Programmed Rearrangement in Ciliates: *Paramecium*

MIREILLE BETERMIER¹ and SANDRA DUHARCOURT²

¹Institute of Integrative Biology of the Cell (I2BC), CNRS, CEA, Université Paris Sud, 1 avenue de la Terrasse, 91198 Gif-sur-Yvette cedex, France; ²Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, Paris, F-75205, France

ABSTRACT Programmed genome rearrangements in the ciliate *Paramecium* provide a nice illustration of the impact of transposons on genome evolution and plasticity. During the sexual cycle, development of the somatic macronucleus involves elimination of ~30% of the germline genome, including repeated DNA (e.g., transposons) and ~45,000 single-copy internal eliminated sequences (IES). IES excision is a precise cut-and-close process, in which double-stranded DNA cleavage at IES ends depends on PiggyMac, a domesticated piggyBac transposase. Genome-wide analysis has revealed that at least a fraction of IESs originate from *Tc/Mariner* transposons unrelated to piggyBac. Moreover, genomic sequences with no transposon origin, such as gene promoters, can be excised reproducibly as IESs, indicating that genome rearrangements contribute to the control of gene expression. How the system has evolved to allow elimination of DNA sequences with no recognizable conserved motif has been the subject of extensive research during the past two decades. Increasing evidence has accumulated for the participation of noncoding RNAs in epigenetic control of elimination for a subset of IESs, and in trans-generational inheritance of alternative rearrangement patterns. This chapter summarizes our current knowledge of the structure of the germline and somatic genomes for the model species *Paramecium tetraurelia*, and describes the DNA cleavage and repair factors that constitute the IES excision machinery. We present an overview of the role of specialized RNA interference machineries and their associated noncoding RNAs in the control of DNA elimination. Finally, we discuss how RNA-dependent modification and/or remodeling of chromatin may guide PiggyMac to its cognate cleavage sites.

INTRODUCTION Ciliates belong to a monophyletic group of unicellular eukaryotes within the Alveolate branch (1). The species that have been used as model organisms are free-living organisms, but parasitic or endosymbiotic ciliates have also been characterized (2, 3). A handful of ciliates have been studied, and common features could be deduced (reviewed in 4). They all carry motile and sensory cilia at the cell surface that allow swimming, food uptake, and the sensing of environmental signals. They present a characteristic nuclear dimorphism and undergo spectacular, genome-wide programmed rearrangements during development. The study of the mechanisms and regulation pathways underlying genome rearrangements has revealed a great diversity in the strategies used by different ciliates (5). The present chapter will focus on *Paramecium*, a widespread group of species that can be found on all continents. The sequences of the somatic genomes of two species, *Paramecium tetraurelia* and *Paramecium caudatum*, have been published recently (6, 7, 8). *P. tetraurelia*, which belongs to the *Paramecium aurelia* group of sibling species (9), is by far the most extensively studied species at the genomic level, and will be the main focus of the present chapter.

Nuclear dimorphism

*Paramecium* exhibits two functionally distinct types of nuclei that coexist in its cytoplasm at all stages of its life...
cycle. This unique nuclear dimorphism actually reflects the physical separation of the germline and somatic functions of chromosomes (10). P. tetraurelia harbors two identical diploid micronuclei (MICs), which divide by mitosis at each cell division. Under laboratory conditions, Paramecium grows vegetatively when fed on bacteria inoculated in plant infusions (11). During vegetative growth, the MICs are not transcribed and appear to be dispensable (12, 13). They are referred to as the germline nuclei because they undergo meiosis during the sexual processes and transmit the genetic information to the progeny. The single, highly polyploid somatic macronucleus (MAC, 800n) is responsible for gene transcription and governs the cell phenotype. During vegetative growth, the MAC divides in an unconventional manner (14), splitting in two halves without evidence of either chromosome condensation or formation of a classical mitotic spindle, which led to the hypothesis that MAC chromosomes do not carry centromeres. Even though its presence is essential for cell survival, the MAC is fragmented and lost during sexual processes, and a new somatic MAC has to differentiate from a copy of the zygotic nucleus in each cell of the next sexual generation.

Upon mild starvation, Paramecium reproduces sexually through conjugation between reactive partners from compatible mating types. In each species of the P. aurelia complex, two mating types—O and E—have been reported (15). At the cytological level, the cellular and nuclear reorganization that takes place during conjugation has been particularly well described in P. tetraurelia (Fig. 1A) (16). Mating starts with MIC meiosis, which takes place in each conjugating partner and gives rise to eight haploid nuclei. While MICs undergo meiosis, the parental MAC is fragmented into ~30 pieces, in which replication stops rapidly but transcriptional activity is maintained (17). Following meiosis, one haploid gametic nucleus, most likely selected randomly, moves to a specific compartment, the paroral cone, close to the interaction surface with the mating partner, and divides once to yield two genetically identical gametic nuclei. Fertilization involves the reciprocal exchange of one migratory gametic nucleus between the two partners. During karyogamy, fusion between the incoming migratory nucleus and the resident gametic nucleus gives rise to the diploid zygotic nucleus. The zygotic nucleus undergoes two successive mitotic divisions: the first takes place in the paroral cone of each mating partner, while the second takes place along the longitudinal axis of each cell, right after the separation of exconjugants. The latter division drives polar localization of the four mitotic products: the longitudinal orientation of mitotic spindles drives the positioning of two nuclei at the cell posterior pole, which triggers their differentiation into new MACs, while the two nuclei at the anterior pole will become the new MICs (16). The exact nature of the signal(s) that drive(s) MIC and MAC determination is not clear. Grandchamp and Beisson noticed that the second division of the zygotic nucleus is accompanied by a transient longitudinal contraction of the exconjugants, which become ~30% shorter in length. Their unpublished data (Fig. 1B) suggested that cell contraction activates asymmetrically located membrane mechanoreceptor channels (specific for Ca2+ at the anterior pole and for K+ at the posterior pole), which may form an intracellular ionic gradient between the two cell poles: Ca2+ at the anterior pole would favor MIC determination, while K+ at the posterior pole would trigger MAC determination (S. Grandchamp and J. Beisson, personal communication). Following nuclear determination, the two future new MACs adopt a central position in the cell where they start their differentiation in each exconjugant. This process, also called MAC development, involves massive endoduplication of the genome from 2n to 800n, during a period extending over two cell cycles (17). Four to five synchronous replication peaks take place during the first cell cycle, concomitant with large-scale programmed genome rearrangements (see below). At the first cell division (or caryonidal division), the two MICs divide through mitosis, and one developing new MAC segregates into each daughter cell. Replication in the new MAC switches to a more continuous mode after the first cell division, until the final ploidy is reached (17).

The peculiar genetics of P. aurelia

P. aurelia species are ideally suited for genetic analyses because of their two alternative modes of sexual reproduction. In conjugation, reciprocal exchange of one of the gametic nuclei between two cells of complementary mating types, followed by karyogamy, produces F1 cells that are always genetically identical, whatever the genotypes of the parental cells. The presence of different phenotypes in the two genetically identical F1 cells immediately reveals non-Mendelian characters. Known cases of maternally inherited phenotypes have been described, including the inheritance of mating-type determination in P. tetraurelia (18, 19). In addition to conjugation, P. aurelia species undergo autogamy, a self-fertilization process in which the two genetically identical gametic nuclei in each cell fuse within the same cell. This results in an entirely homozygous zygotic genome, which, in addition to the extremely low mutation rate observed in P. tetraurelia (20), considerably facilitates genomic analyses.
**FIGURE 1** MIC and MAC determination during the *P. tetraurelia* sexual cycle. (A) Nuclear reorganization during the *P. tetraurelia* sexual cycle. Conjugation occurs between cells with compatible mating types. Nuclear reorganization events take place in both mating partners, but details are represented in only one cell for clarity, with the parental MAC in blue. Following mating of two reactive cells (I), MIC meiosis starts, while the MAC remains intact (IIa). At meiosis II, eight haploid gametic nuclei are produced and the MAC begins its fragmentation process (IIb). In each partner, one meiotic product divides once to give rise to two identical gametic nuclei, and fertilization takes place through reciprocal exchange of one gametic nucleus (III). The remaining seven meiotic products are degraded. In each conjugating cell, a diploid zygotic nucleus is formed through the fusion of a resident and a migratory haploid nucleus (IV). The zygotic nucleus undergoes two mitotic divisions (Va and Vb), and exconjugants separate between the first and second divisions. During the second division of the zygotic nucleus, exconjugants shorten dramatically in size (Vb), which triggers the determination of two new MICs at the anterior cell pole (black) and two new MACs at the posterior pole. Programmed genome rearrangements take place in the developing new MACs (brown in VI). At the first cell division, the new MICs divide mitotically, and each of the two developing new MACs segregates into a daughter cell (VII), where it continues to amplify the rearranged genome to a final ploidy of ~800n. (B) MIC and MAC determination during cell contraction. To simplify the figure, the old MAC is not represented. Mechanical stimulation experiments have indicated that K⁺ mechanoreceptor channels (brown circles) are mostly located at the posterior pole while Ca²⁺ mechanoreceptor channels (red stars) are at the anterior pole (114). For wild-type cells (WT), an attractive hypothesis would be that the transient (~15 min) shortening of exconjugants exerts a pressure on the membrane, which activates mechanoreceptor channels and increases membrane permeability for Ca²⁺ at the anterior pole (where MICs are determined) and K⁺ at the posterior pole (where MACs are determined). Two unpublished experiments confirm that a pre-existing determinant, already present at the cell posterior pole before the polar positioning of the products of the second mitotic division of the zygotic nucleus, drives MAC determination during cell contraction (S. Grandchamp and J. Beisson, personal communication). (i) Amputation of the cell posterior part right before the shortening of exconjugants results in a large excess of progeny with four MICs. (ii) Manipulating the Ca²⁺ or K⁺ intracellular concentration using specific ionophores strongly disturbs nuclear determination: Ca²⁺ ionophores induce an excess of MICs, while K⁺ ionophores trigger an excess of MACs. doi:10.1128/microbiolspec.MDNA3-0035-2014.f1
STRUCTURE OF THE GERMLINE AND SOMATIC GENOMES

The somatic genome: breakthrough from genome sequencing programs

Although they both differentiate from identical mitotic copies of the zygotic nucleus, the MIC and MAC exhibit striking differences in their chromosome organization and genome structure (Fig. 2) (21).

According to pulse-field electrophoresis, MAC chromosomes are linear molecules of variable size ranging from 50 kbp to 1 Mbp (22). The ends of MAC chromosomes are capped by a mixture of G3T3 and G4T2 telomeric repeats, which are added by an error-prone telomerase at heterogeneous positions within a telomere-addition region that extends over ∼1 kb (23, 24). An additional level of heterogeneity has been attributed to the presence of several telomere-addition regions, distant by several kilobase pairs from each other, at the ends of MAC chromosomes (25, 26, 27). Thanks to joint efforts of the international community of Paramecium laboratories (28, 29, 30), a draft assembly of the MAC genome of P. tetraurelia was released in 2006 (7) and a polished version in 2012 (6). One haploid equivalent of the MAC genome consists of 72 Mbp of 72% AT-rich DNA, distributed into 188 scaffolds ranging from 45 kbp to 1 Mbp, ∼150 of which are terminated by telomeric repeats and therefore represent full-length somatic chromosomes. With regard to gene content, it has been known for a long time that Paramecium does not use the universal genetic code and translates UGA and UAG stop codons into glutamine (31, 32). Taking this into account, annotation of the MAC genome sequence revealed a strikingly high coding density (78%), with short intergenic regions (350 bp on average) and very little repeated DNA, except for a few segmental duplications (6, 7). Around 40,000 genes were annotated in the somatic genome of P. tetraurelia, as a consequence of three successive whole-genome duplications (7). Sequencing of the MAC genome from other species of the P. aurelia complex indicated that the most recent whole-genome duplication immediately preceded the burst of this group of species (7, 33). Inside genes, around 90,000 tiny introns (between 20 and 34 nt in length) were predicted and largely confirmed through expressed sequence tag sequencing (34) and high-throughput cDNA sequencing (O. Arnaiz, personal communication). Paramecium introns appear to be under selective pressure for the elimination of incorrectly spliced mRNAs through nonsense-mediated decay: indeed, 3n introns without in-frame stop codons (relative to their upstream exon) are significantly under-represented among the population of all introns (34).

FIGURE 2 Programmed genome rearrangements during MAC development. (A) General organization of MIC and MAC chromosomes and chromosome fragmentation. A representative MIC chromosome is shown on top, with repeated germline sequences (e.g., transposons) drawn as colored double-headed arrows. The exact structure of MIC telomeres (green boxes) is at present not known. During MAC development, each MIC chromosome is amplified ∼400-fold. Programmed heterogeneous elimination of repeated DNA (represented by vertical grey boxes) is associated with two alternative genome rearrangements: chromosome fragmentation and telomere addition to new MAC chromosome ends (blue boxes), or imprecise joining of the two chromosome arms that flank the eliminated region (dotted line). (B) Precise excision of internal eliminated sequences (IESs). The map represents a gene (white box) and its flanking regions. IESs are drawn as colored boxes and their precise excision is represented by dotted lines. Note that the scale is different from that used in (A). doi:10.1128/microbiolspec.MDNA3-0035-2014.f2
A glimpse of the organization of MIC chromosomes

The exact number of MIC chromosomes in *P. tetraurelia* is not known. An early cytological study based on azure A staining was performed on three other species of the *P. aurelia* group (33). Depending on the strain and the species, each MIC contains between 30 and 60 pairs of small poorly characterized germline chromosomes, barely distinguishable under the microscope during the first metaphase of meiosis. It should be noted that tiny chromosomes probably would have escaped detection under these experimental conditions, and accurate counting of MIC chromosomes will require the use of higher-resolution techniques. Since then, unraveling the structure of germline chromosomes has been the subject of intensive work. In 1992, J. Preer and colleagues used a cell fractionation procedure to purify vegetative MICs from *P. tetraurelia* and constructed the first MIC phage library. This library enabled them to compare the MIC and MAC versions of a limited number of genomic loci, leading to the landmark discovery that germline-specific interstitial sequences, called internal eliminated sequences (IESs), are absent from the MAC genome (36, 37). Thanks to the MIC library, a few dozen IESs were identified inside genes or in intergenic regions: IESs are flanked by two TA dinucleotides in the MIC, while a single TA is found at their excision site in the MAC (reviewed in 38). As the purification of pure MIC DNA remained a technical challenge, several tricks were used to gain further insight into the genome-wide distribution of IESs. Analysis of individual MAC sequence reads allowed the identification of several thousand polymorphic “TA indels,” present in only a fraction of reads for each given locus and representing putative IES excision errors, such as IES retention on some MAC copies of the locus or erroneous excision of genomic regions (39, 40). Significant progress was made in 2012, following the discovery of the PiggyMac gene (PGM), which encodes a domesticated transposase required to initiate programmed genome rearrangements, including IES excision (41). Following Pgm depletion, the germline genome is amplified to high ploidy levels in the developing new MAC (see below), thus providing a source of genomic “PGM DNA” enriched for non-rearranged sequences and suitable for next-generation sequencing (6). The complexity of PGM DNA was found to be around 100 Mbp, consistent with previous estimates of the complexity of the MIC genome (10). This indicates that 25 to 30 Mbp of MIC-restricted sequences are eliminated from the somatic MAC. Alignment of PGM DNA reads with the reference MAC genome allowed the identification of a genome-wide set of ~45,000 single-copy IESs (6). Strikingly, 47% of all genes are interrupted by at least one IES in the germline genome, which makes precise excision of IESs essential for the reconstitution of functional somatic genes.

Based on the estimated number of MIC chromosomes (30 to 60 pairs, which represents a minimal value) and the estimated complexity of the haploid germline genome (100 Mbp), the average size of MIC chromosomes would be around ~1 to 2 Mbp at most. This size range is significantly larger than that of MAC chromosomes (~150 kbp to 1 Mbp, based on MAC DNA assembly [7]). The difference can be attributed to chromosome fragmentation, which takes place during MAC development. Indeed, molecular analysis of a couple of fragmentation loci indicates that chromosome fragmentation does not depend upon the presence of a specific DNA sequence motif but is associated with imprecise elimination of large germline regions (over several kilobase pairs) located downstream of telomere addition sites (6, 27). Telomere addition at MAC chromosome ends or the formation of internal deletions through end joining were proposed to be alternative products of the same elimination event (27, 39, 42). A body of molecular evidence indicates that the eliminated germline regions located downstream of chromosome fragmentation sites harbor multicopy transposable elements and other repeated elements (6, 27, 41, 43).

All information available so far indicates that, during MAC development, *P. tetraurelia* discards 25 to 30 Mbp of its germline genome, representing a wide diversity of sequences (long or short, single-copy or repeated DNA, eliminated in a precise or heterogeneous manner). It is noteworthy that all IESs taken together represent 3 Mbp of germline DNA, which corresponds to only 10% of total eliminated DNA (6). Due to the presence of repeated DNA, the remaining germline eliminated sequences, which thus represent 90% of all eliminated DNA (22 to 27 Mbp), could not be assembled properly from the PGM DNA sequence reads. These reads could not be mapped to genomic regions collinear with MAC chromosomes. Therefore, these sequences most likely are included in chromosome fragmentation regions. Only the complete assembly and annotation of the germline genome sequence will provide a complete view of the structure of MIC chromosomes. Of particular interest will be the identification of transposable elements, germline-specific genes, and other chromosome features (discussed in 44).

Characteristic features of IESs

Among the different types of eliminated sequences that have been described in *Paramecium*, IESs have been by
far the most extensively studied. Analysis of the genome-wide set of 45,000 IESs (6, 43) confirmed the conclusions drawn from the study of the first IESs that were identified in the MIC phage library (reviewed in 38 and 46). IESs are invariably flanked by one TA dinucleotide at each end, and only a loosely conserved 8-bp consensus sequence (5′-TAYAGYNR-3′), reminiscent of the inverted repeats found at the ends of Telmariner transposons (47), could be deduced from the nucleotide sequence of their ends. Variations in this consensus sequence may exist according to IES size (46) or to specific requirements for efficient excision (48; see below).

In the mid-1990s, L. Klobutcher and G. Herrick proposed their “invasive, bloom, abdicate, and fade” (IBAF) model, according to which Paramecium IESs have derived from ancestral cut-and-paste transposons, which would have invaded the germline genome and decayed over time through internal deletions and base substitutions, while being kept under selection pressure for being precisely excised from the somatic genome (49). Interestingly, Telmariner transposons integrate preferentially into TA dinucleotides, which they duplicate on each side of the newly integrated copy (50): conservation of the ancestral target site duplication at IES boundaries indicates that the TA dinucleotide has become essential for IES excision. In agreement with the IBAF hypothesis, IESs from the genome-wide set are mostly short, single-copy, noncoding sequences, with 93% ranging between 26 and 150 bp long (6). The analysis of whole-genome duplications, which allow dating the arrival of individual IESs at particular genomic loci, confirmed that the shortest IESs also tend to be the oldest. Among the largest IESs (over 500 bp long), clearly recognizable portions of multicycopy Telmariner transposons were identified, providing further support to a transposon origin for at least a fraction of extant IESs. In parallel with the transposon origin of IESs, the system has also been co-opted to excise genomic fragments as IESs, providing an additional layer for the regulation of genome plasticity and gene expression (18).

Although IES excision is highly precise and efficient overall, excision errors may occur, as witnessed by the identification of IES excision polymorphisms (39). Evidence also exists in some cases for the use of alternative excision boundaries, which represent a fraction of the so-called TA indels (our unpublished observations; see also 6 and 51). When localized inside genes, these errors may be incorporated into transcripts, yielding non-functional mRNAs. Thus, even though IESs are excised from DNA and introns are spliced from RNA, interesting parallels may be drawn in P. tetraurelia between these two classes of elements. Both classes carry little sequence information at their boundaries, but their elimination is essentially precise, which raises the question of their recognition. In addition, exactly as was shown for introns, IESs and TA indels appear to be under selection pressure for recognition by mRNA quality-control systems, such as nonsense-mediated decay: 3n IESs (or TA indels) that would not introduce an in-frame premature stop codon are significantly under-represented among the population (6, 39).

**THE IES EXCISION TOOLBOX**

**Paramecium: a powerful model for the mechanistic study of programmed genome rearrangements**

The molecular mechanisms involved in programmed genome rearrangements in *Paramecium* have been particularly well studied for IES excision, essentially because of the precision of the reaction at the nucleotide level. The search for essential genes involved in IES excision started from a detailed molecular analysis of DNA intermediates produced *in vivo* during the course of MAC development and benefited greatly from the availability of the fully annotated somatic genome (7) and of the transcriptome during sexual processes (52). All genome-wide data have been integrated into a user-friendly database, ParameciumDB, which is continuously curated by the community (53, 54). The development of a simple and efficient RNA interference (RNAi) technique, now used routinely in *Paramecium* laboratories (55), has been of considerable help for the functional analysis of candidate genes, in combination with next-generation sequencing to monitor genome-wide the effect of each RNAi on IES excision (6, 43, 56). Finally, the use of green fluorescent protein (GFP) translational fusions has allowed the localization of each corresponding candidate protein to be monitored (Fig. 3). Several essential components of the core IES excision machinery have now been identified (Table 1), providing further support to a mechanistic connection between programmed DNA elimination and cut-and-paste transposition.

**Insights from the molecular analysis of IES excision intermediates**

Molecular characterization of IES excision intermediates produced during MAC development revealed that IESs are first amplified within the bulk of the germline genome during three to four endoduplication cycles, before excision starts (57). IES excision is essentially completed.
when caryonidal division takes place, although the excision machinery may still be active after this particular cell division (58). Ligation-mediated polymerase chain reaction (LM–PCR) experiments further established that IES excision is initiated by staggered double-strand cleavages centered on each flanking TA, generating DNA double-strand breaks (DSBs) with 4-nt 5’ overhangs (59). The conserved flanking TAs are essential for DNA cleavage, as confirmed by the inhibitory effect of a point mutation within one TA at either end of a given IES on IES excision (60, 61, 62, 63). Other positions within the consensus sequence may also be important (64, 65). Molecular analysis of a small set of IESs showed that programmed DSBs are detected over different periods of time depending on the IES (41, 63), suggesting that excision of particular IESs may be programmed to be completed early or late with regard to MAC development. Understanding whether DNA replication influences IES excision will require genome-wide monitoring of programmed genome rearrangements relative to DNA amplification.

Following DNA cleavage, IESs are released as linear molecules (66) and circularized in a second step to form precisely closed and supercoiled minicircles that do not replicate and are actively degraded at later stages during MAC development (46, 57, and M. Bétermier, unpublished data). At each chromosomal excision site, IES removal leaves a DSB that needs to be repaired efficiently and precisely to ensure that functional somatic chromosomes are assembled correctly. During both chromosome repair and IES circularization, the currently available model (Fig. 4A) proposes that broken DNA ends are aligned through the pairing of the TA dinucleotides that are present on each 5’ overhang, and undergo limited 5’ and 3’ processing before they are joined in a highly precise manner (59).

**piggyBac to the rescue**

The establishment of an evolutionary connection between *P. tetraurelia* IESs and *Tc/mariner* transposons raised a puzzling issue with regard to the origin of the endonuclease responsible for the cleavage of IES boundaries. The original IBAF hypothesis proposed that exaptation of a *Tc/mariner* transposase gene allowed ancestral transposons to decay and give rise to extant IESs. However, as discussed previously (46), the transposition of *Tc/mariner* transposons leaves a characteristic footprint at the donor site from which the transposon is excised before integrating at another target site (50), making it unlikely that a *Tc/mariner*-related protein allows the precise excision of *Paramecium* IESs.

It was possible to unravel this puzzle following a survey of the somatic genome annotation, which allowed identification of the gene encoding PiggyMac (Pgm), a domesticated PiggyBac transposase that is expressed specifically during MAC development and localizes exclusively in the new developing MACs, at the time when IES excision starts (41). When compared with PiggyBac transposases, Pgm harbors at least three distinct domains (Fig. 4B): a putative catalytic domain containing a triad of aspartic acids (DDD) typical of PiggyBac transposases, a short cysteine-rich region similar to the
TABLE 1  Genes involved in genome rearrangements in the new developing MAC

<table>
<thead>
<tr>
<th>Name</th>
<th>Putative function</th>
<th>Inductiona</th>
<th>Nuclear localizationb</th>
<th>RNAi or mutant phenotypec</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCL2</td>
<td>Dicer-like ribonuclease III scnRNA biogenesis</td>
<td>Early</td>
<td>A</td>
<td>Phenotype of Dcl2/Dcl3-depleted cells: Retention of a subset of IESs(^d)</td>
<td>(56, 91, 92)</td>
</tr>
<tr>
<td>DCL3</td>
<td>Dicer-like ribonuclease III scnRNA biogenesis</td>
<td>Early</td>
<td>ND(^f)</td>
<td>No scnRNAs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td></td>
</tr>
<tr>
<td>EZL1</td>
<td>H3K27 and H3K9 histone methyl transferase</td>
<td>Early</td>
<td>B</td>
<td>Retention of a subset of IESs(^e) No H3K27me3, no H3K9me3 Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td>(43)</td>
</tr>
<tr>
<td>NOWA1/2</td>
<td>GW repeat protein RNA binding</td>
<td>Early/intermediate</td>
<td>B</td>
<td>Retention of a subset of IESs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td>(85)</td>
</tr>
<tr>
<td>PTIWI09/01</td>
<td>Piwi protein scnRNA accumulation</td>
<td>Early</td>
<td>B</td>
<td>Retention of all IESs(^e) No scnRNA accumulation(^e) Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td>(92)</td>
</tr>
<tr>
<td>PTMB.220</td>
<td>DEAH-box RNA helicase</td>
<td>Early</td>
<td>B</td>
<td>Retention of a subset of IESs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td>(115)</td>
</tr>
<tr>
<td>SUMO I/II/III</td>
<td>protein modifier</td>
<td>Early</td>
<td>B</td>
<td>Retention of all IESs(^e)</td>
<td>(116)</td>
</tr>
<tr>
<td>DCL5</td>
<td>Dicer-like ribonuclease III iesRNA biogenesis</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of a subset of IESs(^e) No effect on transposon elimination(^e) No iesRNAs(^e)</td>
<td>(43, 56)</td>
</tr>
<tr>
<td>DIESa/b</td>
<td>unknown</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of all IESs(^e)</td>
<td>(117)</td>
</tr>
<tr>
<td>KU70a</td>
<td>Pgm partner C-NHEJ</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of all IESs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e)</td>
<td>(74)</td>
</tr>
<tr>
<td>KU80c</td>
<td>Pgm partner C-NHEJ</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of all IESs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e) No DSB(^e)</td>
<td>(74)</td>
</tr>
<tr>
<td>LIG4a/b</td>
<td>DNA ligase IV C-NHEJ</td>
<td>Early</td>
<td>C</td>
<td>DSB accumulation at IES ends(^e) No repair</td>
<td>(66)</td>
</tr>
<tr>
<td>PGM</td>
<td>domesticated transposase DNA cleavage</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of all IESs(^e) Inhibition of chromosome fragmentation/transposon elimination(^e) No DSB(^e)</td>
<td>(41)</td>
</tr>
<tr>
<td>UBA2</td>
<td>SUMO activating E1 enzyme</td>
<td>Intermediate</td>
<td>C</td>
<td>Retention of all IESs(^e)</td>
<td>(116)</td>
</tr>
<tr>
<td>DNA-PKcs</td>
<td>C-NHEJ</td>
<td>Early/intermediate</td>
<td>ND</td>
<td>Retention of a subset of IESs(^e) Lower DSB repair efficiency(^e)</td>
<td></td>
</tr>
<tr>
<td>mtf</td>
<td>nd</td>
<td>ND</td>
<td>ND</td>
<td>Retention of a subset of IESs(^e)</td>
<td>(118)</td>
</tr>
<tr>
<td>XRCC4</td>
<td>Lig4 partner</td>
<td>Early</td>
<td>ND</td>
<td>DSB accumulation at IES ends(^e)</td>
<td>(66)</td>
</tr>
</tbody>
</table>

\(^a\) Transcriptome analysis from microarray hybridization data (52). Early, induction peak during meiosis; intermediate, induction peak during MAC development; intermediate induction, transcription starts during MAC development and increases gradually until late stages.

\(^b\) Localization of GFP fusions expressed following transgene microinjection into the vegetative parental MAC. A, meiotic MAC; B, old then new MAC; C, new MAC only (see Fig. 3). For each pair of ohnologs, only the first gene of the pair (as indicated in the table) was fused to GFP.

\(^c\) When only a subset of IESs is affected in RNAi-treated or mutant cells, the exact identity and number of sensitive IESs and their levels of retention differ largely from one gene to the other (see text). Data on chromosome fragmentation and/or transposon elimination are indicated only for those genes for which this particular type of genome rearrangement was examined.

\(^d\) Whole-genome next-generation sequencing data.

\(^e\) Data obtained for a limited number of elements (PCR, Southern blot hybridization, LM-PCR, analysis of NGS reads).

\(^f\) ND, not determined.

J. K. Nowak, personal communication.

S. Malinsky, personal communication.
FIGURE 4 IES excision in *P. tetraurelia*. (A) Molecular mechanism and protein actors involved in IES excision. Both DNA strands are represented, with the IES in orange and flanking MAC-destined DNA in black. In the absence of any information about stoichiometry, the Pgm complex is represented by a red ball. Activation of DNA cleavage is thought to involve a physical interaction between Pgm and a development-specific Ku heterodimer (in grey). At each IES excision site (left), Ku is immediately positioned on the resulting DSBs and recruits all other actors of the classical nonhomologous end-joining pathway. In a first step, 5′ processing of the 4-nt overhangs, mediated by a nuclease that remains to be identified, results in the removal of the 5′-terminal nucleotide from each overhang. A gap-filling step involves addition of one nucleotide to each 3′-recessed end before the ligase IV–Xrcc4 (Lig4/X4) complex closes the junction. The linear excised IES (right) is thought to be circularized through a similar pathway. (B) Domain organization of the PiggyMac domesticated transposase (Pgm) and comparison with the PiggyBac transposase isolated from *Trichoplusia ni*. The boundaries of each domain are indicated (numbers refer to amino acid positions). The core transposase domain is in red, with the conserved catalytic aspartic acid triad (DDD) drawn as vertical bars and an upstream domain conserved in PiggyBac transposases represented by a hatched red box. The cysteine-rich domain that is proposed to fold into a putative PHD finger is in pink, and the C-terminal coiled-coil extension in purple. An additional N-terminal domain (in white) is found in the *Paramecium* protein. doi:10.1128/microbiolspec.MDNA3-0035-2014.f4
PHD finger found at the C terminus of the PiggyBac transposases (67), and a long C-terminal coiled-coil extension, which seems to be an innovation of ciliates (68). Strikingly, the DSBs characterized at IES ends have the same geometry as those catalyzed in vitro by PiggyBac transposases, with 4-nt 5′-protruding ends carrying a central TA (41, 67). Conservation of the DDD catalytic triad makes Pgm a good candidate for DNA cleavage during IES excision, although its catalytic activity still has to be demonstrated in vitro. Consistent with this hypothesis, no DSBs are detected at IES ends in the new developing MACs of cells silenced for expression of the PGM gene, while DNA endoduplication occurs normally (41). Supporting the hypothesis that Pgm is the catalytic subunit of the IES excision complex, the conserved DDD triad found in Tpb2p, the homologous protein from the ciliate *Tetrahymena thermophila*, was shown to be essential for in vitro DNA cleavage activity (68, 69).

PGM RNAi results in IES retention in the genome of the new developing MACs, causing subsequent massive cell death in the progeny. As described above, next-generation sequencing of the DNA of Pgm-depleted autogamous cells led to the identification of a reference set of 45,000 IESs (6). This set probably includes all IESs present in the MAC-collinear fraction of the germline genome, as indicated by the sequencing of Preer’s phage library, but whether Pgm is responsible for excision of all IESs needs to be confirmed by the sequencing of purified MIC DNA.

The discovery of Pgm revealed that the IES excision system in *Paramecium* has brought together two distinct transposon families: *Telmariner* (DNA substrates) and *piggyBac* (endonuclease). Pgm was also shown to be involved in the heterogeneous elimination of transposons and other repeated DNA that is associated with chromosome fragmentation (41), but whether this role is direct or indirect still has to be established. During IES excision, the formation of a transpososome-like complex is supported by genetic evidence indicating that the DNA cleavage step requires crosstalk between the two ends of each IES (63). Analysis of the genome-wide set of IESs revealed a remarkable periodicity in the size distribution of these sequences: 35% of IESs are found in a first peak (26 to 32 bp in length) and, up to 150 bp, regularly spaced peaks are conspicuous every 10 to 11 bp, which coincides with the phasing of the DNA double helix (6). Intriguingly, one peak (centered at 35 to 36 bp) is largely under-represented in the distribution, suggesting the existence of a “forbidden” size range. The biological significance of both the 10-bp periodicity and the forbidden peak will require further investigation. As discussed previously (6), the distribution of IES sizes may reflect strong constraints on the assembly of the excision complex, such as the bridging of IES ends through the binding of Pgm subunits that might interact directly for very short elements (26 to 28 bp), or the formation of a DNA loop—perhaps assisted by accessory DNA bending factors—for larger IESs (44 to 46 bp). Major issues will also have to be addressed regarding the molecular mechanisms involved in IES excision, in particular what directs the precision of DNA cleavage at IES boundaries and whether DNA hairpins similar to those introduced by piggyBac transposases are formed during DNA cleavage.

**Association of the classical nonhomologous end-joining pathway (C-NHEJ) with the Pgm endonuclease couples DNA cleavage and precise DSB repair**

When compared with cut-and-paste transposition, IES excision should rather be viewed as a “cut-and-close” reaction, during which the excision site is repaired precisely and the excised DNA is destined for degradation. As discussed in (45), precise and efficient DSB repair at IES excision sites ensures that MAC chromosomes are assembled in the right order and open reading frames are reconstituted with high precision. Using RNAi-mediated functional analysis of candidate genes, the ligase IV–Xrcc4 complex, an essential actor of the C-NHEJ pathway (70), was shown to carry out DSB repair at chromosomal IES excision sites and during the circularization of excised IESs (66). Interestingly, right before ligation, gap filling of recessed 3′ ends was inhibited in cells depleted for ligase IV (Fig. 4), suggesting that recruitment or activation of C-NHEJ DNA polymerases is coupled to the final ligation step (66).

The demonstration that the ligase IV–Xrcc4 specialized ligation complex carries out DSB repair at IES excision sites made *Paramecium* a relevant model to support the notion that the NHEJ pathway can be intrinsically precise, depending on the geometry of the broken ends that have to be joined (discussed in 71). Given the high number of IESs (one every 1 to 2 kbp along the germline genome), a major issue is how the NHEJ repair machinery is efficiently recruited to IES boundaries to carry out the joining of MAC-destined ends following Pgm-dependent cleavage. During general C-NHEJ-mediated DSB repair, broken ends are first bound by the Ku heterodimer and protected against extensive resection (reviewed in 72). Ku recruits the
DNA-dependent protein kinase DNA-PKcs to form the DNA-PK complex, which stabilizes the bridging of broken ends at the DSB, regulates end resection, and activates downstream C-NHEJ actors (73). *Paramecium* encodes two closely related Ku70 and three Ku80 homologs (66), and one homolog of DNA-PKcs (S. Malinsky, personal communication). Using a combination of reverse genetics and molecular approaches, it was established that a development-specific Ku70/Ku80 heterodimer is required in vivo for DNA cleavage at IES ends and interacts physically with Pgm in heterologous cell extracts (74). Depletion of DNA-PKcs results in the persistence of free chromosomal ends at abnormally late stages during MAC development for a subset of IESs, consistent with the expectation that DNA-PKcs is involved in DSB repair at IES excision sites (S. Malinsky, personal communication). Intriguingly, nonexcised copies of DNA-PKcs-dependent IESs persist in the replicating chromosomes upon DNA-PKcs depletion, suggesting a possible defect in DNA cleavage. Thus, although the details of the Ku-Pgm interaction and the mechanisms involved in Ku-mediated activation of Pgm-dependent DNA cleavage need to be analyzed, integration of DSB repair proteins into the catalytically active IES excision machinery provides an ideal way to channel cleaved IES boundaries towards precise C-NHEJ repair right after DNA cleavage.

**Homology-dependent maternal control of genome rearrangements**

An early clue that genome rearrangements are regulated in a homology-dependent manner came from studies in *Paramecium* showing that modifying the DNA content of the maternal macronucleus can profoundly alter the rearrangement profile in the new macronucleus of progeny after sexual events.

The first described example of maternal inheritance of a rearrangement profile came from a study of the d48 epimutant, in which a surface antigen gene is reproducibly deleted from the MAC genome although its germline MIC genome is wild type (25). The gene deletion in the MAC genome is inherited through sexual generations in a non-Mendelian manner as shown by genetic analyses (77). This alternative rearrangement profile is transmitted from the maternal MAC genome to the new MAC genome of the sexual progeny through the cytoplasmic lineage. Introduction of the missing gene in the MAC of d48 epimutant was shown to restore the wild-type rearrangement profile in the new MAC genome of sexual progeny (78, 79, 80). This provided the demonstration that presence of the gene in the maternal MAC genome is needed to maintain the gene in the new MAC genome of sexual progeny. Other examples of MAC deletions of nonessential genes were also shown to be maternally inherited (81, 82). As for d48, reintroduction of the missing gene by transformation of the maternal MAC rescued the wild-type rearrangement profile in the new MAC genome of sexual progeny (83), supporting the idea that the new MAC only maintains sequences that are present in the maternal MAC.

A similar regulation by the maternal MAC was observed for IES excision. Introduction of an IES sequence in the maternal MAC led to the retention of the homologous zygotic IES sequence in the new MAC of sexual progeny, but the excision of all other IESs occurred normally (84). Once induced by transformation, the retention of this IES in MAC chromosomes can further inhibit excision of the homologous germline IES in subsequent sexual generations. This is a perfect example of epigenetic inheritance because one particular IES is no longer excised, although the germline genome remains entirely wild type and the initial trigger for retention is absent. This sequence-specific maternal inhibition of IES excision was observed for a third of the IESs that were tested (5/13), which were called maternally controlled IESs (mcIESs) (48). Among this set of mcIESs, only one was submitted to epigenetic inheritance. A quantitative analysis of IES excision inhibition in a population of post-autogamous cells provided evidence that the higher...
the IES copy number in the maternal MAC, the higher the IES retention rate in the new MAC. Yet, excision inhibition is a stochastic process in each individual developing MAC, because a wide variability is observed among isolated clones derived from a single transformed clone. At the population level, the efficiency of inhibition varies from 0 to 100% among IESs, for a given IES copy number in the maternal MAC, suggesting that IESs differ in some intrinsic properties (48). Yet no obvious characteristics that could distinguish mcIESs from non-mcIESs were identified, even though a variant TATT consensus was found at the ends of some mcIESs. The classification of IESs into mcIESs or non-mcIESs categories, according to whether inhibition was observed, may be misleading as it does not reflect the gradient of quantitative differences observed between IESs. A putative RNA-binding protein, Nowa1, was shown to be essential for transposon elimination and IES excision (85). Because all known mcIESs require Nowa1 for excision, it was then proposed that Nowa-dependent IESs are under maternal control, although this has to be tested at a genome-wide level.

Because identical germline genomes can produce alternative rearrangement profiles in the zygotic genome—deletion/retention of a gene (or of an IES)—the fate of germline sequences appears to be not solely determined by cis-acting determinants but controlled, at least to some extent, by the DNA content of the maternal somatic MAC, which is still present in the cytoplasm during the development of the new one. These observations raised the idea that reproduction of pre-existing rearrangement profiles was achieved by a global comparison of germline and somatic genomes, which would eventually lead to the removal of all germline sequences that were absent from the maternal somatic MAC genome (reviewed in 86).

An RNA-dependent mechanism is involved in the maternal control of genome rearrangements

An important breakthrough was the identification of two classes of noncoding RNA (ncRNA) molecules with antagonistic functions that were shown to participate in the epigenetic regulation of genome rearrangements (Fig. 5).

A class of noncoding transcripts, produced by the maternal somatic genome, appears to protect homologous sequences against elimination. In Paramecium, reverse transcription-PCR experiments have provided evidence for the existence of ncRNAs produced at low levels from both strands by the somatic MAC genome. RNAi-targeted degradation of IES-containing transcripts prevented maternal inhibition of IES excision in the new macronucleus after sexual events (87). Similarly, RNAi-mediated silencing of a nonessential gene during vegetative growth triggered deletion of the homologous gene in the new MAC genome after sexual reproduction (83).

Targeting a small region of the gene is sufficient to delete the entire gene, suggesting that RNAi-mediated silencing leads to the degradation of the transcripts covering the entire gene, although these are not necessarily mRNAs. This is consistent with earlier studies showing that the ability of a maternal transgene to rescue MAC deletions of the homologous gene in sexual progeny does not require the expression of a full-length, functional mRNA (88, 89). Thus, in order to be maintained in the developing MAC, a sequence needs not only to be present but also to be transcribed in the maternal MAC. Based on these observations, it was proposed that somatic ncRNAs protect homologous zygotic sequences against DNA elimination during MAC development. This would rely on generalized transcription of the maternal somatic genome, which has not yet been investigated genome-wide in Paramecium. In support of this hypothesis, increasing evidence of pervasive transcription has been provided in various organisms (90).

A second class of ncRNAs is composed of small RNAs, called scnRNAs, that resemble the metazoan piRNAs. scnRNAs are produced from the maternal germline genome by a meiosis-specific RNAi pathway and promote elimination of homologous sequences in the developing new MAC. Production of scnRNAs, approximately 25 nt in length, requires two Dicer-like proteins, Dcl2 and Dcl3 (56, 91). Sequencing of scnRNAs reveals a 5′-UNG signature and indicates that scnRNAs are cleaved from double-stranded precursors to yield RNA duplexes with 2-nt 3′ overhangs at both ends (91). The scnRNA population is highly complex and does not correspond exclusively to MiC-specific sequences. Indeed, scnRNAs are produced from a large fraction of the germline genome (18, 56, 91), suggesting that the entire micronuclear genome is transcribed and produces scnRNAs. Knockdown of both Paramecium DCL2 and DCL3 results in the disappearance of scnRNAs, DNA elimination defects, and death in the sexual progeny (56, 91). Thus, Dicer-dependent scnRNAs are required for the elimination of germline-specific sequences. Further support of the role of scnRNAs in the DNA elimination process came from the functional analysis of two Piwi proteins, Ptiwi01 and Ptiwi09, necessary for scnRNA accumulation and DNA elimination (92). The most direct evidence
that scnRNAs can trigger DNA elimination in the developing MAC was obtained by injection of 25-nt synthetic RNA duplexes that mimic the structure of scnRNAs (87).

If, as suggested by the evidence above, scnRNAs are produced from the entire germline genome and are capable of promoting elimination of homologous DNA in the zygotic MAC, one needs to explain how the specificity of scnRNA induced deletions is achieved. The “genome-scanning” model posits that the highly complex population of scnRNAs is filtered out by base-pairing interactions with noncoding transcripts produced by the maternal somatic genome, resulting in the selective inactivation of those able to find a perfect

FIGURE 5 The genome-scanning model. The progression of nuclear events during autogamy is represented on top. The corresponding steps of the scanning model are schematized at the bottom. (i) MAC noncoding transcription. Noncoding transcripts are constitutively produced from the entire MAC genome. (ii) Biogenesis of scnRNAs. Early in meiosis I, 25-nt scnRNA duplexes are produced by the Dicer-like proteins Dcl2 and Dcl3 from the whole MIC genome. These duplexes are then transported to the cytoplasm where they are loaded onto Ptiwi01–Ptiwi09 complexes, leading to removal of the “passenger” strand. (iii) MIC-specific scnRNA selection. Ptiwi–scnRNA complexes are transported to the maternal MAC. Homologous pairing between scnRNAs and MAC noncoding transcripts, enhanced by the putative helicase Ptmb.220 and assisted by the putative RNA-binding proteins Nowa1–Nowa2, leads to selection of MIC-specific scnRNAs that are then transported to the developing MAC. (iv) DNA elimination. Recognition of scnRNA-homologous sequences in the developing MAC, enhanced by IES-specific small RNAs, guides recruitment of the endonuclease PiggyMac, allowing elimination of MIC-specific sequences. doi:10.1128/microbiolspec.MDNA3-0035-2014.f5
match, and thus in the selection of MIC-specific scnRNAs (Fig. 5). How MIC-specific scnRNA selection is achieved is not yet clear. High-throughput sequencing supports the idea that scnRNAs are progressively enriched in MIC-restricted sequences during MAC development (18, 56), and favors the hypothesis that the scanning procedure leads to the selective degradation of MAC scnRNAs. Once selected by this “scanning” procedure, which likely occurs in the maternal MAC, MIC-specific scnRNAs would be exported to the developing zygotic MAC to target homologous sequences, thereby recruiting the excision machinery. The idea that scnRNAs travel from the maternal MAC to the developing MAC is supported by the observation that GFP fusions of Nowa1 and Ptivi09 accumulate in the maternal MAC before the end of meiosis of the MIC and are later found in the developing MAC (Fig. 3 and Table 1).

The “genome-scanning” model was initially designed to account for the epigenetic inheritance of genome rearrangements, and it indeed provides a satisfactory explanation for all known cases of maternal inheritance: imprecise DNA elimination events and induced deletions, as well as regulation of mcIES excision. This process also accounts for the maternal inheritance of mating types in *P. tetraurelia*. The mating types, O and E, are not genetically determined in the MIC but are maternally inherited and controlled by the maternal MAC (19, 93). A recent study demonstrated that mating-type determination is achieved by an alternative rearrangement of the mtA gene during MAC development: in mating-type O clones, the mtA promoter is excised during MAC development by the scnRNA pathway, preventing expression of the mtA gene, while in mating-type E clones, the nonexcision of the mtA promoter drives the expression of the transmembrane protein mtA (18). The scnRNA pathway, very much like the piRNA pathway in metazoans, appears to be a genome defense mechanism that allows the removal from the somatic genome of transposons and their relics but can also regulate cellular genes and mediate epigenetic inheritance of phenotypic polymorphisms.

However, a “dark side” to this mechanism would in theory be unveiled during conjugation. Let us consider a cross between cells harboring two allelic IESs that have diverged to the point that scnRNAs produced by one allele can no longer recognize the other allele. Because of maternal inheritance of genome rearrangements, it is assumed that there is no significant exchange of cytoplasm between conjugating cells. Therefore, scnRNAs are independently sorted in each conjugating cell. Thus, in each exconjugant, scnRNAs homologous to the resident allele will be unable to excise the IES from the allele it has received from its conjugating partner, resulting in IES retention in the new MAC. Retention of an IES can lead to disruption of an essential gene and death of the sexual progeny. Given that IESs appear to evolve rapidly and that 16% of them have been acquired since the last whole-genome duplication (6), such mating incompatibilities may not be so rare. Since the *P. aurelia* group of 15 morphologically identical but sexually incompatible species arose shortly after the last whole-genome duplication (7, 33), it would be interesting to test whether the scnRNA pathway has played a role in driving speciation.

**Is the scnRNA pathway the tip of the iceberg for control of IES excision?**

Depletion of the Dicer-like proteins Dcl2 and Dcl3 impairs the excision of all known mcIESs. In agreement with the “genome-scanning” model, scnRNAs are thus required for the excision of mcIESs (43, 56, 91). However, high-throughput sequencing indicates that only a small fraction of IESs, less than 10%, are retained in the new MAC after co-depletion of Dcl2 and Dcl3, suggesting that the number of IESs subject to maternal control might be smaller than anticipated based on previous studies (48). In fact, DCL2 and DCL3 knockdowns have revealed that the vast majority of IESs are correctly excised in the complete absence of scnRNAs. Therefore, excision of non-mcIESs may not depend on scnRNAs (43, 56). Yet the Piwi proteins Ptivi01 and Ptivi09 are required for the excision of the small set of IESs that have been tested so far (92). Even though genome-wide studies of the effects of Ptivi01 and Ptivi09 depletions are needed to reach a definitive conclusion, all IESs appear to rely on Ptivi01/09 proteins for their efficient excision. One possibility is that excision of non-mcIESs relies on other small RNAs, which are not produced by Dcl2 and Dcl3 but are still loaded onto Ptivi01 and Ptivi09.

Recently, a novel class of 26 to 30 nt long, IES-specific small RNAs, called iesRNAs, was reported (56). They accumulate during late MAC development and require the Dicer-like protein Dcl5 for their biogenesis. Dcl5 depletion partially impairs the excision of about 10% of IESs (43), but no lethality is observed in the sexual progeny, likely because only a partial retention is observed for each affected IES. Dcl5-dependent IESs do not correspond to those retained after Dcl2 and Dcl3 co-depletion, and the precise role of iesRNAs in IES excision remains to be elucidated. Other sRNAs may still
remain to be discovered, because many Piwi proteins with unknown functions are found in the MAC genome, and some are expressed during sexual events and MAC development (92).

What drives the DNA cleavage machinery to its target sites?

DNA elimination in the developing MAC relies on accurate recruitment of the excision machinery to MIC-specific sequences. How this is achieved remains to be discovered. According to the scanning model, once selected, MIC-specific scnRNAs would be exported to the developing zygotic MAC and used to target homologous sequences, thereby defining the eliminated sequences and recruiting the excision machinery. As observed in small RNA-guided heterochromatin formation in other organisms (94), scnRNAs would promote the deposition of specific chromatin modifications at homologous loci in the developing MAC (Figs 5 and 6A). In T. thermophila, the scnRNA pathway is conserved, and scnRNA-mediated trimethylation of histone H3 on lysine 9 and lysine 27 (H3K9me3 and H3K27me3) (95, 96, 97, 98) is thought to guide recruitment of the Tpb2p endonuclease (69). As in Tetrahymena, imprecise elimination of long, repetitive germline sequences in P. tetraurelia relies on a putative lysine histone methyltransferase, Ezl1, responsible for H3K27 and H3K9 trimethylation in the developing MAC (43).

As discussed in (75), in the case of IESs, the situation is radically different because excision of these 45,000 DNA segments must require a marking mechanism of considerable precision, allowing the demarcation of these very short, numerous, interspersed germline sequences from adjacent retained somatic sequences. Indeed, the vast majority of IESs are shorter than 150 bp and one-third of them are between 26 and 30 bp in length. IESs are thus not even as long as the DNA wrapped around a single nucleosome. Yet high-throughput sequencing showed that Ezl1-dependent H3K27/K9 methylation is required for the precise excision of about 70% of IESs (43). Ezl1 might not trigger heterochromatin formation on IESs but instead might act locally and methylate histone H3 on one or a few nucleosomes that encompass the IES sequence. Precise mapping of histone marks along the genome will be needed to determine whether chromatin modifications are specifically associated with IESs.

ends (in particular the TA dinucleotide), DNA accessibility, and small RNAs that might guide DNA modifications or/and the Pgm endonuclease to its TA cleavage sites. doi:10.1128/microbiolspec.MDNA3-0035-2014.f6
How chromatin modifications might recruit the excision machinery is not yet clear. The Pgm endonuclease might have the capacity to recognize the marks loaded on eliminated sequences. Indeed, Pgm, like its Tetrahymena counterpart, contains a putative PHD finger, potentially able to directly bind histone marks (41, 69). However, other chromatin-interacting proteins might bridge histone modifications to the Pgm endonuclease via its putative coiled-coil domain (Fig. 6B). In Tetrahymena, conjugation-specific chromodomain proteins Pdd1p and Pdd3p can bind H3K27me3 and form higher-order heterochromatic structures at the nuclear periphery (98, 99). Clearly, the interplay with chromatin features requires further investigation in Paramecium, and in particular the involvement of other histone modifications and higher-order chromatin conformation needs to be explored.

One important question is to understand the mechanisms by which chromatin structure may position the excision machinery for precise DNA cleavages. The Pgm complex may be specifically targeted to IES ends due to its association with chromatin transition zones that mark the boundary between the internal side of IESs and their flanking regions (Fig. 6B). Strikingly, iesRNAs preferentially map to IES ends and may participate in the definition of IES boundaries (56). Where the DNA cleavages occur precisely may be determined, at least in part, by the DNA sequence found at IES ends, in particular the absolute requirement for the TA dinucleotide. As previously discussed (75), access to DNA cleavage sites internal to a nucleosome might be possible through the action of chromatin-remodeling complexes that could promote DNA “breathing” on the nucleosome, nucleosome sliding, or nucleosome eviction (Fig. 6C). In the case of short IESs, targeting might involve a particular nucleosome positioning, whereby nucleosome-free IESs would leave DNA segments accessible to the excision machinery. One possibility would be that sRNAs (scnRNAs, iesRNAs, or others) trigger DNA modifications that position the excision complex right at its cleavage site (discussed in 75). Another, nonexclusive hypothesis is that small RNAs directly guide the Pgm complex to its TA cleavage sites (Fig. 6C), as described for the RNA-dependent endonuclease Cas9 of the bacterial CRISPR system (100).

CONCLUSION

Recent advances in ciliate research have provided a remarkable example of the impact of domesticated transposases on genome plasticity. For other organisms, increasing evidence already supported the notion that genome rearrangements take place in somatic tissues during development, cell differentiation, or aging. Some rearrangements were proposed to involve the activation of transposon-related elements, as reported in mouse and Drosophila brain (101, 102), and/or the programmed elimination of variable amounts of germline DNA (103, 104, 105, 106, 107). Although the significance of somatic mosaicism remains poorly understood, genome rearrangements may contribute to plasticity and regulate gene expression by removing germline-restricted genes or allowing the assembly of functional genes. The molecular mechanisms triggering genome rearrangements are also largely unknown, except for a few cases. The best-documented example is the assembly of immunoglobulin genes in vertebrates, which takes place during lymphocyte differentiation and generates the diversity of the acquired immune response (108). The demonstration that the RAG1 component of the RAG1/RAG2 endonuclease responsible for V(D)J recombination is related to a Transib transposase provided the first evidence of transposase domestication in a cellular recombination system (109). Genome sequencing programs have revealed the existence of numerous other domesticated transposases in various organisms, but the function of these proteins has largely remained elusive (110, 111, 112, 113). The study of programmed genome rearrangements in Paramecium represents a unique opportunity to unravel the role of domesticated piggyBac transposases in genome plasticity, in a system in which DNA sites are not recognized through conspicuous sequence-specific protein–DNA interactions. Understanding how ncRNAs guide PiggyMac to its cleavage sites within the chromatin context of the developing new MAC will certainly provide new insights into the mechanisms that drive maternal inheritance of variant rearrangement patterns across sexual generations.

ACKNOWLEDGMENTS

We thank Jean Cohen and Janine Beisson for transmitting their invaluable knowledge of the biology of P. tetraurelia, and all members of our laboratories and those of Jean Cohen, Laurent Duret, Eric Meyer, and Linda Sperling for long-standing collaborations and stimulating discussions. We are grateful to Janine Beisson, Jean Cohen, and Linda Sperling for critical reading of our manuscript, to Simone Grandchamp, Janine Beisson and Jacek Nowak for allowing us to cite their unpublished results, and to Sophie Malinsky and Olivier Arnaiz for sharing their data prior to publication. Work in our two laboratories was supported by intramural funding from the CNRS, and by grants from the Agence Nationale de la Recherche (ANR 2010-BLAN-1603 and ANR-12-BSV6-0017). M.B. was supported by a grant from the Fondation ARC (#SFI20121205487) and S.D. by an ATIP Plus grant from the CNRS. We acknowledge the support of the CNRS.
Programmed Rearrangement in Ciliates: Paramecium

GDRE “Paramecium” Genome Dynamics and Evolution” and the European COST network “Ciliates as model systems to study genome evolution, mechanisms of non-Mendelian inheritance, and their roles in environmental adaptation network” (COST Action BM1102).

REFERENCES
Bettermier and Dubharcourt


Programmed Rearrangement in Ciliates: Paramecium


