established between these processes and RNA interference (RNAi) and transposon domestication. This chapter will concentrate on progress made during this period, with a brief summary of earlier work to provide an introduction.

“Chromatin diminution” was first reported by Boveri in 1887 (2) and has since been observed in a number of diverse eukaryotic groups including ciliates, nematodes, crustaceans, and vertebrates. This developmental process eliminates a large portion of the inherited genome from all somatic nuclei of an organism. The phenomenon represents a dramatic departure from the rule of genome constancy that is the foundation of organismic development and differentiation, and raises questions regarding its regulatory mechanisms, biological roles, and evolution origin. Answers to these questions are beginning to emerge, and ciliates, being the simplest group, have been providing the bulk of information.
Programmed genome rearrangements were first reported in ciliates in 1971 in the spirotrichous ciliate *Stylonychia*, in which a drastic reduction in somatic DNA composition (as determined by CsCl density gradients) and molecular size were observed (3, 4). DNA renaturation kinetics studies of the distantly related *Tetrahymena* soon followed (5), revealing a 10 to 20% reduction in sequence complexity in the somatic genome. These findings expanded the phenomenon of programmed genome rearrangements in organisms beyond the Ascarid group, and established ciliates as model organisms for its molecular analysis. In the following decades, studies of the molecular details of this process ensued, some of which are summarized below.

**TETRAHYMENA LIFE CYCLE**

*Tetrahymena* lives freely in fresh water. It was among the earliest protozoa to be mass cultured and studied biochemically. The species *Tetrahymena thermophila* was chosen for genetic and molecular studies in the mid-20th century (6). Like most ciliates it displays nuclear dualism. A *Tetrahymena* cell contains two nuclei: a large polyploid macronucleus and a small diploid micronucleus, which separately carry out the somatic and germ-line functions (Fig. 1). *Tetrahymena* cells multiply through binary fission, typically in axenic peptone medium in the laboratory. During this vegetative growth phase, the micronucleus divides by mitosis, but the macronucleus divides by amitosis, an unusual division process without apparent chromosome condensation or spindle formation (7). RNA transcription occurs exclusively in the macronucleus, whereas the micronucleus remains condensed and silent throughout this phase. Upon starvation, *T. thermophila* cells of different mating types pair and initiate the sexual phase of the life cycle. In the laboratory, this process can be induced to occur in large quantities and with good synchrony. During this time, the micronucleus enters an elaborate meiotic division process that includes dramatic nuclear elongation and active transcription (Fig. 2). It produces four haploid nuclei; three of them are degraded and the remaining one divides once to generate two haploid gametic nuclei. The mating partners exchange one gametic nucleus, and the two gametic nuclei from the opposite partners fuse to form a diploid zygotic nucleus in each cell, thus completing the sexual process of genetic exchange. The zygotic nucleus then initiates a series of nuclear events to generate the new nuclei for the following vegetative growth phase. It divides twice mitotically to give rise to precursors for two new macronuclei, one new micronucleus, and a fourth nucleus that degenerates. During this time, the old macronucleus goes through a degenerating process that involves genes related to apoptosis (8) and autophagy (9). The mating pair separates and rests in this state with two new macronuclei and one new micronucleus per cell (“2 MAC 1 MIC”). The whole process takes no more than 20 h to complete. In the laboratory, the stages of mating are usually represented by hours after mixing starved cells of different mating types (hours post-mixing, hpm).

FIGURE 1. An image of *Tetrahymena thermophila* during vegetative growth. The fixed cell was labeled with antibodies against centrin (green) to show basal bodies on the cell cortex and a modified histone (red) to show the micronucleus. The macronucleus and micronucleus were stained with 4′,6-diamidino-2-phenylindole (DAPI; blue). The image was captured using Delta Vision. doi:10.1128/microbiolspec.MDNA3-0012-2014.f1
The cell resumes growth when nutrients become available. The first cell division occurs with the division of the micronucleus and the distribution of the two macronuclei to restore the state of two nuclei per cell.

**OVERVIEW OF TETRAHYMENA DNA REARRANGEMENTS**

New macronuclear development initiates at around 8 hpm, and the rearrangement of the genome occurs mostly at around 12 to 14 hpm (10). During the first three decades of studies, a number of genome-altering events were discovered, including chromosome breakage and telomere formation, selective amplification of the rRNA gene, and internal DNA deletion and elimination. These are described briefly below.

**Chromosome breakage and telomere formation**

Chromosome breakage in *Tetrahymena* was discovered as the initiation step of rRNA gene amplification (11). It was later found to occur at roughly 200 sites in the genome, breaking the five pairs of chromosomes down into smaller fragments with new telomeres (12), which were maintained in the mature macronucleus throughout the vegetative life. Recent macronuclear genome sequence analysis has confirmed this configuration and suggested the presence of 187 macronuclear DNA molecules (13). A well-conserved 15-nucleotide (nt) sequence, CBS (for chromosome breakage sequence), marks the breakage sites and is necessary and sufficient for breakage to occur (12, 14, 15). The same sequence is found in all *Tetrahymena* species analyzed (16), suggesting a
mechanism arising before the species had diverged. Breakage is highly efficient in T. thermophila, with no exceptions reported. After breakage, CBS and about 20 bp of DNA on both sides are lost, and the new ends are capped with new telomeres. Thus, the process reduces chromosome size drastically with the loss of relatively little DNA. The protein(s) that recognizes CBS and breaks chromosomes has not yet been found. It has been speculated that a homing endonuclease may have been domesticated to serve this function (17), but direct evidence is still lacking.

Ribosomal RNA gene amplification

Soon after the new macronucleus begins to enlarge, its DNA content increases through endoduplication. Most genomic sequences are replicated roughly equally, but the ribosomal RNA gene (rDNA) is selectively amplified (18). It was the first programmed gene-altering event known for Tetrahymena. T. thermophila contains a single copy of rDNA in the germline genome (19), which is unique among eukaryotes and rare among organisms analyzed. This copy is released from the chromosome through breakage at the CBS that is present in both ends; it becomes a head-to-head dimer (a palindrome) with new telomeres added to both ends and is selectively amplified. In the mature macronucleus, there are roughly 200-fold more copies of the rDNA than other sequences (18). The DNA ploidy level of the macronucleus has been determined by cytological methods. At the G1 stage, it contains about 23 times as much DNA as the diploid micronucleus and has been referred to as being 45C (20). Since the macronucleus has a reduced genome size (by 34%, see below), there are about 68 copies of most genomic sequences and about 13,000 copies of the rDNA in each macronucleus at the G1 stage. Palindrome formation appears to occur through an intramolecular recombination process guided by a pair of 42-bp inverted repeats that are present at one end of the rDNA (21). A similar process was later found to also occur in the yeast Saccharomyces cerevisiae (22) and in Chinese hamster ovary cells (23), and thus is probably widespread among eukaryotes. It could be a major contributing mechanism for gene amplification in cancer cells (24).

Internal DNA deletion and elimination

DNA renaturation kinetics studies revealed the loss of 10 to 20% of germline sequences from the macronucleus (5). The molecular basis for this was not known until about a decade later when cloning, sequencing, and hybridization methods became available. Focusing on separate genomic regions, two studies uncovered the deletion of specific DNA segments that occurred reproducibly as a programmed process (25, 26). One of these regions contains the M- (for middle) and R- (for right) element, which are relatively small in size (0.6 and 1.1 kb, respectively), and has been used extensively for analysis since then as a model of DNA deletion. By analyzing 20 randomly chosen DNA segments, a rough estimate was made that put the total number of deletion segments at 5,850 or more for the genome, with an average size of 2 to 3 kb to account for the total amount of DNA eliminated (26). Impressively, this is not too far from the actual number and sizes, which it is now possible to measure using micronuclear genomic sequence data (see below). In the following two decades, a number of deletion elements (or IESs for internal eliminated sequence) were cloned and characterized, which revealed several common features (27, 28, 29, 30, 31, 32, 33, 34). First, they were all simple deletions not accompanied by any additional sequence change at the junction. Secondly, the IES were mostly noncoding sequences but also included transposons and their degenerated remnants. No cellular genes were found in these studies and no shared sequence features could be identified. Thirdly, all IESs resided in noncoding regions including spacers and introns, and none was found to interrupt coding sequences. Fourthly, the majority of these elements had invariable boundaries, although some clearly had variable boundaries. These studies revealed a remarkable process that cuts, deletes, and rejoins DNA precisely at thousands of specific genomic locations, all within a period of approximately 2 h (12 to 14 hpm) during nuclear differentiation. At the same time, they have offered little information as to how the specificity of deletion is determined and how the process may affect cellular functions. Solving this mystery remains an interesting challenge. Recent genome sequence studies (35) (http://www.broadinstitute.org/annotation/genome/Tetrahymena) have offered a near-complete picture of all deletion elements, which basically confirms these general features with possible exceptions. Based on the draft genome sequences released, about 34% of the micronuclear genome is eliminated from the macronucleus (reducing the genome from 157 to 104 Mb), including more than 10,000 IESs. Some IESs may occur within coding regions and some genes may reside within IESs, although these exceptions are yet to be confirmed. The rich sequence information offers opportunities for further understanding, which awaits detailed analysis.

Since the publication of Mobile DNA II in 2002, major advances have been made in our understanding of
DNA deletion in *Tetrahymena* and *Paramecium*. The process is shaping up as a specialized form of RNAi that has evolved through the domestication of transposases to delete all transposons from the somatic genome. In the following sections, we will review these lines of progress in *Tetrahymena* and offer our views on its mechanisms, functions, and evolution.

**LINKS TO NONCODING RNA**

**Transcription of internal eliminated sequences**

Perhaps the most significant discovery made on ciliate DNA deletion in the past decade is the link to RNAi (Fig. 3). *Tetrahymena* IESs contain mainly noncoding sequences, yet they are transcribed in an unusual manner. The first observation was reported in 2001, soon after the landmark publication of RNAi in nematodes (36). In studying the well-characterized M-element, it was found through Northern blot hybridization that the element was transcribed from both strands to produce heterogeneous-sized RNAs averaging several hundred nucleotides in length that have no defined ends or poly(A) tails (37). They occur only during conjugation, starting shortly after pairing begins (2 hpm), reaching peak levels just after meiosis (4 to 5 hpm), and continuing through the middle stages of macronuclear development (12 hpm). Similar transcription was also found for the R-element, the repetitive element 2512, and several additional elements in later studies (38). The

![FIGURE 3](ASMscience.org/MicrobiolSpectrum)
exact timing of transcription differs among elements, with some occurring only after the new macronucleus begins to develop. It was thought that double-stranded RNA (dsRNA) transcription occurred in most IESs but not in sequences to be retained in the macronucleus (known as MDSs for macronucleus-destined sequences), but definitive evidence has been hard to obtain partly due to the low and varying abundances of these transcripts (37). High-throughput sequencing analysis may provide additional insights but has not yet been reported. Analysis of small RNA (sRNA), which presumably is derived from this dsRNA, however, has supported this notion (see later sections). The importance of this transcription to DNA deletion was supported by actinomycin D treatment studies. Incubation with actinomycin D for 3 h during these stages (starting every half hour between 4.5 and 7 hpm) produced progeny with partial defects in M-element deletion (37). Although subject to other interpretations, the study offered a strong support for the role of dsRNA in DNA deletion. This unusual form of transcription is probably carried out by polymerase II, since a subunit, RPB3(CNJC), is highly expressed during meiosis and the protein is localized to the meiotic nucleus (39). Direct evidence is yet to be obtained to prove this point.

sRNA and dsRNA

The discovery of RNAi in eukaryotes (36) opened up new possibilities, and a study of Piwi proteins in Tetrahymena provided the first clear link to this process. Mochizuki et al. (40) discovered the presence of a new class of sRNA expressed specifically during Tetrahymena conjugation. This sRNA is around 27 nt (26 to 31 nt) in size and is highly abundant (detectable even by ethidium bromide staining after electrophoresis). It appears early during conjugation and is present until late stages (2 to 16 hpm). A piwi-related gene (TWI1), expressed early during conjugation, is necessary for the maintenance of this sRNA. Twi1p first appears in the cytoplasm, then mainly in the old macronucleus, and finally only in the new macronucleus (Figs 2 and 4). In somatic knockout strains, conjugation is arrested at the stage with two new macronuclei (MAC) and two micronuclei (MIC) (“two-MAC two-MIC”), typical of most mutants blocked in DNA rearrangements. Deletions of the M- and R-elements are completely blocked and chromosome breakage is partially affected at the site (Trt819) examined. This study provided a strong link for two main players of RNAi, namely a Piwi protein and sRNA, with DNA deletion, and led logically to the conclusion that DNA deletion involved the RNAi machinery. Based on these and earlier data, the authors went on to propose a provocative model for DNA deletion, in which sRNA (which they named scanning RNA or scnRNA) was generated from the entire genome and “scanned” the old macronuclear genome to move all similar sequences, and the remaining “germline-specific” sRNA was then used to direct DNA deletion. This model is appealing in proposing a role for sRNAs in controlling DNA deletion and explaining how the sequence specificity of DNA deletion is achieved. It has dominated the field for the past decade. We will return to it again later to examine its current supports.

FIGURE 4 Localization of key proteins during conjugation. The panels summarize the localization patterns of various proteins in different cellular compartments during conjugation. Genes for the proteins involved are listed to the left, with each row indicating the progression of mating (from left to right). Proteins with similar patterns are shown in the same row. The compartments in which proteins have been detected (using antibodies or peptide tags) are colored. doi:10.1128/microbiolspec.MDNA3-0012-2014.f4
The involvement of dsRNA in DNA deletion was demonstrated directly by injecting dsRNA into conjugating cells (41). The RNAs were 500 bp or longer and were produced in vitro from cloned macronuclear DNA templates of either coding or noncoding regions. They were injected into the cytoplasm around the time of macronuclear development, and the progeny cells produced, at rates nearing 50%, showed deletion of the targeted region in the macronuclear genome. Unlike endogenous deletions, these induced deletions had variable boundaries, some reaching beyond the targeted region by several kilobases. This study has an interesting resemblance to the original RNAi experiments in Caenorhabditis elegans (36), except that the consequence is DNA deletion, perhaps the most extreme form of gene silencing. It provided a direct support for the involvement of RNAi in DNA deletion, and showed that dsRNA alone was sufficient to trigger this process.

**Dicer proteins**

These two studies firmly establish the link between DNA deletion and RNAi, and prompted studies of other RNAi players in Tetrahymena including Dicer, Piwi, and related proteins, which are summarized briefly here. The Tetrahymena genome contains three orthologs of Dicer genes, DCR1, DCR2 and DCL1. Only DCL1 is expressed exclusively during conjugation. Genetic knock-out strains lacking DCL1 were able to initiate mating but did not produce scnRNA and appeared to accumulate the heterogeneous-sized dsRNA. These cells were arrested at the two-MAC two-MIC stage and failed to carry out DNA deletion and chromosome breakage at the sites examined (42, 43). Thus, DCL1 is the Dicer gene responsible for generating scnRNA. Tagged Dcl1p appeared mainly in meiotic nuclei, which apparently were the compartments in which dsRNA was processed into scnRNA. However, it was not detected in macronuclear anlagen (42), raising the question of where dsRNA was processed at these stages (8 to 16 hpm) (Figs 2 and 4). Studies of DCR1 and DCR2 revealed no obvious link to scnRNA production. DCR1 is not essential and no clear function has been assigned to it (42, 44). DCR2 is essential for growth. The encoded protein cleaves dsRNA to produce 23 to 24 nt sRNA in vitro (45) and is likely responsible for making this class of endogeneous sRNA in Tetrahymena (46). The Tetrahymena genome contains a single gene, RDR1, for the RNA-dependent RNA polymerase, which is essential for growth (43). The encoded protein physically interacts with Dcr2p and has been implicated in 23 to 24 nt sRNA production. Its role in scnRNA production, if any, is not known.

**Piwi-interacting proteins**

Four Twi1p-interacting proteins have been identified through co-immunoprecipitation experiments (38, 47, 48). Giw1p has no known conserved function. It is required for the localization of Twi1p to the nucleus and the ability to remove one strand of the sRNA. Ema1p is likely a DExH box RNA helicase. It is thought to be involved in the binding of Twi1p to chromatin. Immunoprecipitation experiments suggest that it is involved in the association between scnRNA and long noncoding RNA. Wag1p and CnjBp are two GW repeat proteins. All four proteins are expressed preferentially (CNJB) or specifically (the other three) during conjugation and localize to the parental macronuclei early and the developing macronuclei late during conjugation (Figs 2 and 4). In addition, Ema1p showed strong cytoplasmic localization throughout the stages and CnjBp was found mainly in the meiotic nucleus. These genes are required for the completion of DNA deletion either singly (GIW1 and EMA1) or in combination (WAG1 and CNJB). Interestingly, EMA1 mutants were blocked in the deletion of some (M and Tlr1) but not other (R and Cal) elements tested, revealing IES subgroups that differed in deletion control. Although the underlying mechanism remains unclear, these studies began to offer the basic information necessary toward a comprehensive understanding of scnRNA actions in Tetrahymena.

**CHROMATIN MARKINGS**

**Chromodomain proteins**

The link to RNAi helps explain the dynamics of chromatin during DNA deletion. Analogous to small interfering RNA (siRNA) and Piwi-interacting RNA (piRNA) in other organisms, the scnRNA in Tetrahymena is thought to target chromatin through sequence homology and brings about changes that turn it into “heterochromatin.” The association between heterochromatin and DNA deletion was known before the RNA link was discovered. Earlier work identified several nuclear proteins that were synthesized preferentially during macronuclear development (49). One of the most prominent, p65, was partially sequenced and the gene identified and named PDD1. It is related to the conserved heterochromatin protein HP1, with two chromodomains and a chromo shadow domain (50). The protein is expressed only during conjugation and is localized mainly to the parental macronucleus at early stages and to the developing macronucleus at late stages. In the developing macronucleus, Pdd1p first appears homogenously, becomes...
punctate (10 hpm), and finally forms a few large aggregates with doughnut-like appearances under the electron microscope before it disappears completely (14 hpm) (Figs. 2 and 4). Through in situ hybridization, micronucleus-specific sequences have been shown to have similar distributions (51). A colocalization study indeed found Pdd1p present in aggregates of IESs in the developing macronucleus (50). Thus, IES chromatin appears to be highly dynamic and forms heterochromatin-like aggregates before being eliminated, with Pdd1p as an integral part of this process. Gene knockout studies indeed showed that PDD1 was essential for DNA deletion. Conjugation was arrested at the two-MAC two-MIC stage in strains lacking the macronuclear copies of this gene. Remarkably, tethering Pdd1p to a sequence not normally deleted in a special plasmid system (52) could cause it to be efficiently deleted, arguing for a key role for this protein. These studies set the foundation for all following studies of genes involved in DNA rearrangements. Two other related genes, PDD2 and PDD3 (also a chromodomain protein) were found to have similar properties (53, 54, 55) (Fig. 5). A recent study (56) analyzed PDD1 through substitutions and deletions of various parts and determined that the two chromodomains and the chromo shadow domain are all essential (or highly important) with distinct functions. The first chromodomain can be substituted with a DNA-binding motif, implying a main role in DNA targeting. The other two domains are likely engaged in histone modifications and Pdd1p aggregation separately.

**Histone modifications**

Other heterochromatin marks have also been investigated. The first study addressed the importance of histone deacetylation using the inhibitor trichostatin A, which partially blocked DNA deletion when applied during conjugation (57). Di- and trimethylation of histone 3 lysine 9 (H3K9me2,3) are perhaps the most consistent marks of heterochromatin in eukaryotes. These modifications appear in developing macronuclei and disappear after DNA deletion has occurred (52) (Figs. 2 and 4). Chromatin immunoprecipitation (ChIP) experiments have shown that they are associated specifically with selected IESs at this stage (52), although their overall genomic distributions have yet to be determined. Mutating both copies of the H3 genes in the genome

---

**FIGURE 5** Formation of the Pdd1p aggregates during DNA deletion. The chromodomain protein Pdd1p and other proteins that participate in the formation of the large heterochromatin aggregates are shown. The proteins are guided to IESs through sRNAs and eventually form large aggregates with excised IESs. The left panels show micrographs of mating cells in three successive stages stained with antibodies against Pdd1p (green) and DAPI (blue). doi:10.1128/microbiolspec.MDNA3-0012-2014.f5
to prevent K9 methylation (K9Q) led to failure in DNA deletion and conjugation arrest (at the two-MAC two-MIC stage), establishing the importance of these modifications. In these mutant lines, scnRNA continued to accumulate, supporting the argument that H3K9 methylation acted downstream of scnRNA accumulation (58). In this regard, it is interesting to note that in DCL1 mutants, H3K9 methylation still occurs in developing macronuclei but is no longer enriched in IESs.

Another heterochromatin mark, trimethylation of histone 3 lysine 27 (H3K27me3) also plays important roles. Unlike H3K9me2,3, H3K27me3 occurs in both growing and mating cells and likely in all types of nuclei. In developing macronuclei, it has a distribution pattern very similar to those of H3K9me2,3 and Pdd1p, suggesting a possible link to DNA deletion. Three methylase genes related to this modification are present in the genome, and one of them, EZL1, is expressed only during conjugation. Mutants of EZL1 lack H3K27me3 in parental and developing macronuclei and cannot carry out DNA deletion, indicating an essential role in this process. H3K27me3 occurs before, and likely facilitates, H3K9 methylation, since H3K9me2,3 does not occur in EZL1 mutants but H3K27me3 occurs normally in H3K9Q mutants (59). It should be noted that, unlike H3K9 methylation, which occurs only in developing macronuclei, H3K27me3 also occurs in other nuclei at other stages and could have additional roles. A demethylase gene, JMJ1, has been studied in Tetrahymena. It is partially responsible for the removal of H3K27me3 in the developing macronucleus. It has only a minor role in DNA deletion since knockdown mutants showed only slight defects in DNA deletion but were unable to complete macronucleus development beyond DNA deletion (60).

Other chromatin proteins

Four other proteins have been identified for their associations with heterochromatin-like structures during DNA deletion (61). They were found through a screen of a cDNA library for proteins localized specifically in developing macronuclei during DNA deletion. Five novel proteins were identified, and four of them were found to localize to the Pdd1p structure. Three of these proteins, Lia1p, Lia3p and Lia5p, have been further analyzed (62, 63) and all appear to be important for DNA deletion. LIA1 and LIA5 are essential for DNA deletion and the completion of conjugation, and LIA3 appears to affect boundary determination of some IESs (Figs 2 and 4). Further analysis of these novel proteins will likely expand our understanding of heterochromatin in eukaryotes.

These studies indicate that scnRNA leads to heterochromatin formation in Tetrahymena, involving conserved proteins and histone modifications that are shared by other siRNA-directed transcriptional gene silencing processes (Fig. 3). H3K27 methylation likely occurs before and promotes K9 methylation. Pdd1p (and probably Pdd2p, Pdd3p, and Lia1p) binds to these moieties and eventually causes deletion to occur. It probably also has a role in re-enforcing these modifications. In addition, Pdd1p has an unknown role in scnRNA production, since scnRNA is not produced in its absence (40).

ROLE OF PARENTAL MACRONUCLEI AND THE RNA SCANNING MODEL

Nuclear dualism provides special opportunities for interactions between nuclei. This is especially true during conjugation, when three different types of nuclei, each with its own fate, are present in the same cytoplasm. The interaction between the parental and progeny somatic genomes offers an intriguing aspect of ciliate DNA rearrangements. Early genetic work in Paramecium demonstrated the effect of the parental macronucleus on progeny macronucleus differentiation (64, 65), which was validated and further elucidated with the discovery of DNA rearrangements in this organism (66, 67). Similar effects are known to occur in Tetrahymena and Oxytricha as well, implying a common inheritance strategy accompanying nuclear dualism. In Tetrahymena, the first evidence was derived from experiments that put high-copy-number plasmids containing M- or R-elements in the parental macronucleus, which led to partial inhibition of the respective IES during subsequent conjugation (68, 69). Due to its sequence specificity, this effect argued for an RNA-mediated cross-talk between these two nuclei (1). The discovery of a link to RNAi immediately suggested a role for scnRNA in this cross-talk. The “scanning RNA” model was proposed (40), which argued that the entire genome was transcribed to produce dsRNA during meiosis, which was processed into sRNA and transported into the parental macronuclei to “scan” the genome. All identical sequences were removed or “subtracted” and the remaining sRNA, which now contained sequences present only in the germline genome (and not the somatic genome), was transported into the developing macronucleus to direct chromatin changes and DNA deletion. It provided a simple and elegant solution for the establishment of sequence specificity in DNA deletion and offered a nice explanation for how IESs inserted in parental macronuclei could inhibit...
the deletion of the same sequences in the developing macronucleus.

The RNA scanning model can be tested through some of its molecular assumptions and predictions. The followings are three of the basic requirements of the mechanism. First, the entire genome should be transcribed into dsRNA and processed into siRNAs. Secondly, these siRNAs must enter the parental macronucleus and then move into the developing macronucleus. Thirdly, there is a molecular mechanism to “subtract” siRNAs of similar sequences, thus reducing the sequence complexity after leaving the parental macronucleus. During the past decade, attempts to gather support have met with mixed results. The movements of siRNAs through various nuclear compartments have been difficult to measure directly. However, the appearance of the Piwi protein Twi1P, which is known to bind to this species of sRNA, has been followed, and largely agreed with the expected patterns. Whether the change in protein localization is actually due to movement or to degradation and new synthesis will require additional studies (40, 70).

Another major issue that has been studied is scnRNA abundance throughout conjugation. Several earlier studies offered partial support, although they were not conclusive (40, 43, 70). With the release of the micronuclear genome sequence, high-throughput sequence analysis of sRNA become possible and has now been reported (71). The result of this comprehensive analysis is not in total agreement with the scanning model. First, sRNA was not derived from the entire genome, and secondly, although sRNA abundances did change through conjugation, they were not at the level expected from the model. The study determined sRNA sequences from stages of conjugation before, during, and after the proposed scanning process (3, 4.5, 6 and 8 hpm). It appeared that most (~80%) of the sRNA (26 to 32 nt) at the earliest stage measured (presumably before any “scanning” could have occurred) was derived from the ~25% of the genome that can be uniquely assigned to IESs. The sequences to be retained (MDSs) were also transcribed but to a much lower extent (producing 15% of the sRNA). This result suggested a biased transcription, and not RNA scanning, for the enrichment of siRNAs from IESs. Using a special genetic strain (nullisomic for chromosome 4 in the micronucleus), it was clear that this sRNA was derived from the micronuclear (and not the macronuclear) compartment. This bias was seen also in mutants of TWI1 and EMA1, which should not have carried out “scanning” as proposed. Clearly, sRNA production was not equal over the entire genome, which presumably was the result of selective transcription. Nonetheless, a significant proportion of sRNA was indeed derived from MDSs. Their roles are not clear and will need to be explained. The relative amounts of sRNA at a later stage (8 hpm) appeared to change, with slight increases for those from IESs and significant decreases (about 2-fold) for those from MDSs. This decrease was confirmed in selected loci by hybridization. EMA1 also appeared to be important for some of these decreases. This change indicated a dynamic regulation of sRNA abundance. Whether it was due to “subtraction” by the macronucleus or other unknown processes has yet to be determined.

The model made one clear prediction that is testable, but the results have not been supportive so far. If a sequence is artificially removed from the parental macronucleus, the corresponding micronuclear sequence should now appear as a “micronuclear-specific” sequence. Since no “subtraction” of the corresponding scnRNA sequence can be performed, the sequence should be subjected to deletion. Thus, all somatic knockout strains (and other strains with somatic deletions) should generate offspring with the same somatic sequences deleted. In other words, somatic deletion should be heritable in Tetrahymena. This is a clear and unambiguous prediction, and is indeed reported for at least one gene in Paramecium (72). However, it has never been seen in Tetrahymena.

If the scanning model (in its original form) cannot be validated, one will need modifications or alternative explanations for the influence of the parental macronucleus on DNA deletion as well as the mechanism for selecting a sequence for deletion. One suggestion is that selective transcription of dsRNA could be the primary force for determining a sequence for deletion (73, 74). How the transcription is regulated is not known and will be the key to understanding this process. The parental macronucleus could play a secondary role to augment this sequence selection. If the same sequence is also present in the macronucleus, it will be transcribed, but the transcripts may not be processed into sRNA for the lack of DCL1 in this compartment. This RNA may interact with the sRNA made from the micronuclear genome (similar to the scanning model idea) and prevent it from carrying out deletion (73).

**DNA CUTTING AND TRANSPON DOMESTICATION**

The link to RNAi provides an attractive explanation for the evolution of DNA deletion. Like other eukaryotes, ciliates may have evolved RNAi machineries to defend against invading genetic elements. Tetrahymena has a
23 nt siRNA-mediated, post-transcriptional gene silencing process against endogenous and exogenous RNAs (44, 46). The 27 nt sRNA pathway of DNA deletion, on the other hand, shares features with RNAi processes of transcriptional gene silencing and may have evolved from it. One key step in this evolution would be the acquisition of DNA cutting activities to remove the silenced chromatin. The recent discovery of a domesticated transposase may have provided this crucial link. *Tetrahymena* contains a number of transposons in the genome (31, 35, 75, 76), which are all removed from the macronucleus, mostly through DNA deletion. Interestingly, two genes (TPB1 and TPB2) that show high similarity to piggyBac transposases and contain the catalytic DDD motif are retained in the macronucleus. These genes lack other transposon features and are expressed specifically during mid- to late stages of conjugation (6 to 14 hpm) (Fig. 2). Both encode proteins roughly twice as long as most piggyBac transposases, probably through fusions with additional exons at the N terminus (for TPB1) or C terminus (for TPB2) (77). Further studies showed that TPB2 indeed participated in DNA deletion. Green fluorescent protein-tagged proteins colocalized with the chromodomain protein Pdd1p to form distinct aggregates (Fig. 4). Genetic knockdown strains were unable to carry out DNA deletion or chromosome breakage and arrested at the two-MAC stage, typical of mutants defective in DNA rearrangements. Proteins expressed in *Escherichia coli* showed weak but specific endonuclease activities to generate ends with 4-nm 5' overhangs, a transient feature of M-element deletion (78). These results suggested strongly that Tpb2p participated directly in DNA cutting, and was likely the key enzyme that turned the RNAi process into the DNA deletion process (Fig. 3). Tpb2p probably lost its original specificity for the transposon terminal inverted repeats and became targeted to heterochromatin through the acquired C-terminal sequence. Site-directed mutagenesis studies showed that the zinc-finger domain in the C-terminal half was indeed important for Tpb2p localization (79). A TPB2 ortholog, PGM1, has been found in *Paramecium* and likely serves a similar role (80). Thus, the domestication process probably occurred before these two species split. PGM1 also contains an extra C-terminal half with a zinc-finger domain, although it shares little sequence identity with that of TPB2, suggesting rapid evolution of this sequence. Another gene, LIA5, has recently been described in *Tetrahymena* (61, 63). It has some sequence similarity with the IS4 family of transposons that includes TPB2 but lacks the conserved DDD motif. It also contains a zinc-ribbon domain in the C-terminal half. It is essential for the formation of the Pdd1p aggregates and DNA deletion but is not colocalized with Pdd1p. LIA5 is probably related to TPB2 in evolution and function, which further illustrates the importance of transposase domestication in this process, in a way similar to the domestication of *RAG1* during the evolution of the adaptive immune system of vertebrates. TPB1, on the other hand, is not essential for most DNA deletion and the completion of conjugation. It is more divergent from TPB2 than PGM1, and no ortholog has been found in *Paramecium*. It plays some non-essential role during conjugation, but its nature is not yet clear (72).

The discovery of PGM1 in *Paramecium* and TPB2 in *Tetrahymena* offered a nice explanation for the evolution of DNA deletion. This transposase gene could have been domesticated in the ancestor species, acquired the ability to target all heterochromatin, and modified its activity from transposition to deletion. It thus turned transcriptional gene silencing of RNAi into DNA deletion and heterochromatic DNA into IESs. PiggyBac transposase may not be the only transposase captured for this activity in ciliates. The transposase of a presumably active Mariner family transposon, TBE1, is required for DNA deletion in the distantly related ciliate *Oxytricha* (81). Thus, DNA deletion may have evolved independently in different branches of ciliates, or may have continued to evolve by adopting new transposases for this role after its initial establishment.

**NONHOMOLOGOUS END JOINING AND DNA DELETION**

After DNA cutting to remove IESs, the broken ends generated are ligated back together, apparently involving the cellular DNA double-strand break repair machinery (Fig. 3). Ku70 and Ku80 proteins are conserved in eukaryotes for repairing double-strand breaks through the nonhomologous end-joining (NHEJ) pathway, presumably by binding and protecting these ends (82, 83). *Tetrahymena* contains two Ku70 and one Ku80 (TKU80) orthologous genes. An analysis of TKU80 indicated that it was not essential for growth but was essential for completing conjugation. Mutants of TKU80 could not complete DNA deletion and were arrested at the two-MAC two-MIC stage of conjugation. In these mutants, however, cleaved IESs were detected through their circularized junctions, indicating the occurrence of DNA cutting. Without TKU80, the developing macronucleus accumulated broken DNA ends (as revealed by terminal...
deoxynucleotidyl transferase dUTP nick end-labeling assays) and gradually lost most of its DNA, presumably through degradation from these ends (84). TKu80p is not localized to but is required for the formation of Pdd1p aggregates, indicating that DNA cutting is not dependent on the formation of these aggregates (Figs 4 and 5). In mammalian immune systems, Ku is also required for completing V(D)J recombination, but this process differs from IES deletion in that ligation of the deleted segments (the signal ends) also requires Ku (85, 86).

Although aggregates of Pdd1p are the most prominent structural features associated with DNA deletion, their role is not entirely clear. They contain common heterochromatin features including histone H3 modifications and chromodomain proteins, as well as the domesticated transposase Tpb2p (Fig. 5). Their DNA contents include repetitive sequences of IESs and possibly all IESs. Since IESs are widely distributed in the genome, one wonders if they have already been cleaved before appearing in these aggregates. The TKU80 studies mentioned above suggest that these aggregates are not required for cleavage, raising the possibility that they instead are involved in repairing the cleaved ends, a step that would normally require Ku proteins. In this regard, it is interesting to note that some “aggregates” can be induced to form by UV treatment in LIA5 mutants, which normally do not form any such aggregates. This result suggested a strong link to a DNA damage response (63).

SETTING DELETION BOUNDARIES

In Tetrabymena, essentially all IESs are located within noncoding regions including introns (35). Precision of deletion thus is probably not as critical for somatic gene expression as it is in Paramecium, Euplotes, and Oxytrichia, in which many IESs interrupt coding regions. Nonetheless, DNA deletion is still a largely precise process in Tetrabymena. Initial analysis of the M- and R-elements revealed that, among dozens of independent progeny lines analyzed, the rejoined junctions had exactly the same sequences, indicating the precision of deletion at the single-nucleotide level. The few exceptions had junctions that differed by only a few nucleotides (27, 87). This precision is probably a common feature, although some IESs clearly have nonprecise junctions that vary by several hundred nucleotides or more (29, 34). It is likely that the majority of IESs in the genome are deleted precisely, like the M- and R-elements, but a small proportion has highly variable junctions. This raises the question of how the deletion junction is determined. There are at least two issues. First, if the cutting enzyme (presumably Tpb2p) is targeted to heterochromatin as is currently thought, how does it locate the two ends of an IES for cutting? And secondly, how is the single-nucleotide precision achieved through a guiding system that targets nucleosomes?

DNA deletion can be induced to occur anywhere in the genome through dsRNA injection. In this case, no pre-existing cis-acting sequence is required. DNA deletion can also occur to the aph gene of E. coli inserted at any genomic location. Unlike the M- and R-elements, these induced deletions have boundaries that vary by hundreds to thousands of nucleotides (41, 88). Presumably, these dsRNAs (or derived sRNAs) guide chromatin marking over the entire region to be deleted, which in turn attracts Tpb2p to carry out deletion. We suggest that Tpb2p is targeted to the two ends of the element by recognizing the borders and not the entire “heterochromatin” (Fig. 6). After cutting, the IES would be released as one piece, allowing the production of the circularized junctions observed. Junction variations can be explained if the heterochromatin region is imprecise, being somewhat variable in different chromatin.

Earlier studies revealed flanking sequences that set the boundaries of the M- and R-elements. By introducing specifically modified elements into conjugating cells in a rDNA plasmid, cis-acting sequences important for deletion have been determined. A 10-bp motif (5′-AAAAAGGGGG-3′) located about 45 bp away from both ends of the M-element was found to be crucial for setting deletion boundaries (89, 90). Without it, the boundary became variable. Inserting this sequence within the M-element induced new boundaries to form at about 45 bp downstream. Thus, this sequence is necessary and sufficient to set a deletion boundary at a specific distance from it. The R-element also appears to have a pair of flanking sequences that set its boundaries (91). They are AT-rich sequences, although their exact identities have not been determined. It has been suggested that there are many families of IESs, each defined by a particular flanking sequence that they share (91).

We suggest that flanking regulatory sequences set boundaries for heterochromatin, similar to chromatin boundary elements in other eukaryotes (92, 93). Establishing heterochromatin boundaries helps to provide defined target sites for Tpb2p (and other nucleases) binding and cutting. This interaction could set the Tpb2p binding target to a small region, perhaps no more than a few dozen base pairs, within which Tpb2p will cut at a specific nucleotide due to its binding preference. Several proteins including CTCF have been known to help set chromatin boundaries in various eukaryotes (94, 95).
We suggest that these or related proteins are involved here. These, or proteins that interact with them, could recognize and bind the flanking regulatory sequence to set the deletion boundary. A recent study has revealed that Lia3p possessed the expected properties for setting the boundaries of the M-element (C. M. Carle and D. L. Chalker, personal communication) (Fig. 6). Even though \textit{LIA3} mutants were able to carry out DNA deletion, they lost the boundary specificities for M- and a few other elements that contain the flanking 5'-AAAAAGGGGG-3' motif. \textit{In vitro}, the protein is able to bind to this sequence, likely through the formation of a G quadruplex.

If IESs were derived from invading genetic elements such as transposons, they would be without boundary sequences when first inserted into the genome and would be deleted without precise boundaries. Those inserted within or very close to coding regions would have detrimental effects and could not be kept. Those inserted some distance away would still be harmful until \textit{cis}-acting sequences had evolved to set boundaries and prevent chance deletions of neighboring genes. Thus, we suggest that flanking regulatory sequences evolved later, and that IESs with variable junctions are probably younger IESs (or located in regions poor in coding sequences, such as heterochromatic regions).

**MATING-TYPE DETERMINATION INVOLVES A NEW TYPE OF DNA REARRANGEMENT**

Although chromatin diminution affects a large part of the genome, specific phenotypic roles had not been found in association with any specific rearrangement until recently when mating-type determination mechanisms were characterized. In \textit{Tetrahymena} and \textit{Paramecium}, studies over the past four decades have hinted at the involvement of DNA rearrangements in mating-type determination. \textit{T. thermophila} has seven mating types determined by a single locus with several known alleles. A progeny cell can be any one of these seven types, regardless of its parental mating types, and different subclones derived from a single mating pair, while having identical germline genomes, can express different mating types. Once determined, the mating type is very stable and has never been known to change. These features suggest the involvement of somatic DNA rearrangements (96). Cloning and analysis of this gene should help reveal the nature of this event, and this was finally achieved with the aid of genome sequence data (97). In the macronucleus, the MAT locus is composed of a pair of divergently transcribed genes (\textit{MTA} and \textit{MTB}) with different sequences in cells of different mating types, presumably as the result of DNA rearrangements. Sequencing of the micronuclear genome revealed the nature of these rearrangements. The germline locus is around 91 kb. At its two ends are the C-terminal halves of \textit{MTA} and \textit{MTB}, which are shared (and thus invariable) among all mating types, and the internal region contains an array of the different N-terminal halves of all potential mating-type genes. For the allele that was sequenced (the B allele), there are six mating-type genes (all except type I) arranged in the order 2-5-6-4-7-3, in addition to a number of IESs and spacers (Fig. 7). To generate a

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6}
\caption{Setting deletion boundaries. The cartoon shows a possible heterochromatin structure with associated proteins over the IES to the right and the neighboring euchromatin to the left. The domesticated transposase Tpb2p is targeted to the junction for DNA cutting. In the example represented here, the M-element is shown to have a flanking regulatory sequence (5'-AAAAAGGGGG-3') that is recognized by Lia3p to set the heterochromatin boundary and specify the cutting site. doi:10.1128/microbiolspec.MDNA3-0012-2014.f6}
\end{figure}
particular mating-type gene pair during conjugation, all the IESs and all except one of the six partial gene pairs in the internal region are deleted. The remaining partial pair is joined to the “constant” C-terminal regions at both ends to generate the complete gene pair. This is a remarkable rearrangement that is apparently subject to strict regulations. It is interesting to note that the rearrangement occurs by recombination between the constant regions and the internal regions in sections that share sequence homology for hundreds of nucleotides. This site-specific homologous recombination is distinctive from IES deletion, which also occurs. How it is regulated to produce only one complete gene pair per chromosome remains an interesting problem to solve. The study nicely explains the determination of mating type in *Tetrahymena* and reveals a new DNA rearrangement mechanism distinctively different from IES deletion and chromosome breakage. It involves homologous recombination at specific ectopic sites, which potentially could also occur in other genomic regions.

**CONCLUDING REMARKS**

The past decade has witnessed transforming progress in our understanding of ciliate genome rearrangements. The link to RNAi has provided a plausible evolutionary explanation, suggested likely functional roles, and revealed crucial details in mechanism for DNA deletion. DNA deletion can now be viewed as a genome defense system against invading genetic elements including transposons. It has evolved from the RNAi process through the domestication of transposases and has turned the transcription silencing mechanism into the DNA deletion process. This intriguing phenomenon illustrates the powerful forces of transposons in shaping genomes and driving evolution.

Many interesting questions still remain. For instance, how are sequences recognized for deletion? Since dsRNA is the initiating event, control of its transcription becomes the key issue. Are there cis-acting sequences to be recognized, and how could this mechanism detect a foreign sequence? Relative little is known about the control of noncoding RNA transcription in general. Studies in *Tetrahymena* may help to provide some insights. Also, how does the old macronucleus interfere with DNA deletion in a sequence-specific manner? Little molecular detail is known, but RNA most likely plays a key role. Does the sequence recognition involve RNA–RNA or RNA–DNA interactions? This mysterious process may reveal new biological principles. Finally, how do sRNAs find their chromatin targets? In transcriptional silencing systems, it is generally thought that pairing with nascent transcripts guides sRNAs to their DNA targets (98). Little is known in ciliates in this regard. Since injecting dsRNA is sufficient to induce deletion at any locus, transcription of that locus may not be required, raising questions regarding a targeting mechanism that depends on nascent transcripts. Could sRNA directly target DNA?
This appears to be the case in the CRISPR system in bacteria (99, 100) and remains a distinct possibility here. Questions also remain regarding processes other than DNA deletion. How has chromosome breakage evolved in *Tetrahymena* and what endonucleases are used in DNA cutting? What regulates the site-specific homologous recombination process of the MAT locus? These and other questions continue to make *Tetrahymena* a fertile ground in which to uncover mechanisms that remodel the genome.

Massive somatic genome rearrangements such as chromatin diminution occur sporadically in diverse groups of organisms, including limited species of nematodes, crustaceans, and vertebrates, and most species of ciliates. This diversity suggests independent evolutionary origins, which have likely resulted in mechanisms that differ in details. Recent genome sequencing studies in *Ascaris* did not reveal the involvement of internal DNA deletion or sRNA in chromatin diminution (101). In lamprey, sRNA also has not been reported in association with chromatin diminution (102). Nonetheless, the prevalence of transposons and the wide occurrences of RNAi offer excellent opportunities for the independent evolution of similar processes in diverse groups. Further investigation should clarify this possibility.

Among these organisms, ciliates are particularly interesting. While global genome rearrangements occur only in isolated groups in other phyla, they occur in all ciliate species studied. Moreover, multiple forms of rearrangements appear to occur within a single ciliate species. In *Tetrahymena*, in addition to the well-studied DNA deletion process, other programmed processes including chromosome breakage, ribosomal gene differentiation and amplification, and homologous recombination have also been found. This complexity might also be present in other ciliates as well. Ciliates appear to be hotbeds of programmed somatic DNA rearrangements, and we think nuclear dualism may have played a major role. Having evolved two nuclei in the same cell that are separately responsible for gene expression (the macronucleus) and high-fidelity genome transmission (the micronucleus), the organism can tailor these two compartments differently for their respective roles. Programmed DNA rearrangements provide the scissors and glue to restructure the somatic genome for optimal gene expression. This ability may allow the germline genome to accumulate changes that facilitate genome transmission without hampering gene expression. Presumably, similar changes can also evolve in non-dividing tissues of metazoans. The fly salivary glands and mammalian red blood cells perhaps provide such examples.

*Tetrahymena* may have the simplest version of genome rearrangements among ciliates studied. It has relatively modest levels of DNA elimination and chromosome fragmentation and probably the simplest form of DNA deletion. It is also the easiest to grow and manipulate in the laboratory. It should continue to offer the insights needed for a thorough understanding of this remarkable phenomenon of genome dynamics.

**ACKNOWLEDGMENTS**

We thank members of the Yao laboratory for inputs and comments. This work was supported by Academia Sinica and a grant (NSC 102-2311-B-001 –023 –MY3) from the Ministry of Science and Technology (Taiwan) to M.C.Y.

**REFERENCES**


