Smith College

Smith ScholarWorks

Biological Sciences: Faculty Publications

Biological Sciences

2-3-2015

Analyses of Alternatively Processed Genes in Ciliates Provide Insights into the Origins of Scrambled Genomes and may Provide a Mechanism for Speciation

Feng Gao
Smith College

Scott W. Roy
San Francisco State University

Laura A. Katz
Smith College, lkatz@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/bio_facpubs

Part of the Biology Commons

Recommended Citation
Gao, Feng; Roy, Scott W.; and Katz, Laura A., "Analyses of Alternatively Processed Genes in Ciliates Provide Insights into the Origins of Scrambled Genomes and may Provide a Mechanism for Speciation" (2015). Biological Sciences: Faculty Publications, Smith College, Northampton, MA.
https://scholarworks.smith.edu/bio_facpubs/99

This Article has been accepted for inclusion in Biological Sciences: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

RESEARCH ARTICLE

CrossMark

Analyses of Alternatively Processed Genes in Ciliates Provide Insights into the Origins of Scrambled Genomes and May Provide a Mechanism for Speciation

Feng Gao,[a,b] Scott W. Roy,[c] Laura A. Katz[a,d]

Department of Biological Sciences, Smith College, Northampton, Massachusetts, USA[a]; Laboratory of Protozoology, Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, China[b]; Department of Biology, San Francisco State University, San Francisco, California, USA[c]; Program in Organismic and Evolutionary Biology, UMass-Amherst, Amherst, Massachusetts, USA[d]

ABSTRACT **Chromosome rearrangements occur in a variety of eukaryotic life cycles, including during the development of the somatic macronuclear genome in ciliates. Previous work on the phyllopharyngean ciliate *Chilodonella uncinata* revealed that macronuclear β-tubulin and protein kinase gene families share alternatively processed germ line segments nested within divergent regions. To study genome evolution in this ciliate further, we characterized two additional alternatively processed gene families from two cryptic species of the ciliate morphospecies *C. uncinata*: those encoding histidine acid phosphatase protein (*Hap*) and leishmanolysin family protein (*Lei*). Analyses of the macronuclear *Hap* and *Lei* sequences reveal that each gene family consists of three members in the macronucleus that are marked by identical regions nested among highly divergent regions. Investigation of the micronuclear *Hap* sequences revealed a complex pattern in which the three macronuclear sequences are derived either from a single micronuclear region or from a combination of this shared region recombined with additional duplicate micronuclear copies of *Hap*. We propose a model whereby gene scrambling evolves by gene duplication followed by partial and reciprocal degradation of the duplicate sequences. In this model, alternative processing represents an intermediate step in the evolution of scrambled genes. Finally, we speculate on the possible role of genome architecture in speciation in ciliates by describing what might happen if changes in alternatively processed loci occur in subdivided populations.**

IMPORTANCE **Genome rearrangements occur in a variety of eukaryotic cells and serve as an important mechanism for generating genomic diversity. The unusual genome architecture of ciliates with separate germline and somatic nuclei in each cell, provides an ideal system to study further principles of genome evolution. Previous analyses revealed complex forms of chromosome rearrangements, including gene scrambling and alternative processing of germ line chromosomes. Here we describe more complex rearrangements between germ line and somatic chromosomes than previously seen in alternatively processed gene families. Drawing on the present and previous findings, we propose a model in which alternative processing of duplicated micronuclear regions represents an intermediate stage in the evolution of scrambled genes. Under this model, alternative processing may provide insights into a mechanism for speciation in ciliates. Our data on gene scrambling and alternative processing also enhance views on the dynamic nature of genomes across the eukaryotic tree of life.**

Received 16 December 2014 **Accepted** 18 December 2014 **Published** 3 February 2015

Citation Gao F, Roy SW, Katz LA. 2015. Analyses of alternatively processed genes in ciliates provide insights into the origins of scrambled genomes and may provide a mechanism for speciation. mBio 6(1):e01998-14. doi:10.1128/mBio.01998-14.

Editor Patricia J. Johnson, UCLA

Copyright © 2015 Gao et al. This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Laura A. Katz, lkatz@smith.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology.

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

Genomes are incredibly dynamic within diverse lineages across the tree of life (1, 2). Dynamic genomes differ not only in terms of extensive intra- and interspecific variation in genome content and structure but also in genome processing (e.g., DNA elimination and reorganization). Genome rearrangements occur in a variety of eukaryotic cells and serve as an important mechanism for generating genomic diversity. For example, the switching of variant surface glycoprotein (VSG) to generate antigenic variation in *Trypanosoma brucei* occurs in part by DNA rearrangements involving >1,000 VSG genes (3). Similarly, recombination of V(D)J regions generates diversity in immunoglobulins in humans and other vertebrates (4). Moreover, different chromosomal rearrangements of the supergene locus *P*, which contains a cluster of several genes that control different aspects of wing patterning, result in various wing pattern morphs in the polymorphic mimetic butterfly *Heliconius numata* (5). Finally, rearrangements of a single locus underlie the expression of seven mating types in *Tetrahymena thermophila* (6). Here, mating type is determined through a stochastic process in which the macronuclear copy of the mating gene is alternatively assembled from sequences in the micronuclear mating type locus (6).

Although developmentally regulated chromosome rearrangement occurs in a variety of eukaryotes, genome rearrangements may be most pronounced in ciliates. Ciliates are a very diverse

clade of microbial eukaryotes that segregate germ line and somatic functions into two types of nuclei with distinct genome structures: the diploid micronucleus (germ line) and the polyploid macronucleus (soma). Micronuclei and macronuclei differentiate from a genetically novel zygotic nucleus following sexual conjugation. The new zygotic nucleus divides by mitosis. The two descendant nuclei then take on distinct roles, with one developing into a germ line micronucleus and the other into a somatic macronucleus. During development, the macronuclear genome is transformed through a series of chromosomal rearrangements, including fragmentation, DNA elimination, and DNA amplification (7–15).

The types of DNA elimination during macronuclear development are quite diverse, both within a given ciliate species and among different ciliates (9, 12). Precise excision of internal eliminated sequences (IESs) occurs in *Paramecium*, *Oxytricha*, and *Chilodonella*. A more complex form of genome reorganization (termed gene scrambling) is observed in some ciliates, such as *Chilodonella*, *Oxytricha*, and other stichotrichous ciliates: not only must IESs be removed, but also the intervening macronucleus-destined sequences (MDSs) must be reordered. Gene scrambling has been well characterized in genes encoding actin I, telomere end-binding protein subunit α, and DNA polymerase α in spirotrichs (16–19) and actin and β-tubulins in *Chilodonella uncinata* (20).

The mechanism underlying gene scrambling is not well understood, but MDS boundary motifs, macronuclear RNA templates and small RNAs appear to be important. First, splicing appears to involve homologous recombination between pairs of identical short sequence motifs (called pointers) at the 3′ end of one MDS and the 5′ end of the subsequent MDS (15, 21). Second, RNA transcripts from the parental macronucleus have important roles in guiding creation of new macronuclear chromosomes, and small RNAs determine which sequences to retain in the macronucleus (22, 23). These transcripts serve as templates for splicing and also have a role in proofreading of spliced DNAs (24). The importance of the parental macronuclear genome for development of the new somatic genome is underscored by two observations. First, introducing novel chromosomal sequences in the form of new templates into the macronucleus leads to the presence of these novel chromosomal arrangements in the macronucleus in subsequent generations (9, 24, 25). Second, a high frequency of aberrant nanochromosomes appears to be created in the process of macronuclear creation; however, these aberrant nanochromosomes are not found in the mature macronucleus, indicating that they are discarded and/or corrected by a proofreading mechanism (26). Thus, the presence/absence of a sequence in the preceding macronucleus promotes presence/absence in the new macronucleus.

A previous study on the ciliate *Chilodonella uncinata* revealed a highly complex form of chromosome rearrangement, in which some micronuclear segments are used to generate multiple macronuclear sequences (20), a process called alternative processing. For example, the macronuclear β-tubulin genes P1 and P2 are assembled by alternative processing of several micronuclear loci: MIC P1, MIC P2, and MIC SP1 (20). Previous analyses of transcriptome data revealed more than 100 candidate alternatively processed gene families, indicating that alternative processing may be extensive among gene families within *C. uncinata* (27). Alternative processing in the spirotrichous ciliate *Oxytricha trifallax* was subsequently reported (26, 28, 29).

In the present study, we explored two gene families that were previously identified as possibly alternatively processed on the basis of transcriptome data (27, 30): that encoding histidine acid phosphatase family protein (*Hap*) and that encoding leishmanolysin family protein (*Lei*). *Hap* encodes a member of a large functionally diverse group of proteins that play key roles in such varied biological processes as metabolism, development, and intracellular signaling (31). Leishmanolysin was identified as an important virulence factor that was found in the parasite *Leishmania*, where it contributes to a variety of functions allowing host immune evasion (32, 33). The function of these genes in ciliates is as yet unknown. We found that both gene families have three macronuclear copies that are marked by patterns of regions of identity intermingled with divergent regions. We characterized the micronuclear *Hap* sequences, which revealed a complex pattern of alternative processing to produce the three macronuclear sequences. We propose a model in which alternative processing of duplicated micronuclear sequences represents an intermediate stage in the evolution of scrambled genes. Finally, we speculate on the possibility that alternative processing can contribute to high rates of speciation in ciliates.

RESULTS

***Hap* and *Lei* have multiple macronuclear sequences marked by alternating regions of nucleotide divergence and identity.** We identified three macronuclear sequences for both *Hap* (Acc. no. KJ000273-KJ000278) and *Lei* genes (see Table S1 in the supplemental material). For each gene family, comparison between different macronuclear sequences revealed a combination of identical and diverged sequences (Fig. 1). For the *Hap* genes, comparison of two macronuclear sequences (termed MAC P1 and MAC P2) showed three identical regions (indicated by a π value of 0) (Fig. 1) alternating with more divergent regions. Comparison between MAC P1 and the third sequence (MAC P3) also showed three identical regions, but these regions were in different locations (Fig. 1A). For the *Lei* gene, MAC P1 and MAC P2 share four identical regions alternating with more divergent regions, while MAC P1 and MAC P3 share five identical regions with some varying boundaries as compared to MAC P2 (Fig. 1B).

We sought to time the duplication events that led to the different macronuclear sequences relative to the divergence of the Pol and USA strains (Fig. 1C and D). We found that there was more nucleotide divergence between different macronuclear sequences than there was between the two strains' copies of the same macronuclear sequence (see Fig. S1 and S2 in the supplemental material), indicating that for both gene families the duplication events predate the divergence between the strains.

Macronuclear *Hap* sequences are assembled from alternatively processed MDSs from a single micronuclear locus containing duplicated *Hap* genes. To assess the processing between the germ line micronucleus and somatic macronucleus, we used traditional PCR to characterize the micronuclear sequences of *Hap* genes for the Pol strain (ca. 3.6 kb in length) (Fig. 2), using a MAC P2-specific forward primer and a shared reverse primer. This revealed a single micronuclear locus containing three duplicated *Hap* gene sequences. Based on the comparison with the macronuclear sequences, we term these P2 specific (blue in Fig. 2), P3 specific (purple in Fig. 2), and shared (black in Fig. 2).

Comparison of micronuclear and macronuclear sequences of *Hap* gene revealed a complex pattern of alternative processing and gene scrambling. Pointer sequences ranging from 4 to 8 bp were

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

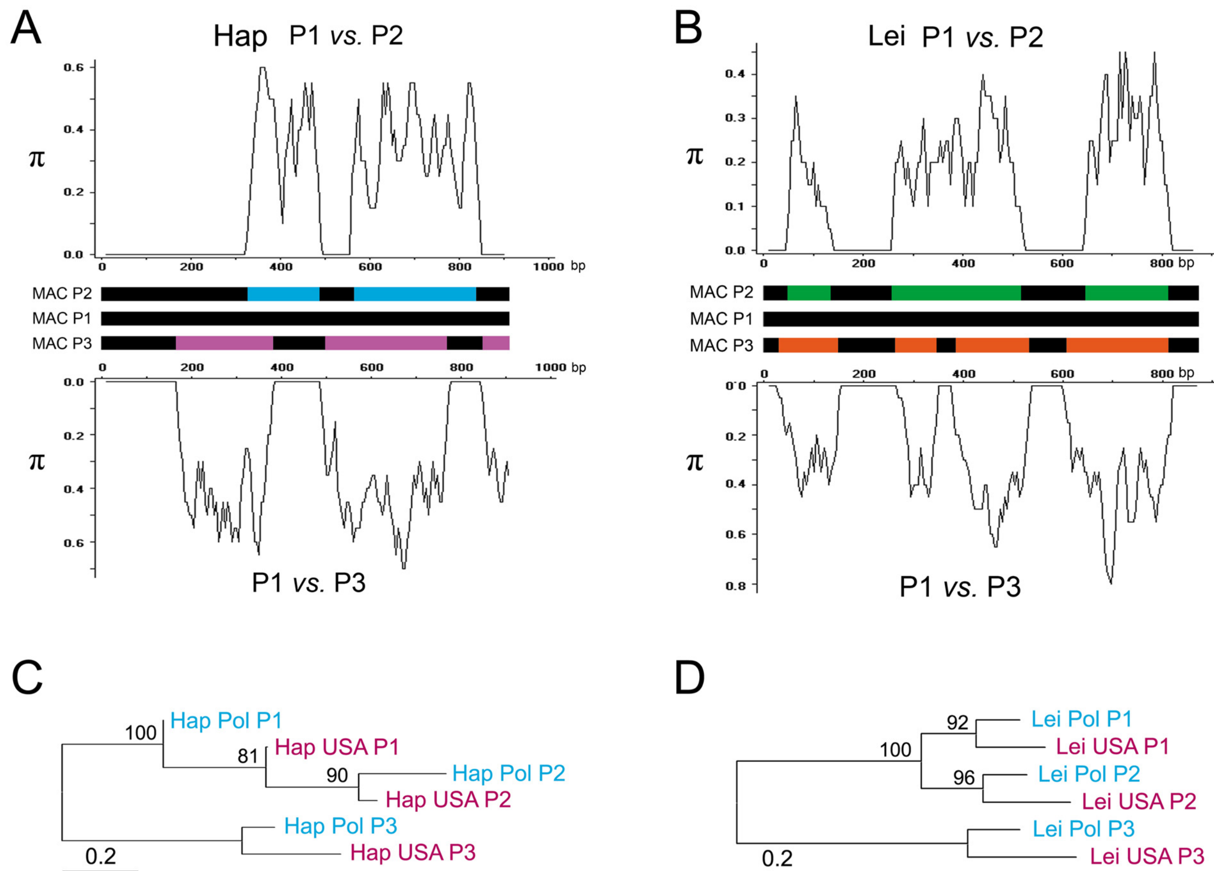

FIG 1 Sequence comparisons among gene family members of *Hap* (A) and *Lei* (B) and genealogies of gene family members from Pol and USA strains for *Hap* (C) and *Lei* (D). (A and B) Graphs are sliding-window analyses of pairwise divergence (π) calculated using DnaSP (59). The top comparison is of macronuclear (MAC) P1 and P2, and the bottom comparison is of MAC P1 and P3. Regions in black at identical positions correspond to shared sequences. (C and D) Topologies were estimated by PhyML (58) as implemented in SeaView (57). Numbers at nodes represent the bootstrap values of maximum likelihood analysis out of 1,000 replicates. Scale bars show substitutions per site.

found at the boundaries of MDSs and IESs (Table 1), supporting the alternative processing of *Hap* gene. The MAC P1 has the simplest pattern and is made up of four MDSs that are located sequentially in a single micronuclear copy (shared) and are separated by three rapidly evolving IESs (see Fig. S3 in the supplemental material). In contrast, both MAC P2 and P3 are scrambled in the micronucleus and are generated by combination of interdigitated sequences from the single micronuclear region. MAC P2 contains sequence from both the shared and P2-specific copies: interestingly, two of the shared MDSs found in MAC P1 are also found in MAC P2 (first and fourth), whereas the other two (second and third) undergo alternative processing with P2-specific sequences. MAC P3 is generated from the shared sequence and yet another sequence (P3 specific) and is more complex yet: (i) no full MDS is shared with either MAC P1 or MAC P2, with only partial shared MDSs being present, and (ii) three of the five P3-specific MDSs are present in the opposite orientation (i.e., on the reverse strand).

We used information on the structure of *Hap* in the Pol micronuclear sequence to design USA-specific primers for characterizing micronuclear copies in this strain. The organization of the USA micronuclear sequence shows a structure similar to that of Pol, except that the fourth MDS of MAC P3 is divided into two MDSs by a 35-bp IES (see Fig. S4 in the supplemental material), implying that this IES was either gained in the USA strain or lost in the Pol strain. The pointer sequences in the USA strain range from 2 to 8 bp, with some MDS-IES junction shifts compared to the Pol strain (Table 1; also, see Table S2 and Fig. S3 in the supplemental material).

Using a similar approach, we were not able to characterize the micronuclear copy(ies) corresponding to the *Lei* gene. Walking PCR for the *Lei* gene yielded sequences that are identical to the macronuclear sequences, indicating that the primers are interrupted by IESs in the micronucleus (we had macronuclear contamination in our micronuclear preps), the gene is highly fragmented or scrambled, and/or the region we characterized does not contain IESs in the micronucleus.

DISCUSSION

This study of two gene families in two strains of the ciliate morphospecies *C. uncinata* leads to three main insights: (i) macronuclear *Hap* and *Lei* gene family members show a combination of regions of identity and highly divergent regions that are suggestive of alternative processing; (ii) the three macronuclear *Hap* members are generated by alternative processing of a single micronuclear region that contains duplicated and decayed *Hap* genes; and (iii) alternative processing is more complex than previously be-

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

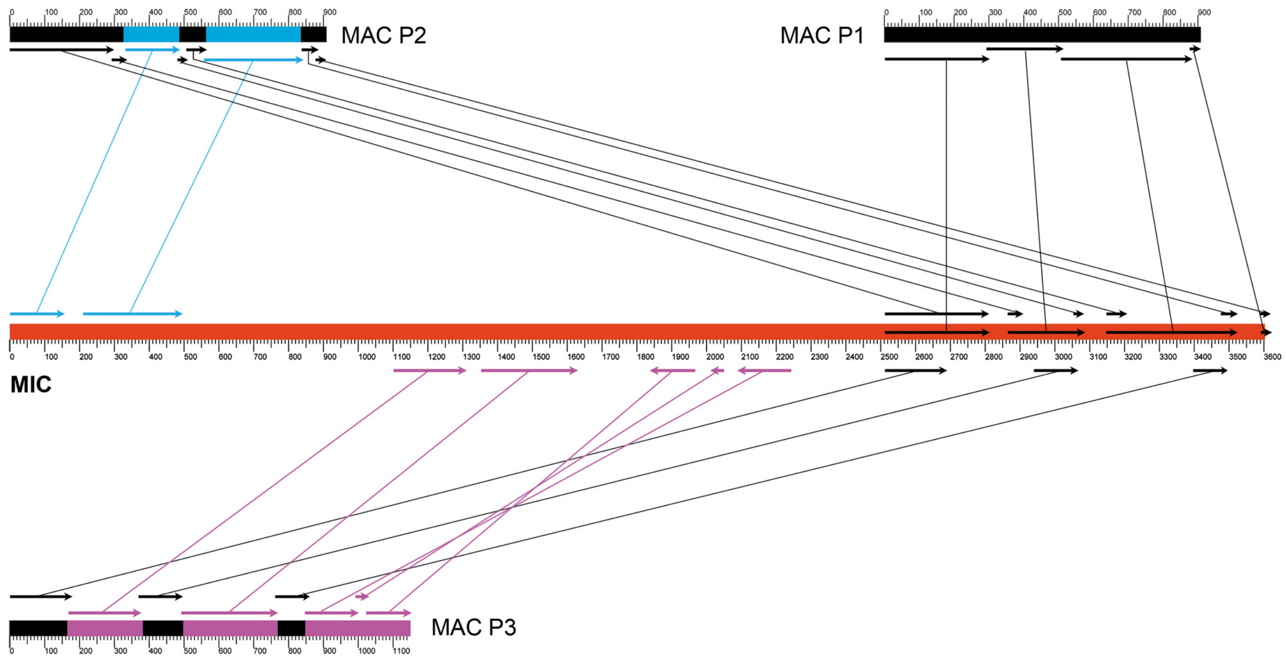

FIG 2 Schematic maps of the somatic and corresponding germ line sequences of *Hap*. The three diverse *Hap* genes are alternatively spliced together from a single micronuclear locus. Colors correspond to macronuclear loci in Fig. 1. MDSs for each macronuclear locus are marked with arrows, and their corresponding sites in the micronuclear locus are also indicated with the same arrows linked with lines. The directions of the arrows indicate the sequence directions in the macronuclear locus.

lieved, as the sharing of micronuclear regions can vary in generating macronuclear products. Drawing on these findings, we hypothesize that alternative processing of duplicated micronuclear sequences may be an intermediate step in the evolution of gene scrambling and may play a role in speciation in ciliates.

Complex processing of Hap and Lei gene family members. Sliding-window analyses of divergence among *Hap* and *Lei* gene family members revealed stretches of identity nested within highly divergent regions. The identical regions are flanked by highly divergent stretches where pairwise differences (π) can be up to 0.60 (*Hap*) or 0.80 (*Lei*), values that are likely underestimates due to multiple hits/saturation (Fig. 1). Previous studies of β-tubulin and protein kinase domain-containing gene families in *C. uncinata* showed similar patterns, with islands of identity within highly divergent macronuclear gene family members (20, 27).

TABLE 1 Characteristics of pointers of the *Hap* gene family from strain Pol of *C. uncinata*

Pointer	Sequence	Start and end	Haplotype(s)
1	TGACAAC	2786-2792/2846-2852	P1/P2
2	CAGAAAC	3059-3065/3130-3136	P1/P2
3	TACCCAAG	3499-3506/3572-3579	P1/P2
4	GATCTTC	133-139/3035-3041	P2
5	AAGATGGA	3182-3189/191-198	P2
6	TTTGCTT	471-476/3458-3464	P2
7	GGTTGCA	2663-2669/1083-1089	P3
8	AGAA	1286-1289/2926-2929	P3
9	GAAACC	3045-3050/1333-1338	P3
10	TCACT	1607-1611/3384-3388	P3
11	TTCG	3466-3469/2223-2220	P3
12	ATTCAAA	2078-2072/2031-2025	P3
13	CCAGAAAG	2002-1995/1949-1942	P3

Analyses of the transcriptome data from *C. uncinata* Pol strain revealed more than 100 gene families that also show similar patterns, suggesting that alternative processing could be common (27).

Three macronuclear *Hap* members are generated by alternative processing of a single micronuclear region that contains duplicated and decayed *Hap* genes. Several lines of evidence support this conclusion: (i) the sharing of identical regions among macronuclear sequences; (ii) the recovery of only one micronuclear sequence containing regions identical to all regions of the macronuclear *Hap* genes; (iii) the presence of pointer sequences at appropriate locations between the micronuclear regions that need to be joined to form macronuclear sequences; and (iv) the fact that the two strains of *C. uncinata* show the same alternative processing patterns. Based on the pattern observed here, we hypothesize that the original *Hap* gene duplicated twice, followed by decay of some of the coding regions and subsequent replacement by recombination of intact homologous regions during macronuclear development (see cartoons in Fig. 3 and 4). The processing of the *Hap* micronuclear locus leads to the three alternatively processed macronuclear sequences in which identical macronuclear regions come from shared micronuclear regions.

Alternative processing is more complex than previously believed, as the sharing of micronuclear regions can vary in generating macronuclear products. Our *Hap* MIC locus adds to the list of alternatively processed genes in *C. uncinata*, which includes β-tubulin gene family (20) and a protein kinase domain containing protein (PKc) gene family (27). Previous analyses of the β-tubulin gene family showed that two members, MAC P1 and MAC P2, are generated using the same alternatively processed MIC SP1 regions (20). The analyses of the PKc gene family also

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

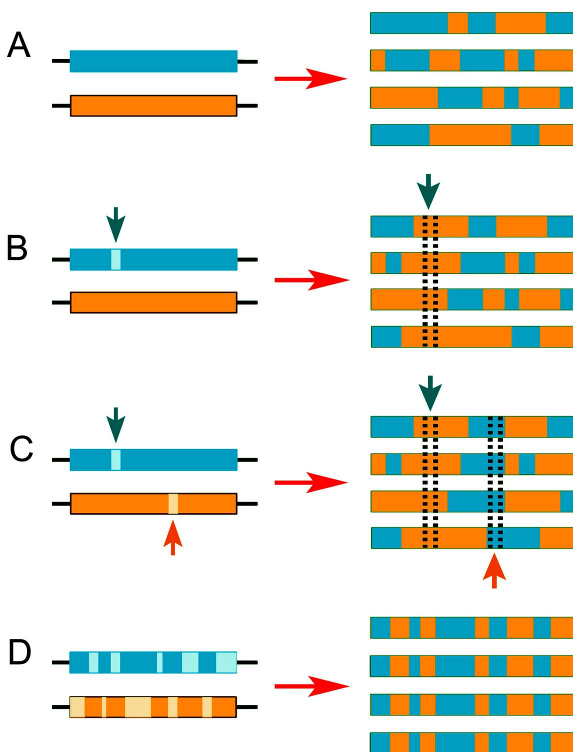

FIG 3 Model for the origins of scrambled micronuclear genes. (A) Following an initial micronuclear duplication, DNA splicing could use a variety of sequences as pointers, leading to identical spliced molecules deriving from various combinations of the two micronuclear duplicates. Blue and orange boxes on the left indicate the two duplicates. Mixed blue/orange boxes on the right indicate various spliced DNAs generated by using a variety of spliced sites. (B) Due to RNA template proofreading, a mutation in one duplicate (arrow) leads to the mutated region becoming restricted to the micronucleus (light color), leading to constitutive usage of sequence from the nonmutated duplicate at that site (all spliced DNAs use orange in the mutated region). (C) A second mutation in the other duplicate leads to constitutive usage of sequence from the other (blue) duplicate at a second site. (D) Accumulation of mutations in the duplicates leads to a scrambled gene.

showed that the shared identical regions are processed using the same MIC regions (27). The present study of *Hap* gene revealed a different pattern in that *Hap* macronuclear gene family members MAC P2 and MAC P3 are generated using different alternatively processed (i.e., shared) MIC P1 regions. This complex pattern of sharing indicates that there must be a controlled and precise rearrangement mechanism to guide the macronucleus-destined sequences into the correct linear order and orientation, as has been found in other ciliates (24, 34).

On the origins and consequences of genome scrambling. Our analyses of patterns among *Hap* and *Lei* gene family members leads to a model on the evolution of gene scrambling whereby duplication of micronuclear regions is followed by a transient period of alternative processing, which is later resolved as gene scrambling (Fig. 3 and 4). The cases of alternative processing reported here and elsewhere (20, 27, 29, 35) share the observation that macronuclear gene family members are generated by recombination between duplicated micronuclear sequences. Such a system may arise through constructive neutral evolution (36, 37), though we recognize the challenges of disentangling the evolutionary forces (e.g., genetic drift and natural selection) at play in the origin of this system (38–42). Hence, we focus on the role of gene duplication in enabling the evolution of alternative processing and, ultimately, gene scrambling.

Following duplication of micronuclear regions, the existence of long stretches of identical sequences provides redundancy in the pointer pairs that direct rearrangements during macronuclear development (Fig. 3A). Alternative usage of various combinations of these nascent pointers could lead to production of macronuclear sequences from diverse combinations of the micronuclear duplicates. Over time, the redundancy in pointer sequences and duplicated coding regions could allow an inactivating mutation in a region of one duplicate to become fixed with no negative fitness effect (i.e., decay) (Fig. 3B). Such mutated regions could be excluded from the macronucleus by scanning during macronuclear development, which ensures that sequences in the newly formed macronucleus reflect those in the previous macronucleus (34, 43, 44); thus, a mutated region of one duplicate could become restricted to the micronucleus. A similar inactivating mutation in the other duplicate could then lead to restriction of that region to the micronucleus, at which point all functional macronuclear regions would be assembled from multiple micronuclear sequences, constituting a newly scrambled gene (Fig. 3C). Further mutations could eventually lead to a pattern of nearly complete reciprocal degradation, with the pointer sequences representing the only remaining regions of sequence redundancy (Fig. 3D). For instance an inactivating mutation within remaining paralogous regions in the black duplicate on the right of Fig. 2 would abolish MAC P1, in which case all remaining macronuclear sequences would be the result of scrambling.

In this scenario, alternative processing could represent a transient stage on the road to full gene scrambling (Fig. 4). This model mirrors classic duplicate gene pseudogenization (45, 46), in which one of a pair of duplicate genes degrades by mutation, though in the case of alternative processing in ciliates, different regions of the duplicates could reciprocally degrade. Another possibility is that some parts of the duplicated gene could be retained in duplicate due to evolution of new functions (neofunctionalization) or partitioning of ancestral functions between the two regions (subfunctionalization) (45, 46). In this case, alternative processing could be evolutionarily stable, with further degradation opposed by purifying selection. In the examples reported here, the persistence of some gene regions in duplicates despite significant sequence divergence suggests that purifying selection is acting to oppose inactivating regions, and thus that they are not simply functionally redundant.

We further speculate that our model of differential degradation of duplicates leading to gene scrambling may provide a mechanism for speciation in ciliates (Fig. 4). If the degradation of regions occurs multiple times in subdivided populations, then this could create a barrier to successful reproduction between resulting strains as offspring between such crosses would not be capable of generating functional gene family members (Fig. 4D and E). In other words, differing patterns of alternative processing of scrambled "options" in subdivided populations would lead to incompatibility in subsequent matings between members, resulting in incipient species. In this scenario, it is possible that reproductive barriers may occur more rapidly than predicted by the accumulation of point mutations, which would explain the disconnect between the rates of morphological and molecular evolution that underlie ciliate species (47–55).

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

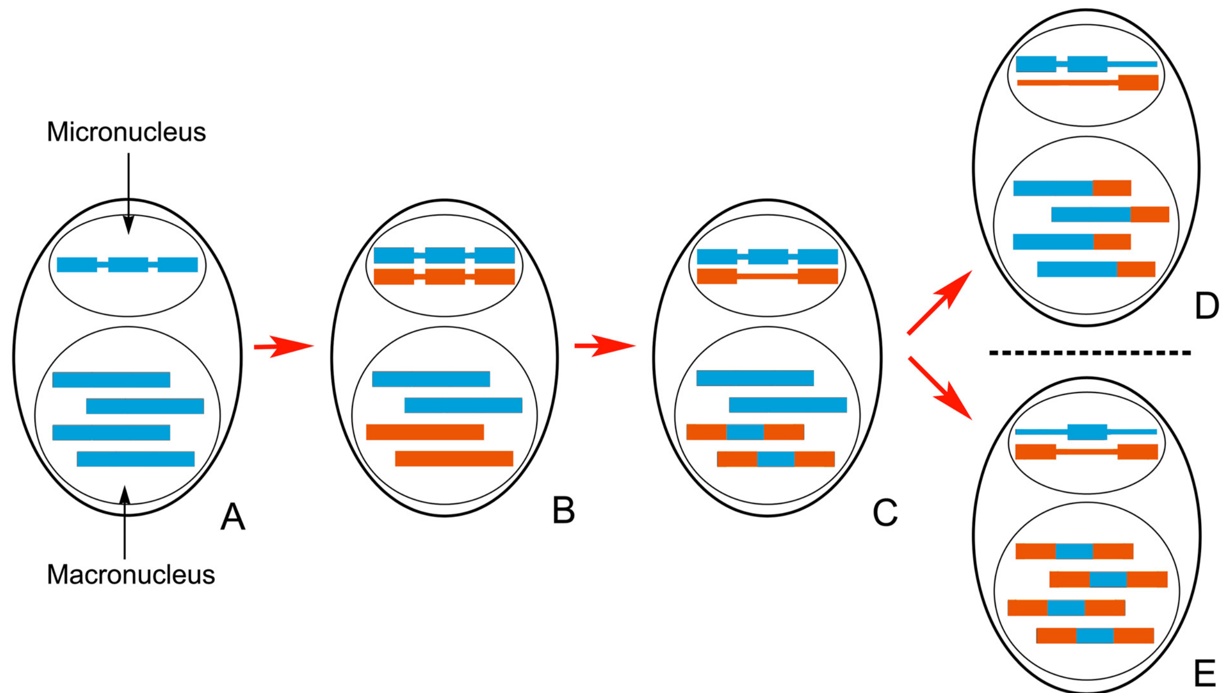

FIG 4 Genome architecture drives evolution in ciliates, resulting in gene scrambling and perhaps even speciation. (A) Each ciliate contains a germ line micronucleus with a canonical eukaryotic genome and a somatic macronucleus represented by a large polyploid nucleus. A single gene with IESs is shown in the micronucleus, and multiple copies of the processed gene are present in the macronucleus. (B) The gene duplicates in the micronucleus followed by divergence, and both copies are processed during macronuclear development. (C) A coding region in the micronucleus degrades and is replaced by recombination of homologous regions from the intact copy, leading to alternatively processed macronuclear chromosomes. Further decay can happen, so that no duplicate homologous regions remain and only one haplotype will be generated during macronuclear development, resulting in gene scrambling. (D and E) Populations that become fixed for different scrambling "options" may become incompatible (i.e., incipient species).

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

MATERIALS AND METHODS

Ciliate culturing and DNA extraction. We maintained two previously characterized cryptic species (referred to here as strains, as they have not been described formally) of the ciliate morphospecies *C. uncinata*, Pol (ATCC PRA-256) and USA-Sc2, following protocols described by Katz et al. (48). To isolate total DNA, cultures were treated overnight with penicillin-streptomycin-amphotericin B (17-745 H; Lonza, Allendale, NJ), and cells were pelleted by spinning at 5,000 rpm for 20 min. Genomic DNA was extracted using phenol-chloroform following standard protocols (56). Micronuclear DNA was isolated according to Katz and Kovner (20). Briefly, micronuclear DNA was gel isolated by gel electrophoresis using low-melting-point UltraClean agarose (15005-50; Mobio, Carlsbad, CA) after digestion with Bal 31 nuclease (M02135; New England Biolabs, Ispwich, MA) to enrich micronuclear DNA. Gel-isolated micronuclear DNA was purified using β-agarase (M03925; New England Biolabs).

Traditional PCR and cloning. We chose two gene families, encoding histidine acid phosphatase family protein (*Hap*) and leishmanolysin family protein (*Lei*), for which multiple RNA transcripts sharing some sequences are present in the assembled *C. uncinata* transcriptome. Primers for both *Hap* and *Lei* genes were designed from these shared regions. The primers were then used on two *C. uncinata* strains, Pol and USA, to amplify the macronuclear sequences. Haplotype-specific primers were designed to amplify the micronuclear sequences. PCR was performed using Phusion Hot Start high-fidelity DNA polymerase (F 540 L; Finnzymes, Finland). Amplified products were cloned using Zero Blunt TOPO kits (K2800; Invitrogen, CA), and screened using the polymerase TaqGold (Applied Biosystems, CA).

Genome walking PCR and cloning. We used Seegene's DNA walking SpeedUp kit (K1052; Seegene, Rockville, MD) to amplify additional regions of *Lei*. PCR amplification was performed following Seegene kit protocol using kit primers and gene-specific primers designed for this study. Genome walking PCR products were cloned using TA TOPO cloning kits (45-0641; Invitrogen) and screened using the polymerase TaqGold (Applied Biosystems, CA).

Sequencing and data analysis. Sequences were generated using the BigDye terminator v3.1 cycle sequencing kit (no. 4337455) from PE Applied Biosystems (Wellesley, MA). Reaction products were cleaned using gel filtration columns (no. 42453) from Edge Biosystems (Gaithersburg, MD) and analyzed on a PerkinElmer ABI-3100 automated sequencer at the Center for Molecular Biology (Smith College, Northampton, MA). Contigs were assembled in SeqMan (DNASTAR), and all polymorphisms were confirmed by eye. SeaView v. 4.2.4 (57) and MegAlign (DNASTAR) were used to create alignments. Genealogies based on nucleotide alignments were estimated using PhyML (58) as implemented in SeaView v. 4.2.4 with the model GTR+gamma. DnaSP (59) was used to perform sliding-window analysis to calculate average pairwise differences (π). Sliding-window analyses were performed with a 20-bp window and a 5-bp step.

Nucleotide sequence accession numbers. The macronuclear sequences for *Lei* genes have been deposited in GenBank database under accession no. KJ000279 to KJ000284. The micronuclear sequence of *Hap* genes for the Pol strain has been deposited under accession no. KJ626297. The micronuclear sequence of *Hap* genes for the USA strain has been deposited under accession no. KJ626298.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01998-14/-/DCSupplemental.

Figure S1, PDF file, 0.5 MB.

Figure S2, PDF file, 0.1 MB.
Figure S3, PDF file, 0.3 MB.
Figure S4, PDF file, 1.2 MB.
Table S1, PDF file, 0.04 MB.
Table S2, PDF file, 0.05 MB.

ACKNOWLEDGMENT

This work was supported by the AREA award from the National Institutes of Health (1R15GM097722) to L.A.K.

REFERENCES

1. **Parfrey LW, Lahr DJ, Katz LA.** 2008. The dynamic nature of eukaryotic genomes. Mol Biol Evol **25:**787–794. http://dx.doi.org/10.1093/molbev/msn032.
2. **Zufall RA, Robinson T, Katz LA.** 2005. Evolution of developmentally regulated genome rearrangements in eukaryotes. J Exp Zool B Mol Dev Evol **304:**448–455. http://dx.doi.org/10.1002/jez.b.21056.
3. **Stockdale C, Swiderski MR, Barry JD, McCulloch R.** 2008. Antigenic variation in *Trypanosoma brucei*: joining the DOTs. PLoS Biol **6:**e185. http://dx.doi.org/10.1371/journal.pbio.0060185.
4. **Nemazee D.** 2006. Receptor editing in lymphocyte development and central tolerance. Nat Rev Immunol **6:**728–740. http://dx.doi.org/10.1038/nri1939.
5. **Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M, Baxter SW, Ferguson L, Wilkinson PA, Salazar C, Davidson C, Clark R, Quail MA, Beasley H, Glithero R, Lloyd C, Sims S, Jones MC, Rogers J, Jiggins CD, ffrench-Constant RH.** 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature **477:**203–206. http://dx.doi.org/10.1038/nature10341.
6. **Cervantes MD, Hamilton EP, Xiong J, Lawson MJ, Yuan D, Hadjithomas M, Miao W, Orias E.** 2013. Selecting one of several mating types through gene segment joining and deletion in *Tetrahymena thermophila*. PLoS Biol **11:**e1001518. http://dx.doi.org/10.1371/journal.pbio.1001518.
7. **Heyse G, Jönsson F, Chang WJ, Lipps HJ.** 2010. RNA-dependent control of gene amplification. Proc Natl Acad Sci U S A **107:**22134–22139. http://dx.doi.org/10.1073/pnas.1009284107.
8. **Katz LA, Lasek-Nesselquist E, Snoeyenbos-West OL.** 2003. Structure of the micronuclear alpha-tubulin gene in the phyllopharyngean ciliate *Chilodonella uncinata*: implications for the evolution of chromosomal processing. Gene **315:**15–19. http://dx.doi.org/10.1016/j.gene.2003.08.003.
9. **Nowacki M, Shetty K, Landweber LF.** 2011. RNA-mediated epigenetic programming of genome rearrangements. Annu Rev Genomics Hum Genet **12:**367–389. http://dx.doi.org/10.1146/annurev-genom-082410-101420.
10. **Goldman AD, Landweber LF.** 2012. *Oxytricha* as a modern analog of ancient genome evolution. Trends Genet **28:**382–388. http://dx.doi.org/10.1016/j.tig.2012.03.010.
11. **Katz LA.** 2001. Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. Int J Syst Evol Microbiol **51:**1587–1592.
12. **Chalker DL, Yao MC.** 2011. DNA elimination in ciliates: transposon domestication and genome surveillance. Annu Rev Genet **45:**227–246. http://dx.doi.org/10.1146/annurev-genet-110410-132432.
13. **Riley JL, Katz LA.** 2001. Widespread distribution of extensive chromosomal fragmentation in ciliates. Mol Biol Evol **18:**1372–1377. http://dx.doi.org/10.1093/oxfordjournals.molbev.a003921.
14. **Chalker DL.** 2008. Dynamic nuclear reorganization during genome remodeling of *Tetrahymena*. Biochim Biophys Acta **1783:**2130–2136. http://dx.doi.org/10.1016/j.bbamcr.2008.07.012.
15. **Prescott DM.** 1994. The DNA of ciliated protozoa. Microbiol Rev **58:**233–267.
16. **Landweber LF, Kuo TC, Curtis EA.** 2000. Evolution and assembly of an extremely scrambled gene. Proc Natl Acad Sci U S A **97:**3298–3303. http://dx.doi.org/10.1073/pnas.040574697.
17. **Mitcham JL, Lynn AJ, Prescott DM.** 1992. Analysis of a scrambled gene: the gene encoding alpha-telomere-binding protein in *Oxytricha nova*. Genes Dev **6:**788–800. http://dx.doi.org/10.1101/gad.6.5.788.
18. **Chang WJ, Kuo S, Landweber LF.** 2006. A new scrambled gene in the ciliate *Uroleptus*. Gene **368:**72–77. http://dx.doi.org/10.1016/j.gene.2005.10.008.
19. **Prescott DM, Greslin AF.** 1992. Scrambled actin I gene in the micronucleus of *Oxytricha nova*. Dev Genet **13:**66–74. http://dx.doi.org/10.1002/dvg.1020130111.
20. **Katz LA, Kovner AM.** 2010. Alternative processing of scrambled genes generates protein diversity in the ciliate *Chilodonella uncinata*. J Exp Zool B Mol Dev Evol **314:**480–488. http://dx.doi.org/10.1002/jez.b.21354.
21. **DuBois ML, Prescott DM.** 1997. Volatility of internal eliminated segments in germ line genes of hypotrichous ciliates. Mol Cell Biol **17:**326–337.
22. **Fang W, Wang X, Bracht JR, Nowacki M, Landweber LF.** 2012. Piwi-interacting RNAs protect DNA against loss during *Oxytricha* genome rearrangement. Cell **151:**1243–1255. http://dx.doi.org/10.1016/j.cell.2012.10.045.
23. **Zahler AM, Neeb ZT, Lin A, Katzman S.** 2012. Mating of the stichotrichous ciliate *Oxytricha trifallax* induces production of a class of 27 nt small RNAs derived from the parental macronucleus. PLoS One **7:**e42371. http://dx.doi.org/10.1371/journal.pone.0042371.
24. **Nowacki M, Vijayan V, Zhou Y, Schotanus K, Doak TG, Landweber LF.** 2008. RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature **451:**153–158. http://dx.doi.org/10.1038/nature06452.
25. **Nowacki M, Landweber LF.** 2009. Epigenetic inheritance in ciliates. Curr Opin Microbiol **12:**638–643. http://dx.doi.org/10.1016/j.mib.2009.09.012.
26. **Möllenbeck M, Zhou Y, Cavalcanti AR, Jönsson F, Higgins BP, Chang WJ, Juranek S, Doak TG, Rozenberg G, Lipps HJ, Landweber LF.** 2008. The pathway to detangle a scrambled gene. PLoS One **3:**e2330. http://dx.doi.org/10.1371/journal.pone.0002330.
27. **Gao F, Song W, Katz LA.** 2014. Genome structure drives patterns of gene family evolution in ciliates, a case study using *Chilodonella uncinata* (Protista, Ciliophora, Phyllopharyngea). Evolution **68:**2287–2295. http://dx.doi.org/10.1111/evo.12430.
28. **Zhou Y, Wubneh H, Schwarz C, Landweber LF.** 2011. A chimeric chromosome in the ciliate *Oxytricha* resulting from duplication. J Mol Evol **73:**70–73. http://dx.doi.org/10.1007/s00239-011-9464-1.
29. **Swart EC, Bracht JR, Magrini V, Minx P, Chen X, Zhou Y, Khurana JS, Goldman AD, Nowacki M, Schotanus K, Jung S, Fulton RS, Ly A, McGrath S, Haub K, Wiggins JL, Storton D, Matese JC, Parsons L, Chang WJ, Bowen MS, Stover NA, Jones TA, Eddy SR, Herrick GA, Doak TG, Wilson RK, Mardis ER, Landweber LF.** 2013. The *Oxytricha trifallax* macronuclear genome: a complex eukaryotic genome with 16,000 tiny chromosomes. PLoS Biol **11:**e1001473. http://dx.doi.org/10.1371/journal.pbio.1001473.
30. **Grant JR, Lahr DJG, Rey FE, Burleigh JG, Gordon JI, Knight R, Molestina RE, Katz LA.** 2012. Gene discovery from a pilot study of the transcriptomes from three diverse microbial eukaryotes: *Corallomyxa tenera*, *Chilodonella uncinata*, and *Subulatomonas tetraspora*. Protist Genomics **1:**3–18. http://dx.doi.org/10.2478/prge-2012-0002.
31. **Rigden DJ.** 2008. The histidine phosphatase superfamily: structure and function. Biochem J **409:**333–348. http://dx.doi.org/10.1042/BJ20071097.
32. **Etges R, Bouvier J, Bordier C.** 1986. The major surface protein of *Leishmania promastigotes* is a protease. J Biol Chem **261:**9098–9101.
33. **Bouvier J, Etges RJ, Bordier C.** 1985. Identification and purification of membrane and soluble forms of the major surface protein of *Leishmania promastigotes*. J Biol Chem **260:**15504–15509.
34. **Bracht JR, Fang W, Goldman AD, Dolzhenko E, Stein EM, Landweber LF.** 2013. Genomes on the edge: programmed genome instability in ciliates. Cell **152:**406–416. http://dx.doi.org/10.1016/j.cell.2013.01.005.
35. **Ardell DH, Lozupone CA, Landweber LF.** 2003. Polymorphism, recombination and alternative unscrambling in the DNA polymerase alpha gene of the ciliate *stylonychia lemnae* (Alveolata; class Spirotrichea). Genetics **165:**1761–1777.
36. **Stoltzfus A.** 1999. On the possibility of constructive neutral evolution. J Mol Evol **49:**169–181. http://dx.doi.org/10.1007/PL00006540.
37. **Stoltzfus A.** 2012. Constructive neutral evolution: exploring evolutionary theory's curious disconnect. Biol Direct **7:**35. http://dx.doi.org/10.1186/1745-6150-7-35.
38. **Doolittle WF, Lukes J, Archibald JM, Keeling PJ, Gray MW.** 2011. Comment on "Does constructive neutral evolution play an important role in the origin of cellular complexity?". Bioessays **33:**427–429.
39. **Keeling PJ, Leander BS, Lukes J.** 2010. Constructive neutral evolution cannot explain current kinetoplastid panediting patterns reply. Proc Natl Acad Sci U S A **107:**E26. http://dx.doi.org/10.1073/pnas.0911933107.
40. **Lukeš J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW.** 2011. How

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest

a neutral evolutionary ratchet can build cellular complexity. IUBMB Life **63:**528–537. http://dx.doi.org/10.1002/iub.489.

41. **Speijer D.** 2010. Constructive neutral evolution cannot explain current kinetoplastid panediting patterns. Proc Natl Acad Sci U S A **107:**E25. http://dx.doi.org/10.1073/pnas.0909867107.
42. **Speijer D.** 2011. Does constructive neutral evolution play an important role in the origin of cellular complexity? Making sense of the origins and uses of biological complexity. Bioessays **33:**344–349. http://dx.doi.org/10.1002/bies.201100010.
43. **Chalker DL, Meyer E, Mochizuki K.** 2013. Epigenetics of ciliates. Cold Spring Harb Perspect Biol **5:**a017764. http://dx.doi.org/10.1101/cshperspect.a017764.
44. **Fuhrmann G, Swart E, Nowacki M, Lipps HJ.** 2013. RNA-dependent genome processing during nuclear differentiation: the model systems of stichotrichous ciliates. Epigenomics **5:**229–236. http://dx.doi.org/10.2217/epi.13.15.
45. **Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J.** 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151:**1531–1545.
46. **Lynch M, Conery JS.** 2000. The evolutionary fate and consequences of duplicate genes. Science **290:**1151–1155. http://dx.doi.org/10.1126/science.290.5494.1151.
47. **Gao F, Katz LA, Song W.** 2013. Multigene-based analyses on evolutionary phylogeny of two controversial ciliate orders: Pleuronematida and Loxocephalidae (Protista, Ciliophora, Oligohymenophorea). Mol Phylogenet Evol **68:**55–63. http://dx.doi.org/10.1016/j.ympev.2013.03.018.
48. **Katz LA, DeBerardinis J, Hall MS, Kovner AM, Dunthorn M, Muse SV.** 2011. Heterogeneous rates of molecular evolution among cryptic species of the ciliate morphospecies *Chilodonella uncinata*. J Mol Evol **73:**266–272. http://dx.doi.org/10.1007/s00239-011-9468-x.
49. **Hall MS, Katz LA.** 2011. On the nature of species: insights from *Paramecium* and other ciliates. Genetica **139:**677–684. http://dx.doi.org/10.1007/s10709-011-9571-3.
50. **Simon EM, Nanney DL, Doerder FP.** 2008. The "*Tetrahymena pyriformis*" complex of cryptic species. Biodivers Conserv **17:**365–380. http://dx.doi.org/10.1007/s10531-007-9255-6.
51. **McManus GB, Xu D, Costas BA, Katz LA.** 2010. Genetic identities of cryptic species in the *Strombidium stylifer*/*apolatum*/*oculatum* cluster, including a description of *Strombidium rassoulzadegani* n. sp. J Eukaryot Microbiol **57:**369–378. http://dx.doi.org/10.1111/j.1550-7408.2010.00485.x.
52. **Esteban GF, Finlay BJ.** 2003. Cryptic freshwater ciliates in a hypersaline lagoon. Protistologica **154:**411–418. http://dx.doi.org/10.1078/143446103322454149.
53. **Huang J, Chen Z, Song W, Berger H.** 2014. Three-gene based phylogeny of the Urostyloidea (Protista, Ciliophora, Hypotricha), with notes on classification of some core taxa. Mol Phylogenet Evol **70:**337–347. http://dx.doi.org/10.1016/j.ympev.2013.10.005.
54. **Li J, Liu W, Gao S, Warren A, Song W.** 2013. Multigene-based analyses of the phylogenetic evolution of oligotrich ciliates, with consideration of the internal transcribed spacer 2 secondary structure of three systematically ambiguous genera. Eukaryot Cell **12:**430–437. http://dx.doi.org/10.1128/EC.00270-12.
55. **Zhang Q, Yi Z, Fan X, Warren A, Gong J, Song W.** 2014. Further insights into the phylogeny of two ciliate classes Nassophorea and Prostomatea (Protista, Ciliophora). Mol Phylogenet Evol **70:**162–170. http://dx.doi.org/10.1016/j.ympev.2013.09.015.
56. **Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K.** 1993. Current protocols in molecular biology. Wiley-Liss, New York, NY.
57. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221–224.
58. **Guindon S, Gascuel O.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696–704. http://dx.doi.org/10.1080/10635150390235520.
59. **Librado P, Rozas J.** 2009. DnaSP V5: a software for comprehensive analysis of DNA polymorphism data. BioInformatics **25:**1451–1452. http://dx.doi.org/10.1093/bioinformatics/btp187.

Downloaded from http://mbio.asm.org/ on April 13, 2021 by guest