8-23-2010

Alternative Processing of Scrambled Genes Generates Protein Diversity in the Ciliate *Chilodonella uncinata*

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

Alexandra M. Kovner
Smith College

Published in final edited form as:
J Exp Zool B Mol Dev Evol. 2010 September 15; 314(6): 480–488. doi:10.1002/jez.b.21354.

Alternative processing of scrambled genes generates protein diversity in the ciliate *Chilodonella uncinata*

Laura A. Katz[1,2,*] and **Alexandra M. Kovner**[1]

[1]Department of Biological Sciences, Smith College, 44 College Lane, Northampton, Massachusetts 01063, USA

[2]Program in Organismic and Evolutionary Biology, University of Massachusetts, 611 North Pleasant Street, Amherst, Massachusetts 01003, USA

Abstract

In ciliates, chromosomal rearrangements occur during the development of the somatic macronuclear genome from the germline micronuclear genome. These rearrangements are extensive in three ciliate classes – Armophorea, Spirotrichea and Phyllopharyngea – generating a macronucleus with up to 20,000,000 gene-sized chromosomes. Previously, we have shown that these three classes also share elevated rates of protein evolution relative to other ciliates. To assess the evolution of germline-limited sequences in the class Phyllopharyngea, we used a combination of traditional and walking PCR to analyze micronuclear copies of multiple genes from two lines of the morphospecies *Chilodonella uncinata* for which we had previously characterized macronuclear sequences. Analyses of the resulting data yield three main results: 1) conserved macronuclear (somatic) regions are found within rapidly evolving micronuclear (germline) regions; 2) gene scrambling exists within this ciliate lineage; and 3) alternative processing of micronuclear regions yields diverse macronuclear β-tubulin paralogs. To our knowledge, this is the first study to demonstrate gene scrambling outside of the non-sister class Spirotrichea, and to show that alternative processing of scrambled genes generates diversity in gene families. Intriguingly, the Spirotrichea and Phyllopharyngea are also united in having transient ‘giant’ polytene chromosomes, gene-sized somatic chromosomes, and elevated rates of protein evolution. We hypothesize that this suite of characters enables these ciliates to enjoy the benefits of asexuality while still maintaining the ability to go through sexual cycles. The data presented here add to the growing evidence of the dynamic nature of eukaryotic genomes within diverse lineages across the tree of life.

Introduction

Genomes vary dramatically in structure and content within diverse lineages across the tree of life (Parfrey et al., '08). In ciliates, this variation is evident in comparisons between the two distinct genomes within each cell: the ‘germline’ micronuclear genome and the ‘somatic’ macronuclear genome. The micronucleus is largely transcriptionally inactive and a meiotic product of the micronucleus is exchanged between cells during conjugation to form a genetically novel zygotic nucleus (McGrath et al., '06, Juranek and Lipps, '07). The resulting zygotic nucleus then divides by mitosis and at least one daughter nucleus develops into a transcriptionally active macronucleus through a series of chromosomal rearrangements (McGrath et al., '06, Juranek and Lipps, '07). The extent of chromosomal processing varies among ciliates: the model ciliate *Tetrahymena* (Oligohymenophorea)

*Corresponding author: Laura Aline Katz – lkatz@smith.edu, Department of Biological Sciences, Smith College, 44 College Lane, Northampton, Massachusetts 01063, USA; phone: 413-585-3825; fax: 413-585-3786.

generates ~180 macronuclear chromosomes from five zygotic chromosomes while ciliates in the class Spirotrichea generate ~20,000,000 ‘gene sized’ chromosomes from ~100 zygotic chromosomes (Prescott, '94, Eisen et al., '06, McGrath et al., '06).

Ciliates in three classes, Armophorea, Spirotrichea, and Phyllopharyngea (the focus of this study), extensively process their genomes to generate a macronucleus with thousands of gene-sized chromosomes, each of which contains only coding domains, untranslated regions, and telomeres (Riley and Katz, '01). These three lineages represent at least two independent origins of extensive fragmentation as phylogenetic analyses reject the monophyly of the Phyllopharyngea with the Spirotrichea (Riley and Katz, '01). These extensively fragmenting lineages share elevated rates of protein evolution relative to all other ciliate lineages (Zufall et al., '06).

Gene scrambling refers to the presence of fragmented coding domains, termed macronuclear destined sequences (MDSs), whose order is scrambled in the micronucleus compared to that of macronuclear chromosomes. This process has been described in only a subset of lineages in the class Spirotrichea (Prescott and Greslin, '92, Curtis and Landweber, '99, Nowacki and Landweber, '09). For example, the actin I gene of the ciliate *Oxytricha trifallax* exists in the micronucleus as MDSs that are ordered 3-4-6-5-7-8-2-1-9 compared to their relative position on the macronuclear chromosome (Prescott and Greslin, '92). During the development of the macronuclear genome, these scrambled MDSs are reordered and intervening sequences (Internally Excised Sequences (IESs)) are removed to generate functional genes (Greslin et al., '89, Prescott, '00). Direct repeats, or pointer sequences, of 2–10 nucleotide sequences at MDS/IES junctions are likely involved in determining the correct order of coding domains (Prescott, '00, Juranek and Lipps, '07). Recent work in the class Spirotrichea suggests that unscrambling may be mediated by RNA scanning coupled with a templated correction of misprocessed molecules (Juranek et al., '05, Mollenbeck et al., '08) and may also involve transposases (Nowacki et al., '09).

To assess the evolution of germline-limited sequences in a member of the class Phyllopharyngea, we analyzed multiple genes from micronuclear enriched DNA preparations generated from two clonal lines of the morphospecies *Chilodonella uncinata*, one isolated from Poland and the other from the USA (Robinson and Katz, '08). Using a combination of traditional and walking PCR, we characterized homologous micronuclear regions of actin, α-tubulin and four β-tubulin paralogs as macronuclear copies of these genes have already been described (Robinson and Katz, '08). Analyses of macronuclear β-tubulin in these two strains of *C. uncinata* revealed a family of five paralogs that vary in their degree of conservation: two β-tubulin paralogs, ‘shared P1’ (SP1) and ‘shared p2’ are highly conserved (<1% different at the nucleotide level and with identical amino acid sequences) while the remaining three paralogs P1, P2 and P3, are more divergent with up to 14% nucleotide differences between the P3s (Robinson and Katz, '08).

Materials and Methods

Micronuclear DNA Extraction

The American line of *C. uncinata* was obtained from the ATCC®50194 while the Polish line was provided by Stefan Radizowski as previously reported (Robinson and Katz, '08) and is now deposited as ATCC®PRA-256. Cultures were treated with Penicillin-Streptomycin-Amphotericin-B (Lonza 17–745H) prior to pelleting cells by centrifugation at 5,000 rpm for 20 minutes. DNA was extracted by standard phenol/chlorophorm protocol (Ausubel et al., '93). Genomic DNA was enriched for micronuclear DNA by digesting with Bal-31 Nuclease (New England Biolabs M02135) according to manufacturer's guidelines. Micronuclear DNA was gel isolated by gel electrophoresis of digested DNA for ~3 hours

using Low Melt UltraCleanTM Agarose (Mobio15005-50). Gel isolated micronuclear DNA was purified using beta-agarase (New England Biolabs M03925).

Traditional PCR & Cloning

All micronuclear genes were amplified using Phusion Hot Start High Fidelity DNA Polymerase (Finnzymes F-540L). Alpha-tubulin, β-tubulin, and actin were amplified using primers described previously (Robinson and Katz, '08, Tekle et al., '08) as well as designed for this study (Table S1). Phusion amplified PCR products were cloned using Zero Blunt Topo cloning kits (Invitrogen 42-0245), multiple (2–8) clones were characterized per PCR reaction, and sequences have been submitted to GenBank (Table S2).

Genome Walking PCR & Cloning

We used Seegene's DNA Walking SpeedUp™ kit (K1052) to amplify additional regions of actin and β-tubulin paralogs. PCR amplification was performed following Seegene kit protocol using kit primers and gene-specific primers designed for this study (Table S1). Genome Walking PCR products were cloned using TA Topo cloning kits (Invitrogen 45-0641), multiple (2–8) clones were characterized per PCR reaction, and sequences have been submitted to GenBank (Table S2).

Sequencing

Sequencing reactions were carried with Applied Biosystems BigDye Terminator Ready Reaction Mix (4337456) and sequenced by the Center for Molecular Biology at Smith College in Northampton, MA. Additional sequencing reactions and sequencing were performed at Penn State Genomics Core Facility in University Park, PA.

Data Analysis

Contigs of paralogs were assembled in Seqman (DNAStar) and sequences were annotated using Seqbuilder (DNAstar). All polymorphisms were confirmed by eye. MacClade (Maddison and Maddison, '05) and Megalign (DNAStar) were used to create alignments. The IESs were identified as insertions in the micronuclear sequences as compared to macronuclear sequences. Direct repeats were inferred from comparisons on micronuclear and macronuclear sequences (Table 1). Meme was used to search for motifs in IESs (http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi). DNAsp was used to perform sliding window analysis to calculate average pairwise differences (Pi). Sliding window analyses of actin, α - tubulin, and SP1 were performed with a 100 base pair window and a 5 base pair step. Sliding window analysis of β-tubulin paralogs P1, P2 and P3 were performed with the windows and base steps of 50 and 5, 15 and 3, and 50 and 3, respectively.

Results

Conserved Macronuclear Destined Sequences Reside in Rapidly Evolving Micronuclear-Limited Sequences

We found micronuclear-limited IESs within coding domains of actin, α-tubulin and four β-tubulin paralogs (SP1, P1, P2 and P3), where they appear as insertions compared to macronuclear sequences (Fig. 1, Table 1). In all cases, IESs are evolving so rapidly that they cannot be aligned with confidence even though MDSs, which correspond to coding domains, are generally highly conserved (*i.e.* π in Fig. 1). For example, the three IESs found at homologous sites within the α-tubulin sequences have an average pairwise divergence (π) as high as 0.40 while coding domains are nearly identical between the USA and Polish strains (Fig. 1A). Even for the divergent macronuclear β-tubulin P3 paralogs, IESs are in homologous positions in the two strains, though their sequences are highly divergent (Fig.

1A). In three of the genes, α-tubulin, β-tubulin P3, and β-tubulin SP1, the MDSs are collinear with the known macronuclear sequences, as indicated by the consecutive numbering in Figure 1A. Direct repeats mark the IES/MDS boundaries for these non-scrambled genes (Table 1).

Scrambled Coding Domain in the Micronucleus

We found evidence of gene scrambling in micronuclear copies of three of the six genes examined from both strains: actin, β-tubulin P1 and β-tubulin P2 (Fig. 1B). The first MDS of actin, including the start codon and 5' UTR, is inverted in relation to the adjacent micronuclear coding domains in both the Polish and USA strains (Fig. 1B). As with the non-scrambled genes, the IES/MDS junctions are at the same location in the two strains and the sequences of the MDSs are generally conserved (Fig. 1B). In contrast, the sequences within the actin IESs are highly diverged (Fig. 1B) and the IES1 of the Polish strain contains a microsatellite that is absent from the USA strain (GenBank nos. GQ453433, GQ45343). Scrambling is also present in the P1 and P2 β-tubulin paralogs as MDSs are arranged in a nonconsecutive order. For example, coding domains corresponding to positions 619–743, 811–893, 956–1001 are missing from the characterized region of the P2 β-tubulin locus (Fig. 1B).

Macronuclear β-Tubulin Paralogs Are Assembled From Alternatively Processed MDSs from Multiple Micronuclear Loci

Two lines of evidence point to the alternative assembly of the β-tubulin paralogs P1 and P2, whereby macronuclear sequences are generated by splicing together segments that are interdigitated on multiple micronuclear loci. First, comparisons of the macronuclear versions of these paralogs plus β-tubulin paralog SP1 reveal multiple identical regions (Fig. 2). Beta-tubulin paralogs P1 and P2 are identical in four regions, as indicated by $\pi = 0$ and shared colors at identical locations in cartoons. (Fig. 2). Similarly, paralogs P2 and SP1 are identical in four regions (Fig. 2). It is possible that a process such as gene conversion generated these identical regions among β-tubulin paralogs. However, walking PCRs from the shared identical macronuclear regions consistently yielded the micronuclear sequences indicated in Figure 1, and no additional micronuclear sequences; this indicates that the identical macronuclear regions are not found on multiple micronuclear loci. A second line of evidence for alternative processing is the conservation of pointer sequences, ranging from 2–8 bp, at the boundaries of the β-tubulin P1, P2 and SP1 micronuclear loci (Table 1; Fig. 3). Conservation of pointer sequences between molecules is consistent with unscrambling (Prescott, '00,Juranek and Lipps, '07). Hence, we propose that macronuclear β-tubulin paralogs P1 and P2 are assembled by alternative processing of MDSs found at multiple micronuclear loci (P1, P2, SP1; Fig. 3).

Germline-Limited IESs Contain Conserved Motifs

We inspected our IESs for conserved motifs that might be involved in IES excision as such motifs have previously been found in *C. uncinata* (Katz et al., '03, Zufall and Katz, '07). Analyses using MEME (http://meme.nbcr.net/meme4_1/cgi-bin/meme.cgi) (Bailey et al., '09) revealed a conserved motif (Fig. 4) that is consistent with the previously reported motif (Katz et al., '03, Zufall and Katz, '07). We find a similar motif when the search is constrained to find only a single motif per sequence or any number of motifs, and when we search through both scrambled and non-scrambled IESs. This indicates that the conserved motif, the consensus of which is KGWKKSTR, is a critical *cis*-acting sequence involved in IES excision in *C. uncinata*. We found no evidence of this conserved motif specific to alternatively generated IESs.

Discussion

The three main results from this study of two strains of the ciliate *C. uncinata* are: 1) conserved macronuclear (somatic) regions exist within rapidly evolving micronuclear (germline) regions; 2) gene scrambling exists within this ciliate; and 3) alternative processing of micronuclear regions yields diverse macronuclear β-tubulin paralogs. For the first observation, the pattern of conservation between macronucleus and micronucleus in the Phyllopharyngean *C. uncinata* is consistent with previous work on species in the classes Spirotrichea and Oligohymenophorea. For example, there is high divergence among IESs in *Oxytricha* species (Seegmiller et al., '96, DuBois and Prescott, '97) and between strains of *Stylonychia lemnae* (Ardell et al., '03), both members of the class Spirotrichea. The limited data on rates of evolution in *Tetrahymena* (Oligohymenophorea) also indicate that micronuclear-restricted sequences are evolving very rapidly (Huvos, '04).

The presence of gene scrambling in the Phyllopharyngean *C. uncinata* plus members of the non-sister class Spirotrichea (Riley and Katz, '01) indicates either that gene scrambling has arisen twice within ciliates or that it is an ancient mechanism that has been lost in some lineages. To date, micronuclear data exist only from these two classes plus the Oligohymenophorea for which there is no evidence of gene scrambling (Fig. 5). Intriguingly, gene scrambling in Spirotrichea and Phyllopharyngea is associated with a suite of characters that do appear unique to these two classes: the presence of ‘giant’ highly polytenized chromosomes during macronuclear development (Ammermann, '87); extensively-processed macronuclear genomes that consist of many gene-size chromosomes (Steinbrück et al., '95, Riley and Katz, '01); and elevated rates of protein evolution (Zufall et al., '06). (This suite of characters may also be present in the poorly studied class Armophorea, which in turn may be sister to the class Spirotrichea (Fig. 5 and Riley and Katz, 2001)). As these latter traits are not known to be present in other ciliate classes (Fig. 5), we suggest that the entire suite of characters has arisen twice: once in the Phyllopharyngea and once in the distantly related Spirotrichea.

Based on two lines of evidence, 1) the sharing of identical regions between macronuclear molecules and 2) the presence of pointer sequences between micronuclear loci, we hypothesize that macronuclear copies of β-tubulin paralogs P1 and P2 are assembled by alternative processing of sequences from SP1, P1 and P2 micronuclear loci (Fig. 3). Under this model, macronuclear loci are generated by ‘unscrambling’ regions from multiple micronuclear loci, and some of these regions are also used to generate other macronuclear loci (Fig. 3). This alternative processing contributes to the diversity of proteins observed in *C. uncinata*.

Finally, the sharing of a suite of characters – scrambling, extensive fragmentation, giant chromosomes, faster protein evolution – in the non-sister classes Phyllopharyngea and Spriotrichea suggests that these features are adaptive. We hypothesize that this suite of characters enables these ciliates to enjoy the benefits of asexuality while still maintaining the ability to go through sexual cycles. Cell division and hence population growth in ciliates occurs asexually and, in all ciliates except the class Karyorelictea, the polyploid macronucleus divides by the poorly-understood process of amitosis (McGrath et al., '06). In turn, amitosis enables selection to increase the frequency of adaptive loci in ciliates in macronuclei (Zufall et al., '06). Even when sex does occur, the next generation macronucleus can be patterned from the previous generation through epigenetic processes: at least in members of the Oligohymenophorea and Spirotrichea, alterations made to the parental macronucleus can be inherited by the daughter macronucleus through an RNA-scanning mechanism (Juranek and Lipps, '07, Mollenbeck et al., '08). Hence, ciliate macronuclei are essentially asexual, enabling maintenance of adaptive genes and gene

complexes. Occasional sex can purge the somatic macronuclei of deleterious mutations either by eliminating variants following recombination during meiosis (Normark et al., '03) or by 'resetting' the polyploidy macronucleus to contain the coding domains from the micronucleus (Kondrashov, '97).

Under this hypothesis, each of the suite of characters shared between the Spirotrichea and Phyllopharyngea (Fig. 5) contributes to the efficiency of selection during asexual cycles. Gene scrambling, which requires both extensive fragmentation and giant chromosomes (*i.e.* multiple copies of micronuclear loci), enables generation of protein diversity though alternative processing (Herrick et al., '87,Jonsson et al., '09). Extensive processing during macronuclear development also breaks down linkage between genes, enabling selection to operate on specific loci without impacting linked genes. As a result, ciliates in these two classes explore protein space in novel manners as evidenced by the statistically significant increase in accumulation of divergent paralogs in numerous gene families (Zufall et al., '06). At the same time, purifying selection is very effective in maintaining conserved macronuclear coding-domains nested within rapidly evolving micronuclear-limited sequences – nearly identical coding domains are maintained for α-tubulin, actin and the SP1 β-tubulin paralogs between strains of the morphospecies *C. uncinata* even though sufficient time has passed for mutations to make IESs highly divergent (π in Fig. 1).

The data presented here indicate that the molecular evolution of coding domains, which can exist as scrambled pieces in a background of rapidly evolving germline DNA, are shaped by selection interacting with the genome architecture of this ciliate. Further, these data add to the growing evidence on the dynamic nature of eukaryotic genomes as genomes vary in structure and content within diverse lineages across the tree of life (Parfrey et al., '08).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by NIH AREA award 1R15GM081865-01. The authors are grateful to Ben Normark, Laura Wegener Parfrey and Dan Lahr, all at the University of Massachusetts Amherst, for thoughtful discussions of earlier drafts of this manuscript.

Literature Cited

Ammermann D. Giant chromosomes in ciliates. Results Probs Cell Differ. 1987; 14:59–67.

Ardell DH, Lozupone CA, Landweber LF. Polymorphism, recombination and alternative unscrambling in the DNA polymerase alpha gene of the ciliate Stylonychia lemnae (Alveolata; class Spirotrichea). Genetics. 2003; 165:1761–1777. [PubMed: 14704164]

Ausubel, FM.; Brent, R.; Kingston, RE.; Moore, DD.; Seidman, JG.; Smith, JA.; Struhl, K. Current Protocols in Molecular Biology. New York: Wiley-Liss; 1993.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009; 37:W202–W208. [PubMed: 19458158]

Curtis EA, Landweber LF. Evolution of gene scrambling in ciliate micronuclear genes. Ann N Y Acad Sci. 1999; 870:349–350. [PubMed: 10415495]

DuBois ML, Prescott DM. Volatility of internal eliminated segments in germ line genes of hypotrichous ciliates. Mol Cell Bio. 1997; 17:326–337. [PubMed: 8972213]

Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM, Tallon LJ, Delcher AL, Salzberg SL, Silva JC, Haas BJ, Majoros WH, Farzad M, Carlton JM, Smith RK, Garg J, Pearlman RE, Karrer KM, Sun L, Manning G, Elde NC, Turkewitz

AP, Asai DJ, Wilkes DE, Wang Y, Cai H, Collins K, Stewart BA, Lee SR, Wilamowska K, Weinberg Z, Ruzzo WL, Wloga D, Gaertig J, Frankel J, Tsao CC, Gorovsky MA, Keeling PJ, Waller RF, Patron NJ, Cherry JM, Stover NA, Krieger CJ, Del Toro C, Ryder HF, Williamson SC, Barbeau RA, Hamilton EP, Orias E. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. PLoS Biol. 2006; 4:e286. [PubMed: 16933976]

Gong J, Stoeck T, Yi Z, Miao M, Zhang Q, Roberts DM, Warren A, Song W. Small subunit rRNA phylogenies show that the class Nassophorea is not monophyletic (Phylum Ciliophora). J Eukaryot Microbiol. 2009; 56:339–347. [PubMed: 19602079]

Greslin, aF; Prescott, DM.; Oka, Y.; Loukin, SH.; Chappell, JC. Reordering of 9 exons is necessary to form a functional actin gene in *Oxytricha nova*. Proc Natl Acad Sci U S A. 1989; 86:6264–6268. [PubMed: 2503830]

Herrick G, Cartinhour SW, Williams KR, Kotter KP. Multiple sequence versions of the *Oxytricah fallax* 81-MAC alternate processing family. J Protozool. 1987; 34:429–434. [PubMed: 3123649]

Huvos P. Modular structure in developmentally eliminated DNA in *Tetrahymena* may be a consequence of frequent insertions and deletions. J Mol Biol. 2004; 336:1075–1086. [PubMed: 15037070]

Jonsson F, Postberg J, Lipps HJ. The unusual way to make a genetically active nucleus. DNA Cell Biol. 2009; 28:71–78. [PubMed: 19196049]

Juranek SA, Lipps HJ. New insights into the macronuclear development in ciliates. Int Rev Cytol. 2007; 262:219–251. [PubMed: 17631190]

Juranek SA, Rupprecht S, Postberg J, Lipps HJ. snRNA and heterochromatin formation are involved in DNA excision during macronuclear development in stichotrichous ciliates. Eukaryot Cell. 2005; 4:1934–1941. [PubMed: 16278460]

Katz LA, Lasek-Nesselquist E, Snoeyenbos-West OLO. Structure of the micronuclear α-tubulin gene in the phyllopharyngean ciliate *Chilodonella uncinata*: implications for the evolution of chromosomal processing. Gene. 2003; 315:15–19. [PubMed: 14557060]

Kondrashov AS. Evolutionary genetics of life cycles. Annu Rev Ecol Syst. 1997; 28:391–435.

Li, C-W.; Chen, J-Y.; Lipps, JH.; Gao, F.; Chi, H-M.; Wu, H-J. Ciliated protozoans from the Precambrian Doushantuo Formation, Wengan, South China. In: Rich, PV.; Komarower, P., editors. The rise and fall of the Ediacaran biota. London: Geological Society; 2007. p. 151-156.

Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25:1451–1452. [PubMed: 19346325]

Maddison, WP.; Maddison, DR. MacClade v. 4.0.8. Sinauer Assoc; 2005. editors

McGrath, C.; Zufall, RA.; Katz, LA. Genome evolution in ciliates. Katz, LA.; Bhattacharya, D., editors. Genomics and Evolution of Eukaryotic Microbes: Oxford University Press; 2006.

Mollenbeck M, Zhou Y, Cavalcanti AR, Jonsson F, Higgins BP, Chang WJ, Juranek S, Doak TG, Rozenberg G, Lipps HJ, Landweber LF. The pathway to detangle a scrambled gene. PLOS ONE. 2008; 3:e2330. [PubMed: 18523559]

Normark BB, Judson OP, Moran NA. Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society. 2003; 79:69–84.

Nowacki M, Higgins BP, Maquilan GM, Swart EC, Doak TG, Landweber LF. A functional role for transposases in a large eukaryotic genome. Science. 2009; 324:935–938. [PubMed: 19372392]

Nowacki M, Landweber LF. Epigenetic inheritance in ciliates. Curr Opin Microbiol. 2009; 12:638–643. [PubMed: 19879799]

Parfrey LW, Lahr D, Katz LA. The dynamic nature of eukaryotic genomes. Mol Biol Evol. 2008; 25:787–794. [PubMed: 18258610]

Prescott DM. The DNA of ciliated protozoa. Microbiol Rev. 1994; 58:233–267. [PubMed: 8078435]

Prescott DM. Genome gymnastics: Unique modes of DNA evolution and processing in ciliates. Nature Reviews Genetics. 2000; 1:191–198.

Prescott DM, Greslin AF. Scrambled Actin I Gene in the Micronucleus of *Oxytricha nova*. Dev Genet. 1992; 13:66–74. [PubMed: 1395144]

Riley JL, Katz LA. Widespread distribution of extensive genome fragmentation in ciliates. Mol Biol Evol. 2001; 18:1372–1377. [PubMed: 11420375]

Robinson T, Katz LA. Non-Mendelian inheritance of two cytoskeletal genes in the ciliate *Chilodonella uncinata*. Mol Biol Evol. 2008; 24:2495–2503. [PubMed: 17890762]

Seegmiller A, Williams KR, Hammersmith RL, Doak TG, Witherspoon D, Messick T, Storjohann LL, Herrick G. Internal eliminated sequences interrupting the *Oxytricha* 81 locus: allelic divergence, conservation, conversions, and possible transposon origins. Mol Biol Evol. 1996; 13:1351–1362. [PubMed: 8952079]

Steinbrück G, Raszikowski S, Golembiewska-Skoczylas M, Sapetto-Rebow B. Characterization of low and high molecular weight DNA in the macronucleus of the ciliate *Chilodonella steini*. Acta Protozool. 1995; 34:125–134.

Tekle YI, Grant JR, Anderson OR, Cole JC, Nerad TA, Patterson DJ, Katz LA. Phylogenetic placement of diverse amoebae inferred from multigene analyses and assessment of clade stability within 'Amoebozoa' upon removal of varying rate classes of SSU-rDNA. Mol Phylogenet Evol. 2008; 47:339–352. [PubMed: 18180171]

Wright A-DG, Lynn DH. Maximum ages of ciliate lineages estimated using a small subunit rRNA molecular clock: crown eukaryotes date back to the paleoprotoerozoic. Archiv Fur Protistenkunde. 1997; 148:329–341.

Zufall RA, Katz LA. Micronuclear and macronuclear forms of beta-tubulin genes in the ciliate *Chilodonella uncinata* reveal insights into genome processing and protein evolution. J Eukaryot Microbiol. 2007; 54:275–282. [PubMed: 17552983]

Zufall RA, McGrath C, Muse SV, Katz LA. Genome architecture drives protein evolution in ciliates. Mol Biol Evol. 2006; 23:1681–1687. [PubMed: 16760419]

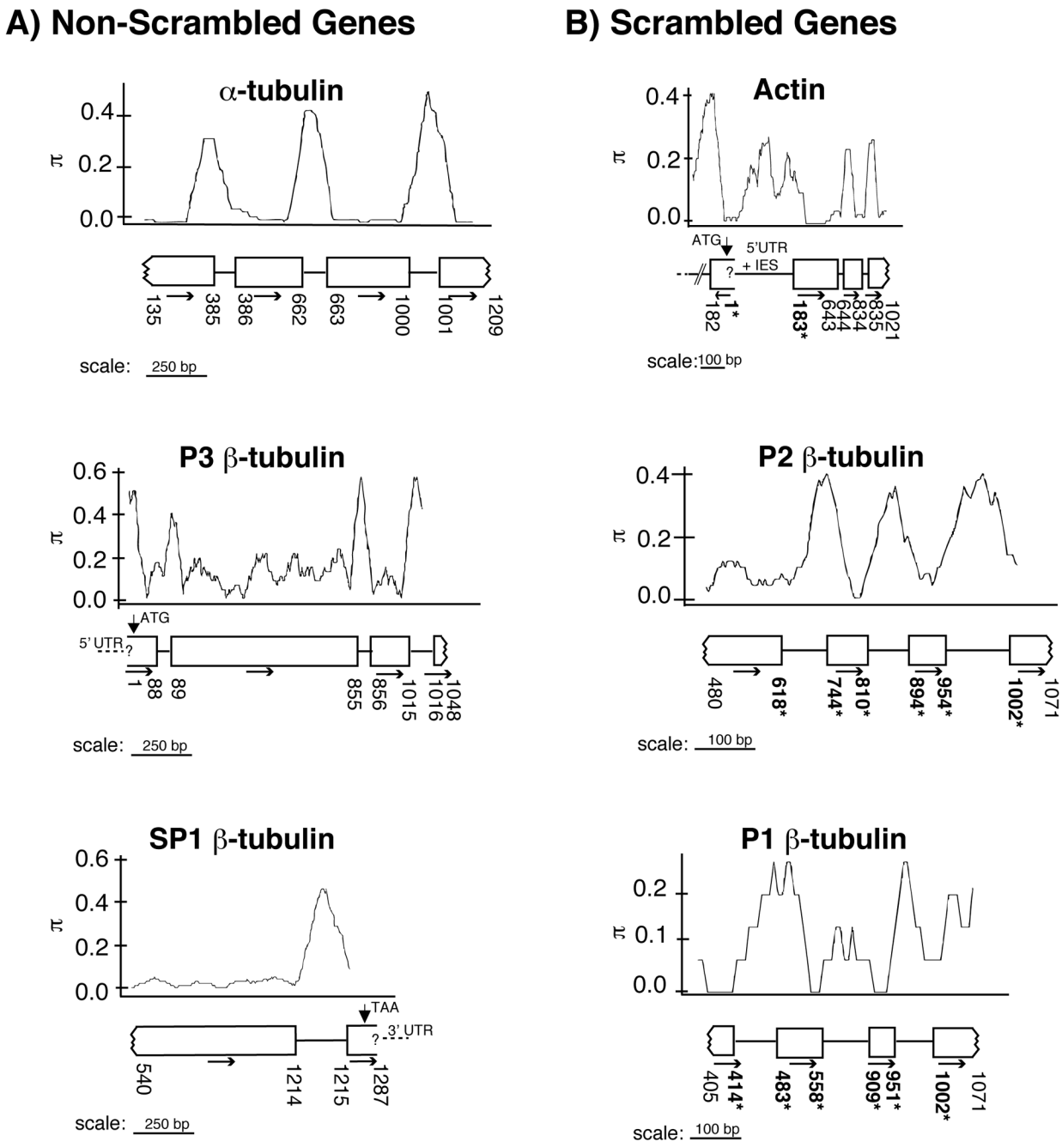

Figure 1. Diagram of micronuclear structures of a) three scrambled genes and b) three non-scrambled genes

The graphs are sliding window analyses of pairwise divergence (π) between sequences from USA and Poland, calculated using DNAsp (Librado and Rozas, '09). The cartoons for each gene indicate the structure of the MDSs (boxes) and IESs (lines) that are shared between the lines. Numbers below the MDSs are position relative to coding domain. Scrambling is indicated by non-sequential numbering of MDSs (bold numbers with asterix), or in the case of actin, by the inversion of MDS numbering and arrow below boxes. Jagged edges reflect missing data and '?' indicate areas where CDS/UTR boundaries cannot be determined.

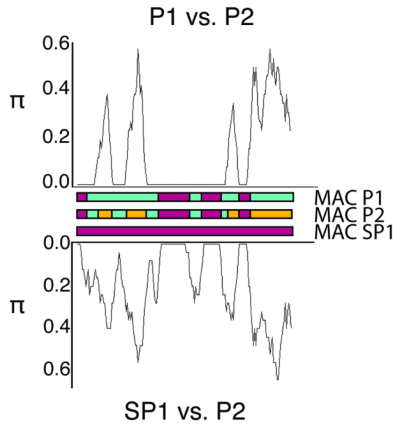

Figure 2. MAC β-tubulin paralogs P1, P2 and SP1 share some identical regions

Graph is sliding window analyses of pairwise divergence (π) calculated using DNAsp (Librado and Rozas, '09). Top comparison is of macronuclear (MAC) P1 and P2; bottom comparison of P2 and SP1. Regions with the same color at identical positions correspond to shared sequences, and colors correspond to micronuclear loci in Figure 3.

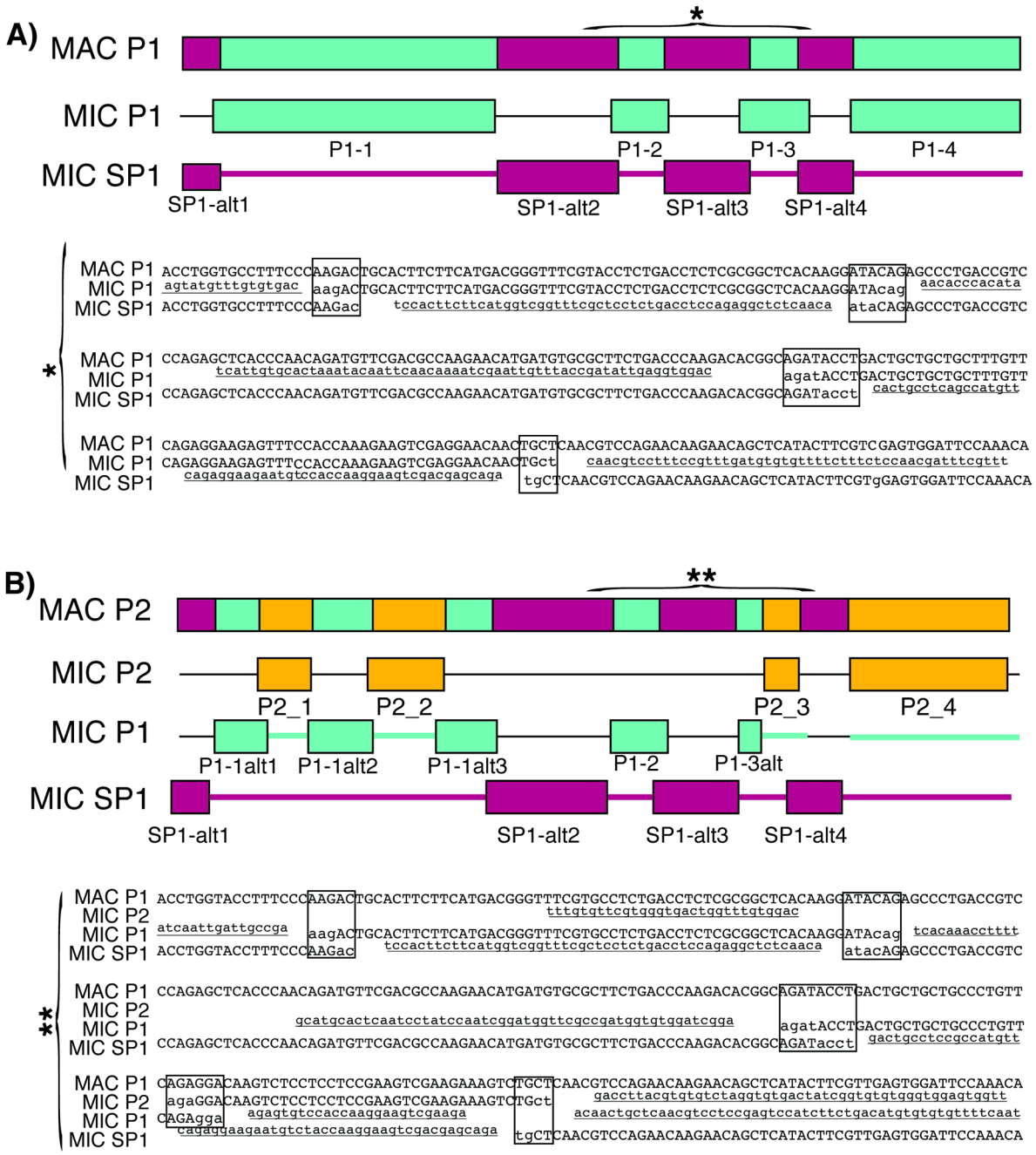

Figure 3. Micronuclear sequences can be alternatively spliced together to make diverse β-tubulins

Each color refers to a micronuclear locus (MIC P1, MIC P2, and MIC SP1), boxes indicate MDSs, black lines are traditional IESs and thick colored lines are alternatively processed IESs. The large MDS used to generate SP1 (Figure 1A) is divided into four MDS (MIC SP1-alt1-4) that are then spliced to MDSs from the MIC P1 locus to generate the MAC paralog P1 (Fig 31). These same four alternatively processed MIC SP1 MDSs are combined with alternatively processed MDSs from MIC P1 plus MDSs from MIC P2 to generate MAC P2 (Fig 3B). The brackets and asterix above the MAC molecules correspond to the region of sequence data below each figure, shown to exemplify boundaries: * = MAC P1, ** =MAC

P2. The nucleotides in boxes are the conserved pointer sequences as detailed in Table 1, with uppercase letters for MDSs and lowercase letters for IESs.

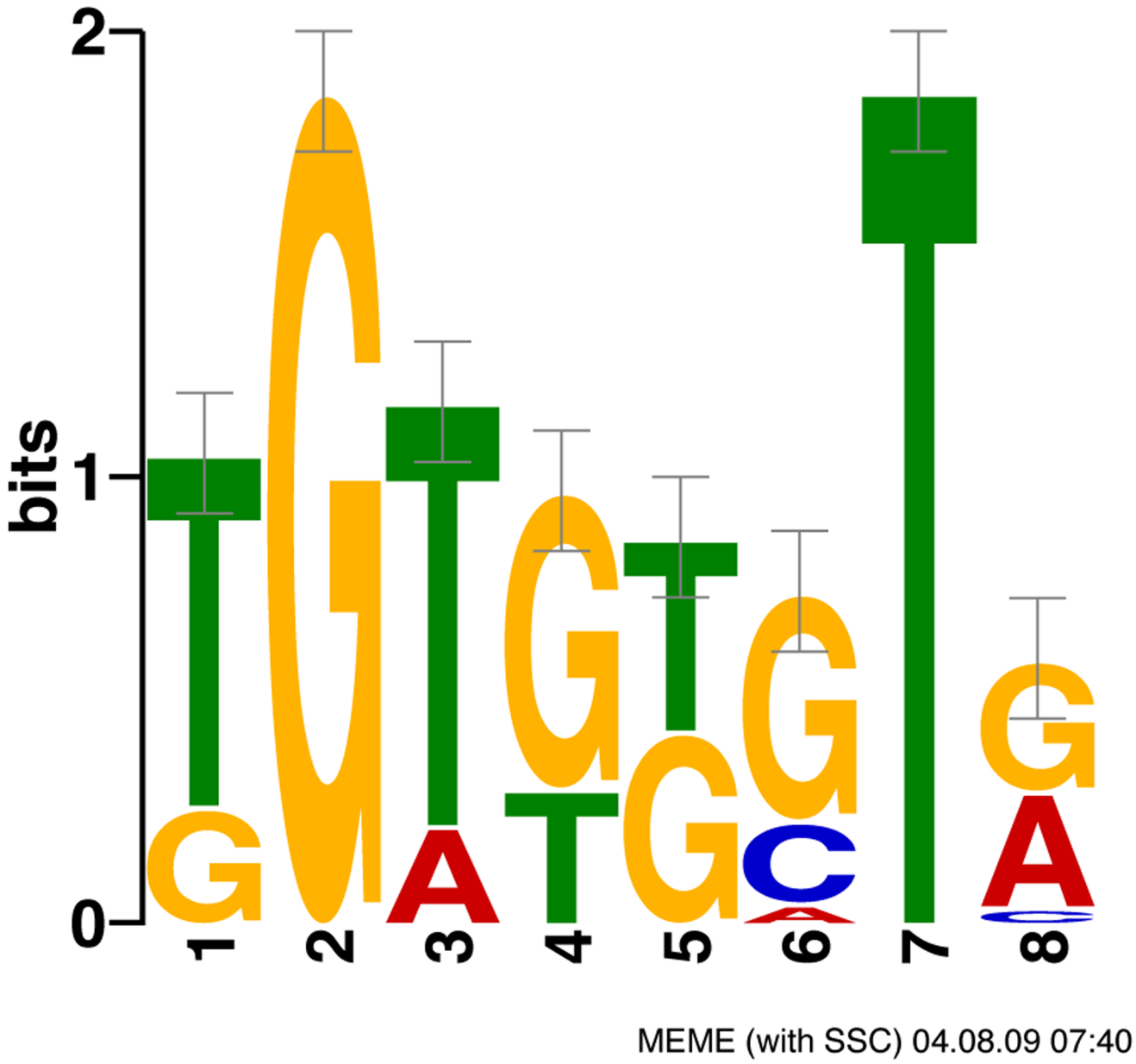

Figure 4. Conserved motif that may be a *cis*-acting signal for excision of both scrambled and non-scrambled IESs

Motif was generated using MEME (Bailey et al., '09) by searching among all IESs from the polish strain of *C. uncinata* with the constraint of identifying only one motif per sequence. The motif is significant as it has an E value of 4.8e+000 while analysis of the same sequences shuffled yields a much higher E value of = 1.2e+005. A similar motif was obtained when analyzing the IESs from the USA line, the scrambled IESs and the non-scrambled IESs, regardless as to whether we constrained the program to find just one motif per sequence or any number of motifs.

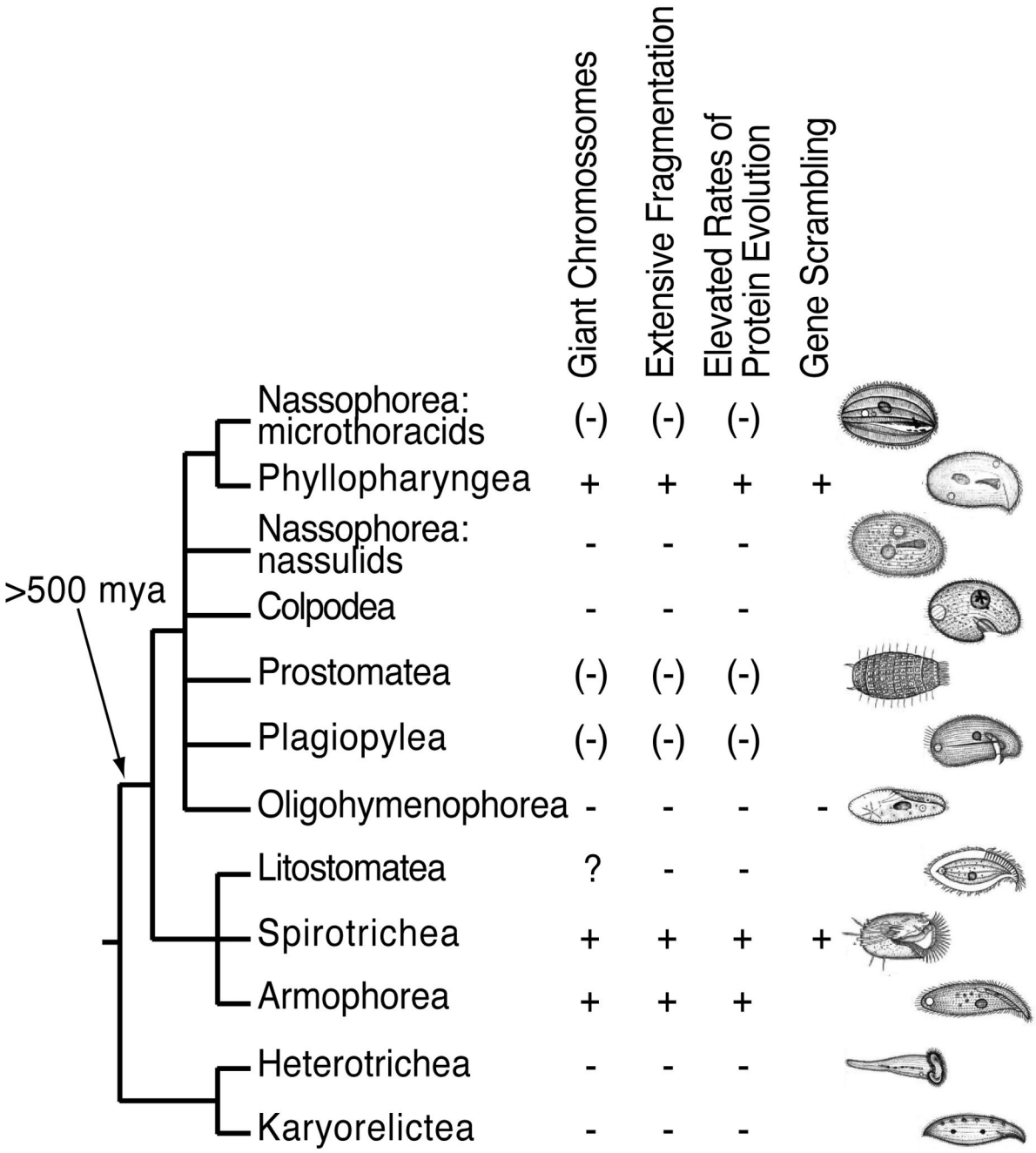

Figure 5. Evolution of genome characteristics among ciliates

A suite of four characters is found in the non-sister classes Phyllopharyngea and Spirotrichea: giant chromosomes, extensive fragmentation, elevated rates of protein evolution and gene scrambling. '+' and '−' refer to presence and absence of characters, respectively. Symbols in parentheses indicate lineages for which there are only few data while no entry means an absence of data. The '?' for giant chromosomes in the class Litostomatea reflect an 1890 report of this feature in the genus *Loxophyllum* that has not been supported by subsequent studies (Ammermann, '87). The topology of the tree is redrawn from recent references and reflects the apparent non-monophyly of the class Nassophorea (Riley and Katz, '01, Gong et al., '09). The age of divergence between

Phyllopharyngea and Spirotrichea is estimated to be at least 580 million years based on fossil evidence (Li et al., '07) while a molecular clock estimate the age of the ciliate clade at >1 billion years (Wright and Lynn, '97). Representative ciliate images redrawn from http://microscope.mbl.edu/.

Table 1

Details of micronuclear loci.

Gene	MDS (bp)	Pointer Sequence	USA IES (bp)	POL IES (bp)
Actin	1–182	GAAGGGTGTC	-	-
	183–643	GAAGTGCT	62	64
	644–835	CAGTGC	55	54
Atub	135–385	ACTGCA	81	81
	386–662	GAGAGACC	88	86
	663–1000	GCCACCA	107	112
	1001–1209			
Btub SP1	<12–93	TGACCCAA	94	-
	94–1259	TCTCTGAG	196	145
	1260–>1295			
Btub P1	<307–340	ATCGACTC	alt.	alt.
	341–660	CCAACTT	89	85
	661–792	AAGAC	alt.	alt.
	793–843	ATACAG	80	74
	844–931	AGATACCT	alt.	alt.
	932–994	TGCT	100	111
	995–1047	ATCAA	alt.	alt.
	1048–1156			
Btub P2	<307–340	ATCGACTC	alt.	alt.
	341–402	CAGAT	alt.	alt.
	403–456	TTTCC	66	65
	457–528	TCAGA	alt.	alt.
	529–600	GT	74	59
	601–660	CCAACTT	alt.	alt.
	661–792	AAGAC	alt.	alt.
	793–843	ATACAG	alt.	alt.
	844–931	AGATACCT	alt.	alt.
	932–954	AGAGGA	alt.	alt.
	955–990	TGCT	61	77
	991–1047	ATCAA	alt.	alt.
Btub P3	1–92	CAGATGA	55	54
	92–905	CTC (T/G) AT	58	56
	906–1065	AAGCAG	89	98

Gene	MDS (bp)	Pointer Sequence	USA IES (bp)	POL IES (bp)
	1066–>1152			

Notes: MDS = macronuclear destined sequence, pointer sequence = direct repeat, bp = length. Numbers for β-tubulins are based on alignment of full length sequences (available upon request) while numbering for other loci are based on partial sequences. ‘Alt’ indicates alternatively processed regions (Fig. 3).