Subtraction by addition: domesticated transposases in programmed DNA elimination

Jason A. Motl and Douglas L. Chalker1
Biology Department, Washington University in St. Louis, St. Louis, Missouri 63130, USA

The ciliate Paramecium tetraurelia must eliminate ~60,000 short sequences from its genome to generate uninterrupted coding sequences in its somatic macronucleus. In this issue of Genes & Development, Baudry and colleagues (pp. 2478–2483) identify the protein that excises these noncoding sequences: a domesticated piggyBac transposase that has been adapted to remove what are likely the remnants of transposon insertions. This new study reveals how addition of a transposase to small RNA-directed silencing machinery can guide major genome reorganization.

In the midst of studying chromosomal behavior during cell division, a body of work that provided support for his chromosomal theory of inheritance, Boveri [1887] described a peculiar genomic phenomenon of massive DNA elimination that he called “chromatin diminution”. Carefully watching embryonic development of the roundworm Parascaris univalens, Boveri [1887] observed that the chromosomes of the developing somatic cells became highly fragmented during early cleavages. The central fragments segregated to the poles, becoming the cells’ somatic chromosomes, while the remaining chromatim remained in the cytoplasm and was eventually degraded (see Goday and Pimpinelli 1984; Muller et al. 1996; Muller and Tobler 2000). This is not the way one expects an organism to treat its genome. Nevertheless, several examples of large-scale DNA elimination events have since been documented in a taxonomically diverse array of organisms, including sciarid flies, lampreys, copepods, hagfish, and ciliated protozoa (Beerman 1977; Kubota et al. 1993; for review, see Prescott 1994; Goday and Esteban 2001; Smith et al. 2009), which suggests such genome-downsizing must offer some advantage when it occurs.

Another thing that these unorthodox phenomena reveal is that large blocks of genomes are expendable, at least in somatic cells. The fact that such DNA elimination events are developmentally programmed indicates that the cells in which they occur have regulated mechanisms to recognize superfluous DNA and remove it. The eliminated DNA, in many cases, consists largely of repetitive sequences, which are often thought of as junk DNA. DNA elimination appears to be these cells’ means to clear the junk out of the attic. More conventional eukaryotes simply store their repetitive DNA in the attic of silent heterochromatin. The results of several recent studies suggest that the recognition of DNA to be packaged into heterochromatin may be a precursor to genome-wide DNA elimination. A report from the research group of Betermier [Baudry et al. 2009] in this issue of Genes & Development reveals that a domesticated piggyBac transposase has been recruited to carry out programmed DNA elimination in the ciliate Paramecium tetraurelia. Previous work in Paramecium and related ciliates has shown that an RNAi-related machinery is also essential for these DNA rearrangements, providing a key link between DNA elimination and heterochromatin formation [for review, see Yao and Chao 2005; Duharcourt et al. 2009]. Thus, the simple addition of a transposase to an existing silencing pathway provides the means to subtract unwanted DNA from the genome.

Large amounts of DNA are eliminated from the somatic macronucleus (MAC)

Ciliates, such as Paramecium, are single-celled protozoa that carry two different copies of their genome, each of which is contained in a functionally distinct nucleus [for review, see Prescott 1994]. Micronuclei (MICs) harbor the transcriptionally silent germline genome, ready to supply their genetic material to the next generation; MACs carry the somatic genome, providing for all gene expression in the current generation. The somatic genome is a streamlined version of the germline, having undergone extensive elimination of DNA sequences and fragmentation of its chromosomes during its differentiation. This differentiation occurs during sexual reproduction upon conjugating with a partner cell or, in some ciliates, a self-fertilization process called autogamy [see Meyer and Chalker 2007]. During conjugation or autogamy, the parental MAC DNA is eventually discarded and the
germline MIC provides copies of the genome for both the new MICs and the MACs of the progeny cells. While both the MIC and MAC progenitors start with identical copies of the genome, the differentiation of the somatic macro-nucleus removes large portions of this DNA (up to 95% of the germline genome is some ciliate species). It seems reasonable to classify much of the DNA that is eliminated from the Paramecium MAC as junk; it is discarded, after all. Most of the repetitive sequences, such as DNA satellites and transposon-like elements, as well as numerous noncoding single-copy sequences, are the target of genome reorganization (for review, see Betermier 2004). The repetitive elements are eliminated through a somewhat imprecise process that deletes hundreds to thousands of base pairs. If the broken ends generated during the excision of a particular germline-limited sequence fail to be joined, the chromosome fragments are stabilized by de novo telomere addition (Le Mouel et al. 2003). The single-copy sequences removed are short (26–886 base pairs [bp]), AT-rich, noncoding sequences, referred to as internal eliminated sequences (IESs), which number from 50,000 to 65,000 per haploid genome. IESs are largely located within the body of genes and must be precisely excised to create uninterrupted coding sequences in the MAC, making their efficient removal an essential process. Both IESs and the repetitive sequences are bound by TA dinucleotides, for which only one copy is retained upon their removal from the genome (see Fig. 1A). For IESs, the TA dinucleotides are imbedded in a larger consensus sequence of 5’-tggTAYAGYNR-3’ that is similar to the terminal inverted repeats (TIRs) of germline-limited Tec elements of the distantly related ciliate Euplotes crassus, which are members of the widespread Tec/mariner family of transposons. This finding provided support for the idea that IESs are derived from transposon insertions whose sequence subsequently degenerated until only abbreviated TIRs are recognizable (see Fig. 1B; Klobutcher and Herrick 1995, 1997).

PiggyMac: the domestication of a transposase

The idea that IESs are the remnants of transposons has always been an attractive hypothesis, but whether transposons actively participate in programmed DNA elimination

Figure 1. Pgm-mediated DNA elimination of an IES and the evolution of IESs and a domesticated transposase. (A) IESs in P. tetraurelia contain TIR containing a conserved TA dinucleotide repeat. Concomitant DNA DSBs of the IES generate 4-bp 5’ overhangs, which are paired at the central TA and are repaired by DNA ligase IV and other host factors. (B) Domestication of an IES transposase. (Top line) A generic transposon undergoes mutation that causes loss of a functional transposase ORF and degeneration of inverted repeats (purple arrow) retaining only the TA dinucleotide and the minimally needed recognition sequence. Transposon undergoes further degeneration (red boxes) to create an IES. (Bottom lines) An ancestral piggyBac transposon becomes fixed in the euchromatin of the MAC during macronuclear development by loss of its TIRs. Acquisition of heterochromatin protein-binding capacity by the retained transposase creates a developmentally regulated chimeric protein that is able to be recruited to silent heterochromatin. Further modulation of the inverted repeat cut site allows more promiscuous recognition of IES boundary elements. Now, upon RNAi-directed heterochromatin formation in ciliates, these heterochromatic sequences are removed from the MAC during development to create a transposon-free somatic MAC.
was not known. This new report from Baudry et al. (2009) shows that Paramecium uses what was once a transposase from a piggyBac element to reorganize its somatic genome. When this putative endogenous transposase was fused to green fluorescent protein, it localized to developing MACs where DNA elimination occurs, leading Baudry et al. (2009) to call this gene PiggyMac [PGM]. To further assess the possibility that Pgm participated in DNA elimination, the group knocked down PGM levels by feeding Paramecium bacterial cells expressing PGM-specific dsRNA prior to inducing autogamy. This treatment resulted in a 20-fold reduction in progeny and thus revealed that the gene product has a critical function during development (Baudry et al. 2009).

If PGM actually encodes the enzyme responsible for IES excision, a process that occurs quite late in development, disrupting late-stage expression should be sufficient to perturb production of progeny. To examine the timing of Pgm function, Baudry et al. (2009) exploited the fact that mating cell pairs separate and begin refeeding before the developing MACs carry out DNA elimination (Berger 1973). Feeding cells PGM dsRNA early (before conjugation starts) or late (after pairs separated) interfered with normal development. In control experiments, RNAi knockdown of the meiosis-specific Spo11 homolog [PtSpo11] only elicited a phenotype if feeding occurred prior to meiosis. Thus, the critical time window when PGM must be active is the period when IES removal from the genome occurs.

Of course, if Pgm is the enzyme that carries out DNA elimination in the Paramecium developing MAC, RNAi knockdown should lead to failed excision. To test this possibility, Baudry et al. (2009) used ligation-mediated PCR to assay the ends of a known IES for the appearance of double-strand breaks [DSBs] that normally occur during DNA elimination. In the PGM knockdown cells, they failed to detect breaks that were readily observed in wild-type cells. This result showed that Pgm is required to remove IESs from the genome. All the data taken together strongly support the idea that Pgm is a domesticated transposase that has been recruited to perform programmed DNA elimination.

Not only has the discovery of this domesticated transposase provided needed insight into the mechanism of DNA elimination in Paramecium, it has helped clarify some apparently disparate observations reported in previous work. For example, the consensus sequence flanking Paramecium IESs is reminiscent of Tc/mariner TIRs; however, the type of DSB characterized at the end of each IES did not match that expected for this family of elements [see Fig. 1A; Gratias and Betermier 2003]. The Betermier group (Gratias and Betermier 2003) had shown previously that IESs are removed by a mechanism that generates a 4-base staggered cut, leaving 5’ overhangs; Tc transposases generate 2-base 3’ overhangs [van Luenen et al. 1994]. Like the cleavage observed at IES ends, piggyBac transposases generate 4-base, 5’ overhangs of the sequence TTAA [Mitra et al. 2008]. Note that these overhangs are centered around a TA dinucleotide. The IES TIRs may look like Tc ends, and may even be derived from such transposons, but it is piggyMac that is able to recognize and cleave at IES boundaries (Fig. 1).

It is interesting to note that Baudry et al. (2009) report that piggyBac-like transposases are encoded in the genome of the related ciliate Tetrahymena thermophila. The IESs of this ciliate do not share a common TIR sequence or even possess conserved TA dinucleotides at their ends. In fact, mechanistic studies in Tetrahymena revealed that the sequences at the boundaries of IESs were not essential. Instead, sequences located a short distance (45–50 bp) outside the excision boundaries, which are in the DNA that remains in the somatic genome, were shown to determine the boundaries of excision [Godiska et al. 1993; Chalker et al. 1999; Patil and Karrer 2000; Fillingham et al. 2001]. These regulatory sequences that specify the excision boundaries share little, if any, similarity with one another, a fact that has always been hard to reconcile with a transposon origin. The one common observation is that both ciliates generated 4-base, staggered cuts upon IES excision. As the Baudry et al. (2009) study suggests, the proteins that excise both Tetrahymena and Paramecium IESs may share a common piggyBac origin. In support of this, a piggyBac protein encoded in the Tetrahymena genome has been found to be essential for DNA elimination [CY Cheng and MC Yao, pers. comm.].

**Going from heterochromatin to DNA elimination: one small step for ciliate-kind?**

The discovery of these domesticated piggyBac transposases identifies key players in programmed DNA rearrangement that have been missing in the understanding of these events. Recent research has focused on elucidating the role of RNAi-directed heterochromatin formation in DNA elimination [for review, see Yao and Chao 2005; Duharcourt et al. 2009]. The exciting discovery that small RNAs, which are homologous to IESs, associate with a Piwi family protein that is essential for DNA elimination in Tetrahymena took the emphasis away from the possible role of transposons in mediating their own excision [Mochizuki et al. 2002]. These small RNAs were shown to direct heterochromatin-associated modifications, histone H3 Lys 9 and Lys 27 methylation, to the IESs, marking them for elimination [Taverna et al. 2002; Liu et al. 2007]. Further characterization of these RNAi-directed mechanisms demonstrated that both Tetrahymena and Paramecium begin early in their development to decipher what sequences should be eliminated, well before the transposases are expressed. At the start of meiosis, the germline MIC genome is bidirectionally transcribed [Chalker and Yao 2001], and the resulting dsRNA transcripts are processed into 25- to 30-nucleotide [nt] small RNAs [depending on the species] by Dicer-like ribonucleases [Malone et al. 2005; Mochizuki and Gorovsky 2005; Lepere et al. 2008, 2009]. Later in development, these meiosis-specific small RNAs move to the developing somatic MACs and, at least in Tetrahymena, direct heterochromatin modifications to IESs. It is not clear whether small RNAs direct chromatin modifications
in *Paramecium*, as their histone H3 proteins possess divergent N-terminal tails, making assessment of modifications challenging. These findings suggest that DNA elimination is driven by the host cells trying to silence transposons (and other junk DNA), not simply transposons moving themselves around.

Massive DNA elimination does not seem nearly so unorthodox when one realizes that the mechanisms ciliates use to mark sequences for removal are analogous to those that many organisms use to silence transposons in the germline and elsewhere. Piwi-associated RNAs (piRNAs) of the fruit fly *Drosophila melanogaster* represent a well-studied case of small RNAs acting in germ cells to keep transposons in check (for review, see Saito et al. 2006; Vagin et al. 2006; Malone and Hannon 2009). Plants also use their RNAi machinery to silence transposons in pollen (Slotkin et al. 2009) and somatic tissues (for review, see Matzke et al. 2009). These mechanisms are obviously quite ancient, as they are found in animals, plants, and protists. The discovery of a domesticated transposase as the mediator of wholesale DNA elimination of the sequences recognized by the RNAi machinery provides a vision of how such a phenomenon might evolve.

With the identification of Pgm as an essential player in DNA elimination, ideas as to the origins and excision mechanisms of IESs have come full circle. IESs were often postulated to be degenerate transposons, and this is still the most likely scenario [Fig. 1B]. The ancestral cells would be expected to be equipped with an RNAi mechanism to silence these invaders and limit their spread. Along the way, an invading *piggyBac* element in the genome lost its TIRs and could no longer mobilize itself. This change or other alterations allowed it to escape silencing, and it eventually became a fixture within euchromatin [Fig. 1B]. At this point, it may still be able to mobilize other elements in *trans*, which could have led to the large expansion of IESs. If this semidomesticated transposase was responsible for spreading sequences around the genome, it should also retain the specificity to remove them, as *piggyBac* elements are cut-and-paste transposons.

It is not hard to imagine how this transposase could become an “excisase” favoring DNA elimination over mobilization. Alteration of the enzyme’s affinity for TIRs could decrease its ability to both cut and paste. To compensate for this decreased ability to recognize TIRs of the future IESs, the protein would need to acquire the ability to interact with the silent chromatin associated with these undesirable sequences. Pgm has a C-terminal extension that Baudry et al. [2009] speculate could give it the ability to interact with chromatin modified by the RNAi machinery. The *Tetrahymena* protein also has an extended C terminus, which could suggest that that addition occurred in a common ancestor. This adaptation could also allow the transposase to recognize a larger variety of IES end sequences, as protein–protein interactions would be the main force recruiting the excisase to its sites of action.

It is somewhat paradoxical that *Paramecium* has recruited a transposase to eliminate transposons and other extra DNA from its somatic genome. Originally, PGM could have been considered part of the junk DNA, but it is junk no longer. It has taken on an essential role in the major events shaping the somatic genome during its differentiation from the germline. The possibility that a transposase is being used to remove undesirable sequences like other transposons should not be overtly surprising, as several examples of transposon protein domestication have been described (for review, see Volff 2006). Perhaps the most well known is the RAG1/RAG2 recombinase of jawed vertebrates that initiates VDJ recombination in B and T cells [Agrawal et al. 1998; Hiom et al. 1998]. Pgm is just the most recent example of the hijacking of a transposase to carry out programmed DNA rearrangement. However, domesticated transposases are by no means limited to acting as transposases as they did in their undomesticated state. For example, a likely former hAT transposase, Daysleeper, acts as a transcriptional activator in *Arabidopsis thaliana* [Bundock and Hooykaas 2005]. Furthermore, the former pogo-like transposase CenP-B is an important centromeric protein conserved in diverse organisms, from fission yeast to mammals [Smit and Riggs 1996; Casola et al. 2008].

Given that ciliates go to a lot of work to eliminate transposons and other nonessential DNA from their somatic genome, they must gain something from it. One simple gain is the removal of the transposons from the transcriptionally active nucleus. This should be an effective way to limit their spread and prevent possible deleterious mutations that might arise due to transposition. Possibly a more important force derives from the polyploid state of the somatic genome. By having polyploid MACs, ciliates possess a potential for high-level gene expression that comes with having tens of thousands of copies of each gene. Thus, these organisms are able to obtain a relatively large size for single cells that in turn allows them to have rather big mouths, thus increasing their range of food options. The high level of gene expression still allows them to divide quickly relative to their size, which is important to survive as a species when lots of other creatures see them as lunch as they are scavenging for theirs. By streamlining their polyploid somatic genome, they have that much less DNA to replicate and can divide that much faster.

This discovery of *piggyMac* provides an interesting glimpse into the evolution of genome-wide DNA elimination. Some common ancestor of *Paramecium* and *Tetrahymena* was likely less drastic in its dealing with the junk in its genome, simply packaging it up into heterochromatin as most eukaryotes do. The maintenance of a separate germline genome and the domestication of a transposase allowed this silenced DNA to be discarded from the soma. All of the ciliates studied exhibit some degree of programmed DNA elimination, so it would be reasonable to assume that the DNA elimination mechanism arose early in this lineage. Even if that is the case, it has been shown recently in *Oxytricha trifallax* that the non-*piggyBac* TBE transposase is necessary for the elimination of the several thousand TBE elements and at least some IESs [Nowacki et al. 2009].
Thus, there have been at least two separate transposon domestication events associated with DNA elimination in ciliates. Extending this observation beyond ciliates, one cannot help but wonder whether other chromatin diminution events [e.g., the fragmentation of chromosomes in *Parasarcis* observed by Boveri [1887]] are the result of domesticated transposons acting on silent heterochromatin. Once cells have a means (e.g., RNAi) to recognize the foreign DNA in their genome, picking up a transposase is a handy way to eliminate them altogether.

**Acknowledgments**

We thank Dr. D. Frank for comments on the manuscript. Both J.A.M. and D.L.C. are supported by NIH grant GM069593 and NSF grant MCB-0642162.

**References**


A piggyBac transposase in *Paramecium*.


Subtraction by addition: domesticated transposases in programmed DNA elimination

Jason A. Motl and Douglas L. Chalker

*Genes Dev.* 2009, 23: Access the most recent version at doi:10.1101/gad.1864609

<table>
<thead>
<tr>
<th>Related Content</th>
<th>PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Céline Baudry, Sophie Malinsky, Matthieu Restituito, et al.</td>
</tr>
<tr>
<td></td>
<td><em>Genes Dev.</em> November, 2009 23: 2478-2483</td>
</tr>
</tbody>
</table>

| References            | This article cites 44 articles, 17 of which can be accessed free at: http://genesdev.cshlp.org/content/23/21/2455.full.html#ref-list-1 |
|                       | Articles cited in: http://genesdev.cshlp.org/content/23/21/2455.full.html#related-urls                                       |

| License               |                                                                    |

| Email Alerting Service| Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here. |

Biofluids too dilute to detect microRNAs? See what to do.