

Transposon debris in ciliate genomes

Yi Feng¹, Laura Landweber¹

¹ Departments of Biochemistry & Molecular Biophysics and Biological Sciences, Columbia University, New York, NY, 10032 USA

Abstract

The germline genomes of ciliated protists are replete with “junk” DNA insertions that need to be removed for gene expression. Unlike introns, these are spliced as DNA. What is their source and why are they so abundant? A new study in *PLoS Biology* supports a classic model of transposon origins.

Body start

A genome encodes all the primary information that DNA-based organisms need through their lifespan. Though it has been assumed to be identical in most cells of Metazoa, some animals, such as nematodes, sea lampreys, and songbirds contain multiple genomes that usually differ between germline and somatic cells [1]. Another example of somatic rearrangement is V(D)J recombination in jawed vertebrates, which permits the production of diverse antibodies and receptors for adaptive immunity [1]. In addition to these examples in Metazoa, ciliates, which are microbial eukaryotes, contain two distinct genomes packaged in a somatic macronucleus (MAC) and a germline micronucleus (MIC) within the same cytoplasm. Only the MAC genome is transcriptionally active during asexual (vegetative) growth. During post-zygotic development, the MAC genome rearranges from a copy of the MIC genome by deleting internal

eliminated sequences (**IESs**) and other repetitive regions from the DNA, together with rejoining of flanking MAC-destined segments. Though analogous to RNA splicing, these rearrangements occur on DNA. *Paramecium tetraurelia* removes ~45,000 IESs during genome differentiation. Most interrupt coding regions [2]. Therefore, precise elimination is necessary for proper gene function. This gives rise to a long-standing question in the field: how did IESs accumulate if their presence is deleterious?

A study published in the current issue of *PLoS Biology* [3] provides evidence to support a model proposed by Klobutcher and Herrick in 1997 that linked the origin of IESs to transposable elements (TEs) in ciliate genomes [4]. TEs play a key role in genome rearrangement in other cells, and are occasionally domesticated by the host. A classic example is the origin of RAG1 recombinase by domestication of a *Transib* transposase [2]. In humans, mutations in the PiggyBac TE derived 5 (PGBD5) gene trigger aberrant genome rearrangement in some childhood cancers, emphasizing the importance of regulating genes derived from transposon domestication [5]. In ciliates, multiple lines of evidence support the role of TE-derived genes in DNA elimination. In *Oxytricha trifallax*, over one billion years diverged from *Paramecium*, tens of thousands of MIC-limited TEs from the Tc1/mariner family are expressed during development, and silencing of transposase genes via RNAi stalls DNA elimination [6]. The *Paramecium* and *Tetrahymena* lineages domesticated transposase genes from the piggyBac family, and their silencing also prevents correct DNA elimination [7,8]. Further, an analysis of 20 *Paramecium* IESs suggested that they could be decayed transposons [9]. The IESs share an 8 bp consensus 5'-TAYAGYNR-3' on the boundary, similar to sequences flanking Tc1/mariner transposons [9]. Based on these observations, it is necessary to revisit the Klobutcher and Herrick model.

Like a whimsical [M.C. Escher metamorphosis print](#), the model describes four stages to transform transposons into IESs: Invasion, Bloom, Abdicate and Fade, or “IBAF” (Figure 1). An autonomous transposon first *invades* the host genome, bringing with it its encoded transposase to cut-and-paste itself. Once there are enough transposon insertions in gene regions, it would become beneficial for the host to permit sufficient translation of transposase protein to excise these insertions. The germline micronucleus provides a safe haven for transposons to flourish (*bloom*), as long as they are safely eliminated before formation of the macronucleus. Indeed, modern ciliate germline genomes may still be prone to transposon invasion, but the trait of ciliate nuclear dimorphism, itself, may have evolved as an evolutionary strategy to corral them. Some hosts could eventually domesticate or repurpose a transposase gene by placing it under control of its own promoters, which would relax the selection pressure on the transposase ORF encoded in TEs, hence *abdicating* their role as excisase and allowing more mutations to accumulate across the crippled TE. The decayed TEs may rapidly atrophy to short IESs and lose the protein-coding ability (*fade*), as long as their boundary sequences can still be recognized and removed efficiently. While this model may explain the origin of IESs, there has been mostly indirect evidence to support it (e.g., the resemblance of sequences flanking IESs to TE target site motifs; 10). Sellis, Guérin and colleagues provide compelling evidence that thousands of IESs derive from TEs by comparative genomics within the *Paramecium* genus [3].

In this study, Sellis *et al.* annotated and dated IESs in 9 diverse species of *Paramecium* [3]. They sequenced both genomes in each species and annotated IESs by comparing short MIC reads to the MAC genome assemblies. IESs were classified as old, intermediate, or recent on a phylogeny. To explore whether IESs derived from TEs, they clustered IESs to reveal shared ancestry. While the majority of IESs are unique, 24 families contained at least 10 homologous

IESs across the *Paramecium* species. Notably, 4 families are homologous to the IS630-Tc1-mariner transposon superfamily. 97.5% of these clustered, mobile IESs seem to be recent insertions, much higher than the 9.5% of overall IESs. This is expected, as older IESs accumulate substitutions, obscuring homology. Two species of different subclades, *P. sonneborni* and *P. tredecaurelia* were found to share several mobile IES families, likely the result of horizontal gene transfer.

Old IESs are less often under epigenetic control than new ones [3]. In *P. tetraurelia*, 70% of IESs require histone methylation or small RNAs for removal, while the rest are independent of either pathway [11]. The length distribution also differs between older and newer IESs. Older IESs are much shorter, compatible with the “fade” step of the model. Moreover, some conserved IESs encode MIC-limited genes [3, 12] or provide a “switch” to regulate gene expression.

This study demonstrates a workflow for studying IES evolution within a lineage. In the future, it will be illuminating to compare mobile IES families across clades. Unlike the consensus “TA” flanking IESs in *Paramecium*, IESs in some other lineages are flanked by diverse microhomologous sequences, though the shortest repeats still resemble transposon ends or target site duplications [10], harkening back to likely transposon origins. Long-read genome assemblies may provide a more complete portrait of transposon-IES evolution in ciliates.

Sellis *et al.* suggest that IESs dispersed across the genome within a clade may derive from TEs. Identifying the initial invading TE insertions is an open question in the IBAF model. This model is also similar to one proposed for intron evolution. Recent evidence demonstrated that DNA transposons have generated hundreds to thousands of introns in two species separated by one billion years [13]. Both programmed DNA deletion and intron splicing offer strategies to remove harmful insertions from functional genes. This opportunity for accurate deletion assuages

their deleteriousness, permitting a stealth bloom of TE insertions, that may also beget innovation, permitting the evolution of domestication [7,8] or mutualism [2,6] between host and genomic parasite.

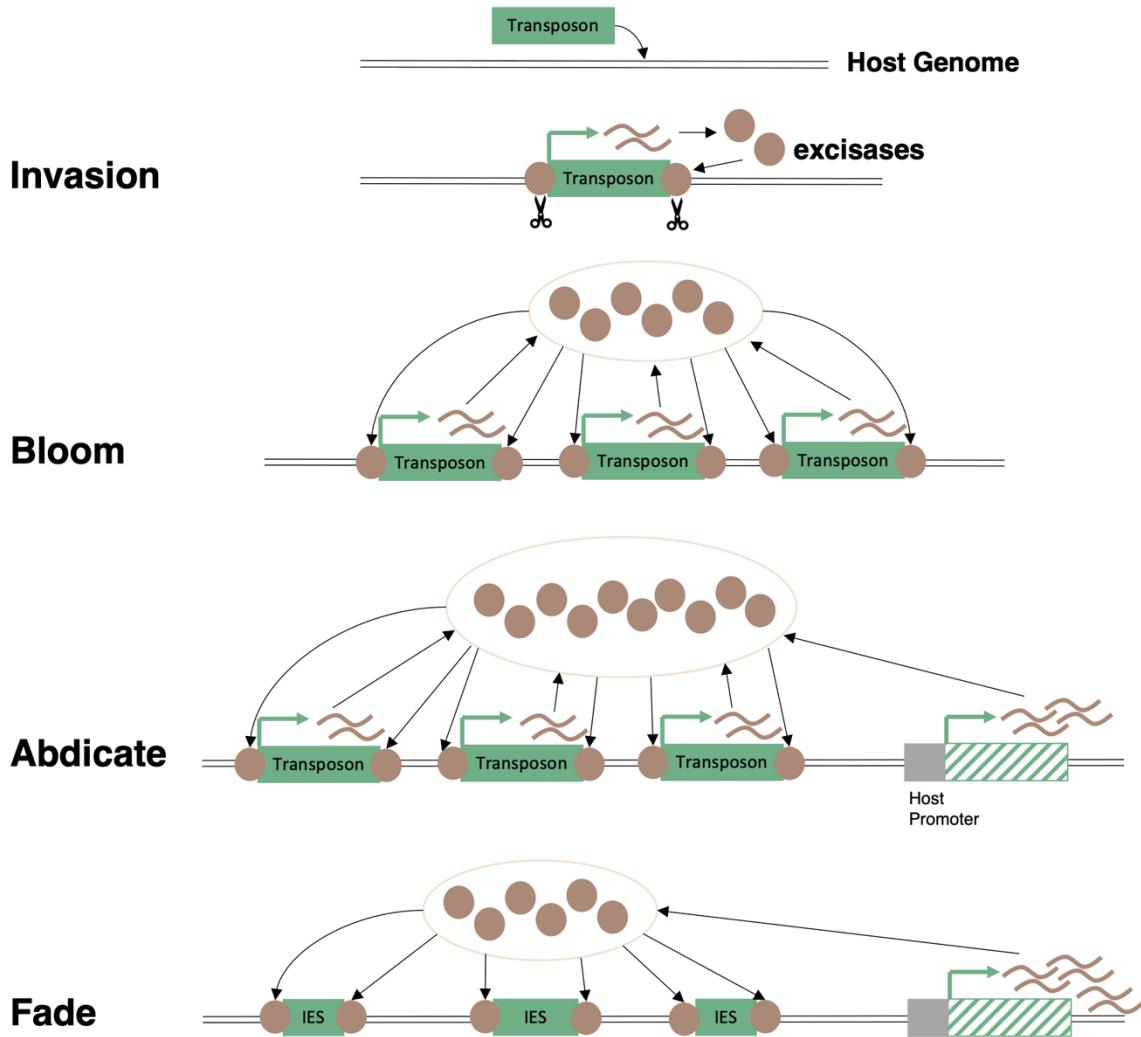


Figure 1. Klobutcher and Herrick's original proposal [4] that most IESs are the decayed remnants of transposons. This model is abbreviated as "IBAF" to describe the gradual conversion of transposons into IESs in four steps: Invasion, Bloom, Abdicate and Fade. Invasion: One or more transposons, perhaps of the IS630-Tc1-mariner superfamily (green box) invaded the host ciliate

germline genome, and many transposons may still be active to this day. Active transposons encode a transposase (brown circles) that permits their excision from or movement within the genome. Bloom: Transposon expansion via replicative transposition across the genome. Their encoded transposase/excisase proteins, capable of acting on each other, become a common good, contributing to the excision of transposons other than one's self. Abdicate: A gene (green hashed box) that encodes the excisase becomes placed under a host promoter for stable and regulated expression, but loss of mobility. In some cases, this gene may also be retained in the macronucleus (e.g., PiggyMac in *Paramecium tetraurelia*). Fade: Mutation accumulation in most transposons gradually converts them into short IESs, non-autonomous elements that retain motifs to permit their recognition and cleavage from the genome by host excisases.

References

1. Wang J, Davis RE. Programmed DNA elimination in multicellular organisms. *Current opinion in genetics & development*. 2014 Aug 1;27:26-34.
2. Vogt A, Goldman AD, Mochizuki K, Landweber LF. Transposon domestication versus mutualism in ciliate genome rearrangements. *PLoS Genet*. 2013 Aug 1;9(8):e1003659.
3. Sellis et al. *This study*.
4. Klobutcher LA, Herrick G. Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog Nucleic Acid Res Mol Biol*. 1997;56: 1–62.
5. Henssen AG, Koche R, Zhuang J, Jiang E, Reed C, Eisenberg A, Still E, MacArthur IC, Rodriguez-Fos E, Gonzalez S, Puiggros M. PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nature genetics*. 2017 Jul;49(7):1005-14.
6. Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. A functional role for transposases in a large eukaryotic genome. *Science*. 2009 May 15;324(5929):935-8.
7. Baudry C, Malinsky S, Restituito M, Kapusta A, Rosa S, Meyer E, Bétermier M. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes & development*. 2009 Nov 1;23(21):2478-83.

8. Cheng CY, Vogt A, Mochizuki K, Yao MC. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in *Tetrahymena thermophila*. *Molecular biology of the cell*. 2010 May 15;21(10):1753-62.
9. Klobutcher LA, Herrick G. Consensus inverted terminal repeat sequence of *Paramecium* IESs: resemblance to termini of Tc1-related and Euplotes Tec transposons. *Nucleic acids research*. 1995 Jun 11;23(11):2006-13.
10. Chen X, Bracht JR, Goldman AD, Dolzhenko E, Clay DM, Swart EC, Perlman DH, Doak TG, Stuart A, Amemiya CT, Sebra RP, Landweber LF. The architecture of a scrambled genome reveals massive levels of genomic rearrangement during development. *Cell*. 2014 Aug 28;158(5):1187-98.
11. Lhuillier-Akakpo M, Frapporti A, Wilkes CD, Matelot M, Vervoort M, Sperling L, Duharcourt S. Local effect of enhancer of zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. *PLoS Genet*. 2014 Sep 25;10(9):e1004665.
12. Miller RV, Neme R, Clay DM, Pathmanathan JS, Lu MW, Yerlici VT, Khurana JS, Landweber LF. Transcribed germline-limited coding sequences in *Oxytricha trifallax*. *G3 (Bethesda)*. 2021 Mar 27:jkab092. doi: 10.1093/g3journal/jkab092.
13. Huff JT, Zilberman D, Roy SW. Mechanism for DNA transposons to generate introns on genomic scales. *Nature*. 2016 Oct;538(7626):533-6.