

## Micronuclear and Macronuclear Forms of $\beta$ -Tubulin Genes in the Ciliate *Chilodonella uncinata* Reveal Insights into Genome Processing and Protein Evolution

REBECCA A. ZUFALL<sup>a</sup> and LAURA A. KATZ<sup>b</sup>

<sup>a</sup>Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204, and

<sup>b</sup>Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063

**ABSTRACT.** *Chilodonella uncinata*, like all ciliates, contains two distinct nuclei in every cell: a germline micronucleus and a somatic macronucleus. During development of the macronucleus from a zygotic nucleus, the genome is processed in several ways, including elimination of internal sequences. In this study, we analyze micronuclear and macronuclear copies of  $\beta$ -tubulin in *C. uncinata* and find at least four divergent paralogs of  $\beta$ -tubulin in the macronucleus. We characterize the micronuclear version of one paralog and compare its internally eliminated sequences (IESs) with previously described IESs in this species. These comparisons reveal the presence of a conserved sequence motif within IESs. In addition, we compare the sequences of  $\beta$ -tubulin from *C. uncinata* with other ciliates and to other alveolates in order to test the hypothesis that the mode of molecular evolution in ciliates obscures phylogenetic signal in protein-coding genes. We find that heterogeneous rates of substitution in  $\beta$ -tubulin across ciliates result in unstable genealogies that are inconsistent with phylogenies based on small subunit rDNA genes and on ultrastructure. We discuss the implications of our findings for genome processing and protein evolution in ciliates.

CILIATES are microbial eukaryotes defined by the presence of two types of nuclei in every cell. The germline micronucleus is essentially transcriptionally inactive, whereas the macronuclear genome is transcriptionally active during vegetative growth. Both types of nuclei develop from zygotic nuclei following conjugation. During development of the macronucleus, chromosomes are fragmented and amplified, and internal sequences are eliminated (reviewed in Jahn and Klobutcher 2002; Katz 2001; McGrath, Zufall, and Katz 2006; Prescott 1994; Yao, Duharcourt, and Chalker 2002). In some ciliates, genome fragmentation results in small chromosomes that often carry a single gene. Such extensive fragmentation is found in three classes of ciliates, Phyllopharyngea (including *Chilodonella uncinata*, the focus of this study), Armophorea, and Spirotrichea (Klobutcher and Herrick 1997; Prescott 1994; Riley and Katz 2001), and has arisen independently at least twice within ciliates (Katz 2001; Riley and Katz 2001).

All ciliates studied to date possess internally eliminated sequences (IESs) that are eliminated from the zygotic genome during development of the macronucleus. Analyses of IESs indicate the presence of conserved motifs within some species, but little conservation of these sequences across species (Chalker et al. 1999; Katz, Lasek-Nasselquist, and Snoeyenbos-West 2003; Ku, Mayr, and Forney 2000; Seegmiller and Herrick 1998). *Cis*-acting sequences are involved in IES excision in some species in the class Oligohymenophorea (reviewed in Yao et al. 2002), but the role of these sequence motifs remains unclear. Internally eliminated sequences have been previously described only for the  $\alpha$ -tubulin gene of *C. uncinata* where an 8-bp conserved motif was identified in three IESs within a single  $\alpha$ -tubulin paralog (Katz et al. 2003).

Previous studies have also shown a correlation between genome processing and rates of molecular evolution in ciliates—ciliates tend to experience increased rates of protein evolution in comparison with other eukaryotes and ciliates that extensively fragment their genomes into gene-sized chromosomes have increased rates of protein evolution in comparison with other ciliates (Katz et al. 2004; Zufall et al. 2006). *Chilodonella uncinata* (Cl: Phyllopharyngea) is a member of one of three classes of cil-

iates with extensively fragmented genomes and shows the characteristic rapid protein evolution (Zufall et al. 2006).

To gain a better understanding of genome processing and protein evolution, we examine the micronuclear and macronuclear versions of the gene family coding  $\beta$ -tubulin in *C. uncinata*. We describe four highly divergent paralogs of  $\beta$ -tubulin found in the macronucleus and characterize the micronuclear copy of one of these paralogs. The IESs found in  $\beta$ -tubulin allow us to reevaluate previously reported IES conserved motifs in *C. uncinata*. In addition, we assess the hypothesis that the heterogeneous rates of molecular evolution in ciliates obscure phylogenetic signal in protein-coding genes.

### MATERIALS AND METHODS

**DNA isolation.** *Chilodonella uncinata* DNA was extracted using phenol and chloroform from a single-cell isolate clonal line, derived from ATCC 50194. Total genomic DNA was separated by electrophoresis on a 0.8% (w/v) low-melt agarose gel. Micronuclear DNA is mobility limited; macronuclear DNA appears as a smear between approximately 500 bp and 5 kb. To isolate micronuclear DNA, we excised the mobility-limited band and purified by digestion with AgarAce enzyme (Promega, Madison, WI) (Katz et al. 2003). Isolation and purification were repeated 2 times.

**Amplification and sequencing of  $\beta$ -tubulin.** Macronuclear versions of  $\beta$ -tubulin were previously identified (Zufall et al. 2006). Paralogs P1–P4 refer to GenBank Accession No. P1: DQ665918; P2: DQ665921; P3: DQ665920; P4: DQ665919. The presence of these sequences was confirmed and new sequences were identified as described in Zufall et al. (2006). Based on emerging data from whole macronuclear chromosomes from this species (McGrath, Zufall, and Katz 2007), conceptual translations were made using TAA as the only stop codon.

Micronuclear versions of  $\beta$ -tubulin were amplified from the purified mobility-limited DNA using Phusion DNA polymerase (New England BioLabs, Ipswich, MA). The first round of PCR was performed using the same primers as for macronuclear amplifications. Some of these amplification products contained a secondary band of approximately 1,450 bp, in addition to a primary band of approximately 975 bp (corresponding to the macronuclear fragment). The larger band, containing micronuclear-limited DNA, was excised and purified with UltraClean GelSpin kit (Mo Bio, Carlsbad, CA). This product was reamplified using the same primers and an additional internal primer,  $\beta$ -tub

Corresponding Author: R. Zufall, Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204—Telephone number: 713 743 8172; FAX number: 713 743 2636; e-mail: rzufall@uh.edu

738+ 5'-GCTTACCCGGNYARCTWAAC-3'. These products were similarly purified and cloned into PCR SMART ((Lucigen, Middleton, WI). Six clones from two independent amplification reactions were sequenced with vector primers.

**IES analysis.** Internally eliminated sequences were identified by aligning micronuclear and macronuclear sequences by hand. Internally eliminated sequences were characterized by comparisons of sequences identified from the large (> 15 kb as assessed by agarose gel electrophoresis) mobility-limited portion of the DNA (micronuclear) but not found in the total genomic DNA. Because macronuclear chromosomes are present in up to 1,000 copies per cell, we do not expect to be able to amplify micronuclear sequences from total genomic DNA. We identified IESs as sequences interrupting the  $\beta$ -tubulin coding domain that were bounded on either side by direct repeats.

Internally eliminated sequences were analyzed for conserved motifs using MEME version 3.5.1 (Bailey and Gribskov 1998). Internally eliminated sequences found in  $\beta$ -tubulin were analyzed independently from and together with three IESs previously identified in  $\alpha$ -tubulin from *C. uncinata* (Katz et al. 2003) and two IESs in an actin gene from the same species (Robinson, Zufall, and Katz, unpubl. data). Motif searches were performed assuming the presence of either one motif per sequence or any number of motifs per sequence. Motif significance was assessed by comparison with motifs found with shuffled characters.

**Phylogenetic analysis.** Macronuclear sequences from *C. uncinata* were compared with  $\beta$ -tubulin from other ciliates and other

well-sampled clades of alveolates (Table 1). Sequences were aligned in Clustal W (Thompson, Higgins, and Gibson 1994) and adjusted by hand. Eight hundred sixty-one nucleotide and 287 amino acid characters were used in analyses. For all maximum likelihood methods, an appropriate model was chosen using ModelTest 3.7 (Posada and Crandall 1998), based on AIC: GTR+I+ $\Gamma$ . Genealogies were constructed based on first and second codon positions using maximum likelihood in PAUP\* (Swofford 2002). Support was assessed with 100 bootstrap replicates. MCMC analyses, using all codon sites, were performed in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). Each analysis was run  $10^7$  generations, with the first  $2.5 \times 10^6$  generations discarded as burn-in. Beta-tubulin genealogies based on amino acid sequences were constructed using the neighbor-joining algorithm in PAUP\* with distances calculated in TreePuzzle 5.2 (Schmidt et al. 2002). Support was assessed using quartet puzzling in PAUP\*. Amino acid alignments were also analyzed in MrBayes, where we allowed model jumping between fixed-rate amino acid models and ran the analyses for  $10^7$  generations and discarded the first  $2.5 \times 10^6$  generations as burn-in.

Additional genealogies were constructed using all  $\beta$ -tubulin and  $\beta$ -tubulin-like sequences from *Tetrahymena thermophila* and *Paramecium tetraurelia* genomes (*Tetrahymena*: Ttherm\_00348510 (BTU1), Ttherm\_00836580 (BTU2), Ttherm\_01104960 (BLT1), Ttherm\_01104970 (BLT2), Ttherm\_01104980 (BLT3), Ttherm\_01120580 (BLT4), Ttherm\_01620840 (BLT5), Ttherm\_00834960 (BLT6); *Paramecium*: GSPATG00021361001

Table 1. Species sampled for genealogical analyses.

Species	Lineage	$\beta$ -tubulin	SSU rDNA
<i>Babesia bovis</i>	Apicomplexa	L00978	L31922
<i>Eimeria tenella</i>	Apicomplexa	U19609	U40264
<i>Plasmodium berghei</i>	Apicomplexa	M58750	M14599
<i>Plasmodium falciparum</i>	Apicomplexa	X16075	M19172
<i>Toxoplasma gondii</i>	Apicomplexa	M20025	L37415
<i>Chilodonella uncinata</i>	Ciliate	DQ665918, DQ665919, DQ665920, DQ665921	AF300281
<i>Colpoda</i> sp.	Ciliate	X94349	AY905498
<i>Euplotes crassus</i>	Ciliate	J04534	AY007440
<i>Euplotes focardii</i>	Ciliate	S72098	
<i>Euplotes octocarinatus</i>	Ciliate	X69467	AJ310489
<i>Gastrostyla steinei</i>	Ciliate	AF510204	AF164133
<i>Hypotrichida</i> sp.	Ciliate	AF510210	
<i>Metopus es</i>	Ciliate	DQ665922	
<i>Metopus palaeformis</i>	Ciliate		M86385
<i>Nyctotherus ovalis</i>	Ciliate	DQ665923, DQ665924	AY007454
<i>Oxytricha granulifera</i>	Ciliate	AF510205	AF164122
<i>Oxytricha longa</i>	Ciliate	AF510206	AF164125
<i>Oxytricha trifallax/Sterkiella histriomuscorum</i>	Ciliate	AF510207	AF164121
<i>Paramecium caudatum</i>	Ciliate	AB070222	AF217655
<i>Paramecium tetraurelia</i>	Ciliate	AJ608919, X67237	X03772
<i>Stylonychia lemnae</i>	Ciliate	AF510208, X06874	AF164124
<i>Stylonychia mytilus</i>	Ciliate	AF188162	AF164123
<i>Tetrahymena pyriformis</i>	Ciliate	X12769, X12768	X56171
<i>Tetrahymena thermophila</i>	Ciliate	L01415, L01416	X56165
<i>Uroleptus gallina</i>	Ciliate	AF510209	AF164130
<i>Amphidinium corpulentum</i>	Dinoflagellate	AF482405	AF274252
<i>Cryptothecodinium cohnii</i>	Dinoflagellate	AF421537	DQ241737
<i>Dinophyceae</i> sp.	Dinoflagellate	AF482424	
<i>Gyrodinium striatum</i>	Dinoflagellate	AF482408	DQ084522
<i>Heterocapsa triquetra</i>	Dinoflagellate	AF482413, AF482414	AJ415514
<i>Kryptoperidinium foliaceum</i>	Dinoflagellate	AY713392	AF274268
<i>Peridinium willei</i>	Dinoflagellate	AF482421	AF274280
<i>Wolozynskia tenuissima</i>	Dinoflagellate	AF482422	

Lineage refers to major alveolate clade. GenBank Accession numbers are listed for each gene.

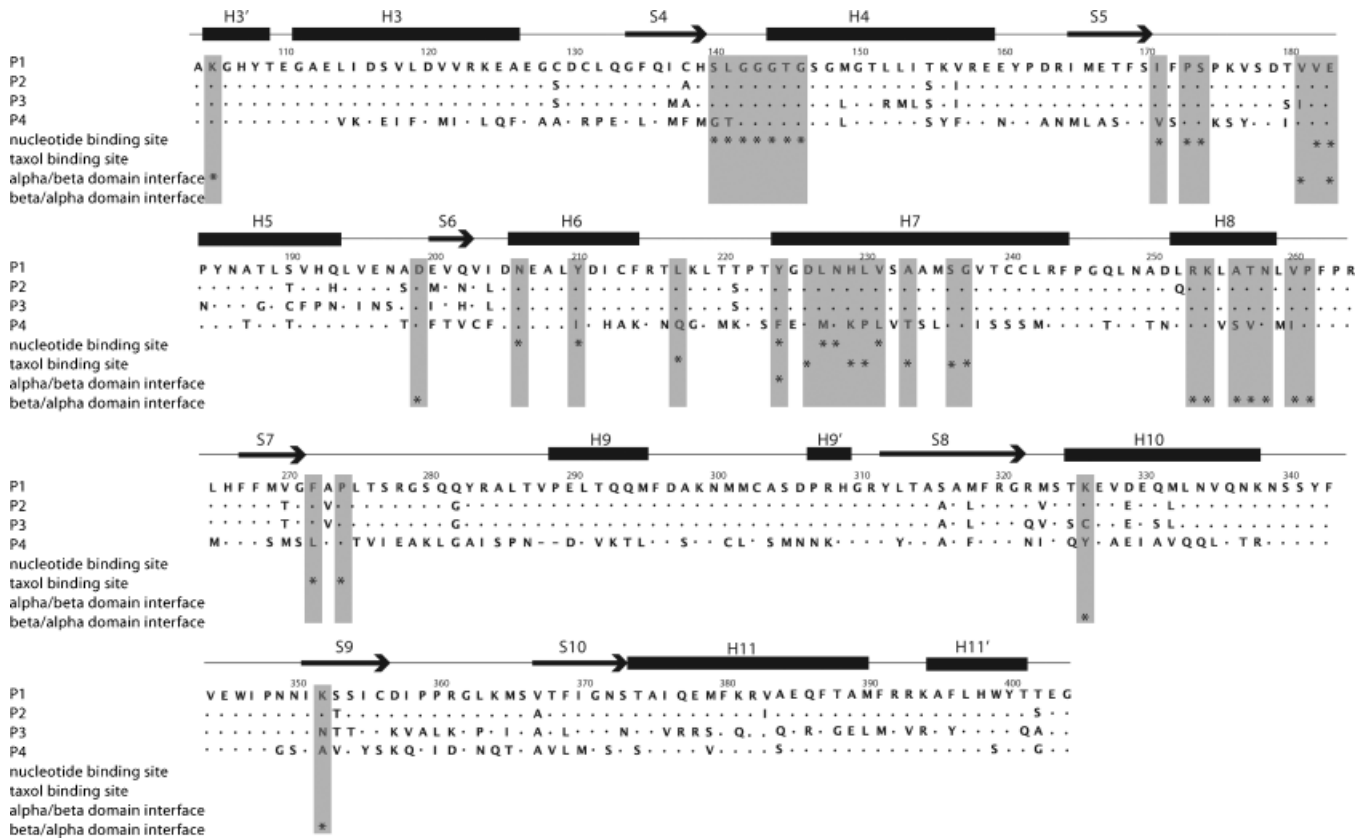


Fig. 1. Alignment of sequenced regions of *Chilonella uncinata*  $\beta$ -tubulin paralogs (P1–P4). A dot (.) indicates a match to P1. Beta-tubulin structure is mapped above alignments (Lowe et al. 2001). Functional domains, based on annotation in the conserved domain database (Marchler-Bauer et al. 2005), are mapped below sequences and highlighted.

(tub\_betaPT1), GSPATG00036831001 (tub\_betaPT2), GSPATG00035237001 (tub\_betaPT3), GSPATG00035238001 (tub\_betaPT3), GSPATG00032790001 (tub\_betaPT4), GSPATG00027639001 (tub\_betaPT5), GSPATG00024605001 (tub\_betaPT6), GSPATG00012209001 (tub\_betaPT7), GSPATG00007518001 (tub\_betaPT8) (Aury et al. 2006; Eisen et al. 2006)). The two sequences representing tub\_betaPT3 from *Paramecium* were not included in these analyses due to their short size (<60% length of the other sequences). Genealogies were reconstructed in MrBayes, using a mixed model for amino acid evolution; the analysis was run for  $10^7$  generations with the first  $2.5 \times 10^6$  generations discarded as burn-in.

We also performed phylogenetic analyses of these taxa based on small subunit (SSU)-rDNA genes for all species included in the  $\beta$ -tubulin analysis except: (a) the following species are not included in the SSU-rDNA analysis: *Euplotes focardii* and *Woloszynskia tenuissima*, for which there were no available SSU-rDNA sequences on GenBank and unidentified species of Hypotrichida and Dinophyceae, and (b) instead of *Metopus es*, *Metopus palaeformis* was included in the SSU-rDNA analysis (Table 1). SSU-rDNA sequences were aligned in Clustal W (Thompson et al. 1994) and adjusted by hand. All ambiguous regions of the resulting alignment were removed from further analyses, leaving 1,486 characters. Models of sequence evolution were assessed in ModelTest (Posada and Crandall 1998) based on AIC and the GTR+I+ $\Gamma$  model was used. Maximum likelihood analyses were performed in PAUP\*, with 100 bootstrap replicates. MCMC analyses were run in MrBayes (Ronquist and Huelsenbeck 2003) for  $5 \times 10^6$  generations with a burn-in of  $1.25 \times 10^6$  generations.

Table 2. Diversity and composition of  $\beta$ -tubulin paralogs from *Chilonella uncinata*.

Paralog	Average distance	%GC	$N_c$
P1	0.19	53.2	29.34
P2	0.21	51.9	41.07
P3	0.28	52.9	44.64
P4	0.45	45.0	52.39

Average distance is the average pairwise nucleotide distance between each paralog and all other ciliate  $\beta$ -tubulins. %GC indicates the GC content in the coding portion of the gene.  $N_c$  is a measure of the effective number of codons. Low GC content and high  $N_c$  are indicative of relaxed functional constraint or positive selection. GC content and  $N_c$  were calculated in codonW version 1.4.2 (<http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html>).

## RESULTS

**Macronuclear sequences.** Comparative analyses of macronuclear sequences reveal high levels of diversity among  $\beta$ -tubulin paralogs in *C. uncinata* (Fig. 1). The average pairwise nucleotide distance between each paralog and all other sequenced ciliate  $\beta$ -tubulins (in Table 1) ranges from 0.19 to 0.45 (Table 2; as calculated in PAUP\* (Swofford 2002)). These distances are correlated with differences in GC content and effective number of codons (a codon usage index) (Table 2).

Mapping differences among paralogs onto protein structure reveals no simple relationship between amino acid substitution

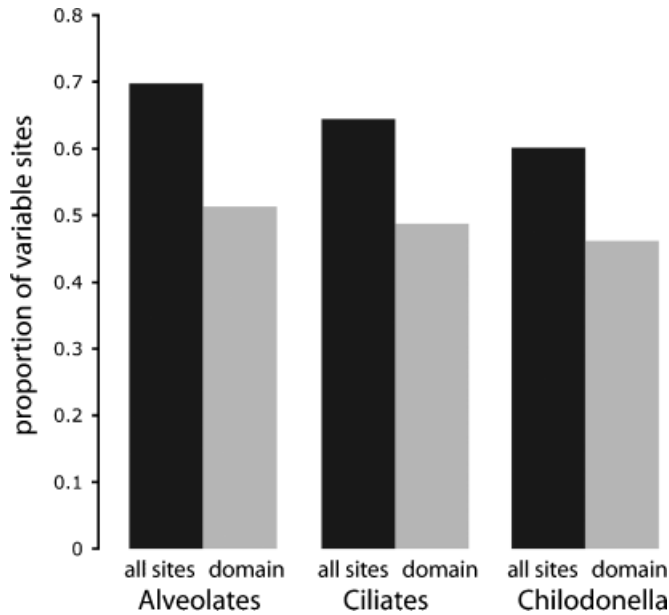


Fig. 2. Proportion of amino acid sites in  $\beta$ -tubulin that are variable among paralogs and across taxa. Black bars indicate the proportion of variable sites across the whole length of the protein examined; gray bars indicate the fraction of variable sites within identified functional domains (Fig. 1). Comparisons are shown at three phylogenetic scales: across alveolates, across ciliates, and among paralogs in *Chilodonella uncinata*. In every case, there is reduced variation at the functional domains suggesting that these regions are subject to purifying selection due to functional constraint.

and structural features (Fig. 1). Amino acid substitutions among paralogs are spread throughout the gene, instead of being concentrated to specific structural regions of the protein. However, if we consider amino acid sites that are part of known functional domains, we find fewer substitutions at these sites than across the

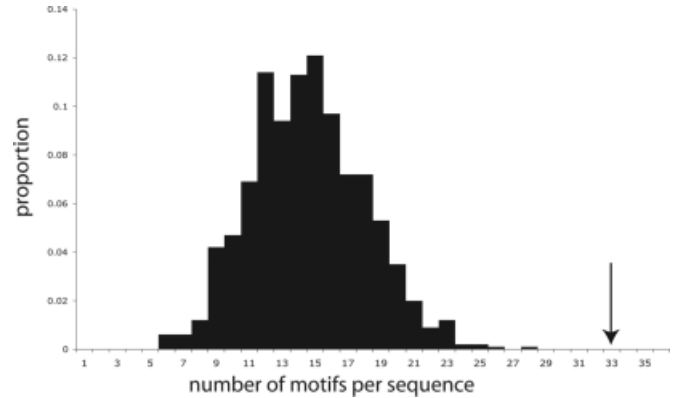


Fig. 4. Random distribution of G(A/T)TT versus observed distribution in internally eliminated sequences (IESs) of *Chilodonella uncinata*. One thousand random sequences were generated with the same nucleotide composition and the combined length of all nine IESs ( $\alpha$ -tubulin,  $\beta$ -tubulin, and actin). The number of occurrences of G(A/T)TT/AA(T/A)C counted in these random sequences are shown in the histogram and compared with the number of this motif found in all IESs (shown as black arrow; 33 occurrences in nine IESs). These simulations demonstrate that this sequence motif is found more often in IESs than expected based on the underlying nucleotide composition.

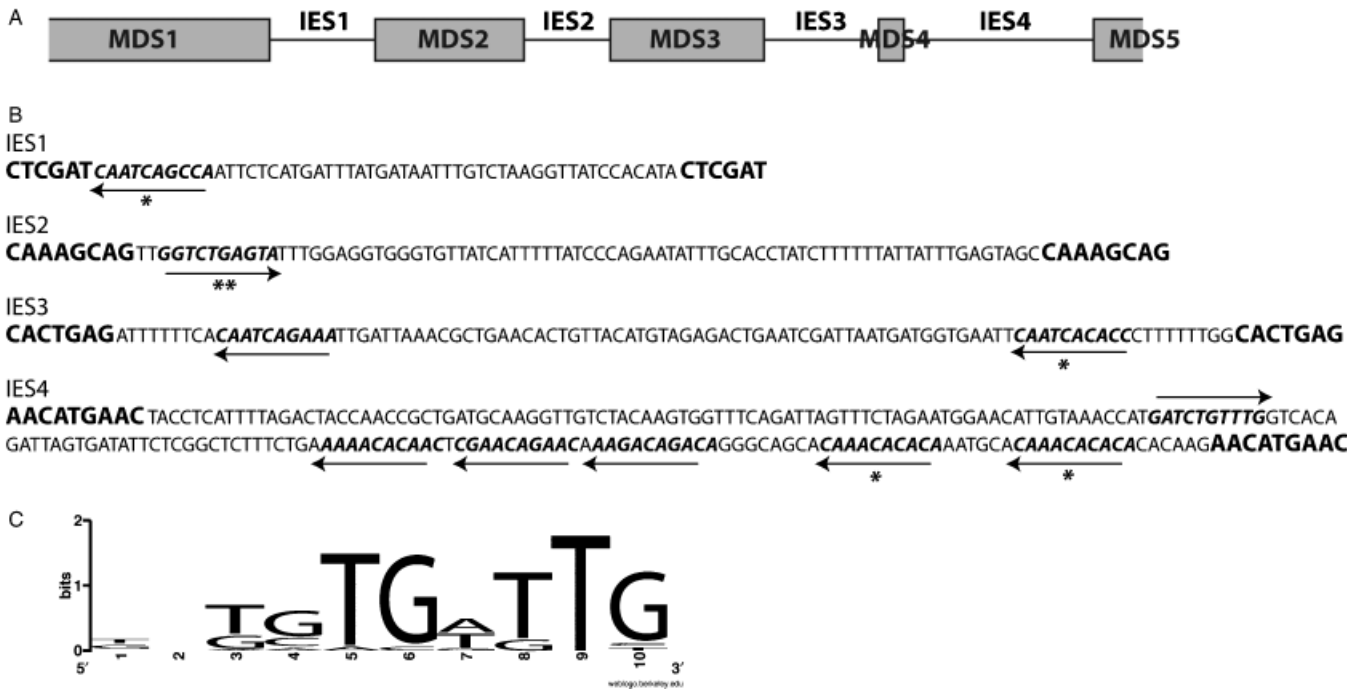


Fig. 3. Micronuclear  $\beta$ -tubulin. (A) Gene structure of micronuclear  $\beta$ -tubulin paralog of *Chilodonella uncinata*. Five macronuclear destined sequences (MDS) are separated by four internally eliminated sequences (IES). (B) Sequence of IESs. Direct repeats are in bold at the ends of the sequences. Locations and orientations of conserved motifs are indicated by arrows. \* Motif identified both by assuming one motif per sequence and any number of motifs per sequence. \*\* Motif identified only when assuming one motif in each sequence. (C) Internally eliminated sequences conserved motif sequence from  $\beta$ -tubulin,  $\alpha$ -tubulin, and actin IESs as represented by WebLogo (Crooks et al. 2004; Schneider and Stephens 1990). This includes all conserved sequences identified under both sets of assumptions.

Table 3. Support for monophyly of major clades based on small subunit (SSU)-rDNA and  $\beta$ -tubulin sequences.

	SSU		$\beta$ -tubulin			
	ML	Bayes	NT		AA	
			ML	Bayes	NJ	Bayes
<b>Ciliates</b>	●	●	X	X	X	X
Oligohymenophorea	●	●	●	●	●	○
Spirotrichea	●	●	X	X	X	X
Armophorea	●	●	X	X	X	X
<b>Apicomplexa</b>	●	●	X	X	○	●
<b>Dinoflagellates</b>	●	●	●	●	●	●

ML, Bayes, and NJ indicate the type of phylogenetic analysis performed: maximum likelihood, Bayesian, and neighbor joining, respectively. NT and AA indicate results of analyses of  $\beta$ -tubulin based on nucleotide or amino acid sequences, respectively. X indicates that a clade is not monophyletic in an analysis. An open circle indicates <50% support; a shaded circle 50–80% support; and a dark circle >80% support for a clade, where support refers to either bootstrap support, Bayesian posterior probabilities, or quartet puzzling.

rest of the protein (Fig. 1, 2). In addition to these  $\beta$ -tubulin paralogs, partial sequence data generated in this study (data not shown) and preliminary data from other strains of *C. uncinata* suggest the presence of additional paralogs (Robinson, Zufall, and Katz, unpubl. data).

**Micronuclear sequences.** We found IESs in one micronuclear  $\beta$ -tubulin paralog (GenBank Accession No. EF205128). This sequence is most similar to paralog P4; however, it contains a 3-bp insertion and differs at 0.91% (nine of 987) of nucleotides in the macronuclear-destined sequences in comparison with the previ-

ously published  $\beta$ -tubulin P4. These differences result in a single amino acid insertion/deletion and two amino acid substitutions. In our analyses of high molecular weight DNA, we also find this variant allele without IESs, corresponding to the macronuclear version. These differences may be due to allelic variation that was previously unsampled from the macronucleus, or a second micronuclear copy that does not contain IESs.

There are four IESs in paralog P4 ranging in size from 58 to 212 bp (Fig. 3). AT content of IESs ranges from 60% to 67%, in contrast to 47%–55% in macronuclear-destined sequences. Like

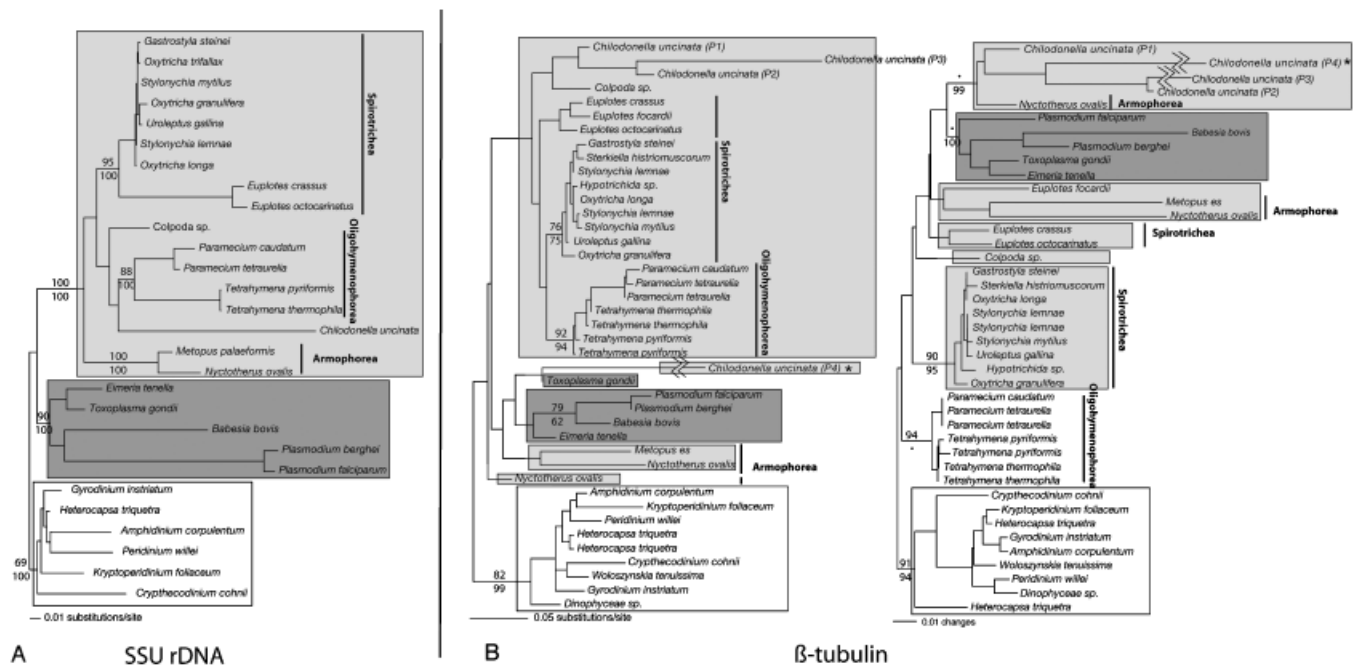


Fig. 5. Phylogenies of alveolates based on small subunit (SSU)-rDNA and  $\beta$ -tubulin. Ciliates are boxed in light gray, apicomplexans in dark gray, and dinoflagellates in white. (A) Phylogenetic reconstruction based on SSU-rDNA. The maximum likelihood topology with support at key nodes shown as maximum likelihood bootstrap support (top) and Bayesian posterior probabilities (bottom). (B) Phylogenetic reconstructions based on  $\beta$ -tubulin nucleotide sequences (left tree) and amino acid sequences (right tree). The nucleotide tree topology is based on maximum likelihood analysis with support shown as maximum likelihood bootstrap (top) and Bayesian posterior probability (bottom). The amino acid tree topology is based on neighbor-joining analysis with support shown as quartet puzzling (top) and posterior probability (bottom). All trees are shown rooted with dinoflagellates. \* Support of less than 50%. \* The paralog similar to that for which we have characterized the micronuclear copy (P4).

$\alpha$ -tubulin in this species (Katz et al. 2003), the shorter IESs are found at the 5'-end of the gene and the longer IESs at the 3'-end. These IESs are flanked by direct repeats (Fig. 3B). The direct repeats differ between IESs and range from 6 to 9 bp. One repeat is presumably eliminated during excision; the other remains in the macronuclear-destined sequence to complete the coding region.

We identified conserved sequence motifs within IESs under two different sets of assumptions about the distribution of regulatory motifs (Fig. 3). In each case, a site in a sequence is considered significant if the probability is less than 0.0001 of a random string with the same background base composition matching the consensus as well or better than the motif (Bailey and Gribskov 1998). If we assume that each IES has a single occurrence of a regulatory motif and analyze  $\beta$ -tubulin IESs jointly with IESs from  $\alpha$ -tubulin and actin, the consensus sequence of that motif is 5'-GGTGATTG-3' ( $E$ -value 74 vs.  $1.6 \times 10^3$  with shuffled characters; Fig. 3). If we instead suppose that there can be any number of motifs per sequence (including zero), a significantly conserved motif is found in three of four  $\beta$ -tubulin IESs and two of three  $\alpha$ -tubulin IESs with the consensus sequence: 5'-TTTGTGTTTG-3' ( $E$ -value  $2.5 \times 10^{-2}$  vs.  $3.0 \times 10^3$  with shuffled characters; Fig. 3). The same motifs are found when  $\beta$ -tubulin IESs are analyzed separately. Similar motifs are recovered using other algorithms, including the Gibbs Motif Sampler (Thompson, Rouchka, and Lawrence 2003; Thompson et al. 2004). In addition, these motifs are not significantly conserved in macronuclear-destined sequences. All of these motifs share a conserved 4-bp core, G(A/T)TT, which is present, and overrepresented, in all *C. uncinata* IESs examined except  $\beta$ -tubulin IES2 (Fig. 4).

We were unable to identify micronuclear-limited sequences within the remaining three paralogs of  $\beta$ -tubulin. This is either because (1) there are no IESs within these paralogs or (2) the IESs are long or otherwise difficult to characterize with our PCR-based approach.

**Genealogical analysis.** We found contrasting patterns between genealogical analyses of SSU-rDNA and  $\beta$ -tubulin. Analysis of SSU-rDNA reveals strong support for the monophyly of ciliates, three classes within Ciliates, Apicomplexa, and Dinoflagellates, consistent with previous analyses (e.g. Riley and Katz 2001; Van de Peer et al. 2000) (Table 3 and Fig. 5A). In contrast, phylogenetic analyses of  $\beta$ -tubulin sequences reveal unstable topologies with no support for the monophyly of ciliates, nor two of three classes within ciliates (Table 3 and Fig. 5B). Our analyses of  $\beta$ -tubulin also result in paraphyletic Apicomplexa, which is contrary to the results of previous studies of this gene that include a smaller sampling of ciliates (Leander, Clopton, and Keeling 2003; Saldarriaga et al. 2003). Under all of the evolutionary models and algorithms that we tried, the ciliate  $\beta$ -tubulins are not monophyletic with respect to the other alveolates.

To test whether these results were due to the comparison of paralogous, rather than orthologous, genes we repeated the genealogical analyses including all  $\beta$ -tubulin and  $\beta$ -tubulin-like sequences from the *Tetrahymena* and *Paramecium* genomes. We find that the *Tetrahymena* and *Paramecium*  $\beta$ -tubulin-like genes form a clade together with *C. uncinata* P4 (Fig. 6). The  $\beta$ -tubulin-like sequences and *C. uncinata* P4 are all on long branches. Therefore, this clustering could be due to either long-branch attraction or ancient duplication; however, we are unable to

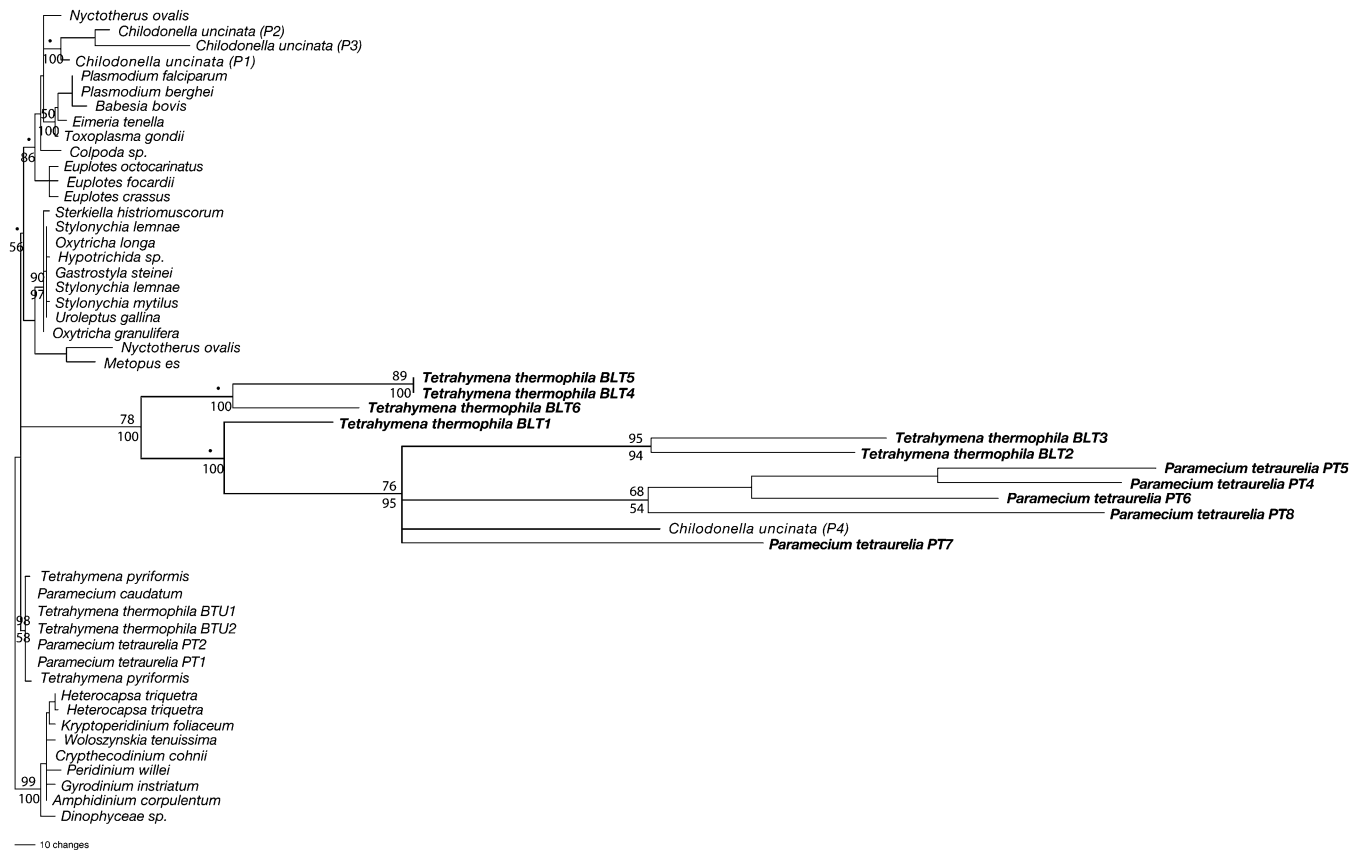


Fig. 6. Genealogy of  $\beta$ -tubulin amino acid sequences demonstrates long branches of *Chilodonella uncinata* P4 and  $\beta$ -tubulin-like sequences from *Tetrahymena thermophila* and *Paramecium tetraurelia*. Sequences in bold are those not used in analyses shown in Fig. 5. Support is shown for quartet puzzling (performed in PAUP\*; top number) and posterior probability (bottom number) at relevant nodes only. • Support of less than 50%.

distinguish between these two possibilities. Based on these results, we removed P4 and the  $\beta$ -tubulin-like genes and repeated the genealogical analyses. Removal of P4 had little effect on the resultant genealogies: ciliates remained polyphyletic (data not shown).

## DISCUSSION

In this study, we characterized micronuclear and macronuclear paralogs of  $\beta$ -tubulin in *C. uncinata*, a ciliate with extensive developmental genome processing. Our results have implications for both models of protein evolution in ciliates and inferences about the evolution of genome processing.

We identified four IESs within one  $\beta$ -tubulin paralog, each of which contains at least one conserved sequence motif shared with IESs in  $\alpha$ -tubulin and actin from the same species. Because this motif is present in the IESs of multiple genes, it may be important in the regulation of IES excision during genome processing. Alternatively, this shared sequence motif may be evidence of a common origin of these IESs, or may result from compositional biases in non-coding regions of the genome. The motif derived from this combined analysis is different, but overlapping with motifs identified in analyses of  $\alpha$ -tubulin IESs alone (Katz et al. 2003).

It is possible that there are numerous conserved *cis* signals within IESs, some of which are found within a gene and some of which are shared between genes (Coyne, Chalker, and Yao 1996). Under this model, we would expect to find many motifs, each shared by just a few IESs. Alternatively, it is possible that epigenetic mechanisms similar to those found in *Paramecium* and *Tetrahymena* (Betermier 2004; Chalker, Fuller, and Yao 2005; Mochizuki and Gorovsky 2004) compensate for the absence of *cis* signals in some IESs. For example, we demonstrate the overrepresentation of the 4-bp sequence G(A/T)TT in IES sequences (Fig. 4), supporting a role for this motif in regulation of IES excision. However, among nine IESs one does not contain this sequence, suggesting that other mechanisms are also likely to be involved in IES excision.

Our comparison of SSU-rDNA and  $\beta$ -tubulin genealogies from parallel taxon samples suggests that the rapid and heterogeneous rates of protein evolution in ciliates may have eliminated phylogenetic signal in protein-coding genes. Specifically, we tested the hypothesis that the heterogeneous rates of evolution in a ciliate protein-coding gene lead to inconsistent and unstable tree topologies. By comparing phylogenies with parallel taxon samples, we demonstrate the lack of phylogenetic signal in ciliate  $\beta$ -tubulin sequences compared with SSU-rDNA. Previous studies have shown inconsistent/discordant genealogies of ciliates based on protein-coding genes, including  $\beta$ -tubulin (Fast et al. 2002). The inability of this protein to recover phylogenetic relationships among ciliate species is consistent with our previous demonstration that the unique genome architecture of ciliates allows for increased rates of protein evolution (McGrath et al. 2006; Zufall et al. 2006).

## ACKNOWLEDGMENTS

We are grateful for the thoughtful comments and advice of several anonymous reviewers. This work was supported by grants from the National Science Foundation to R. A. Z. (DBI 0301610 and MCB 0625272) and L. A. K. (DEB 0092908 and DEB 043115).

## LITERATURE CITED

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut,

- A., Beisson, J., Blanc, I., Bouhouche, K., Camara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouel, A., Lepere, G., Malinsky, S., Nowacki, M., Nowak, J. K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Betermier, M., Weissenbach, J., Scarpelli, C., Schachter, V., Sperling, L., Meyer, E., Cohen, J. & Wincker, P. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*, **444**:171–178.
- Bailey, T. L. & Gribskov, M. 1998. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**:48–54.
- Betermier, M. 2004. Large-scale genome remodeling by the developmentally programmed elimination of germ line sequences in the ciliate *Paramecium*. *Res. Microbiol.*, **155**:399–408.
- Chalker, D. L., La Terza, A., Wilson, A., Kroenke, C. D. & Yao, M. C. 1999. Flanking regulatory sequences of the *Tetrahymena* R deletion element determine the boundaries of DNA rearrangement. *Mol. Cell Biol.*, **19**:5631–5641.
- Chalker, D. L., Fuller, P. & Yao, M. C. 2005. Communication between parental and developing genomes during *Tetrahymena* nuclear differentiation is likely mediated by homologous RNAs. *Genetics*, **169**:149–160.
- Coyne, R. S., Chalker, D. L. & Yao, M.-C. 1996. Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. *Annu. Rev. Genet.*, **30**:557–578.
- Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. 2004. WebLogo: a sequence logo generator. *Genome Res.*, **14**:1188–1190.
- Eisen, J. A., Coyne, R. S., Wu, M., Wu, D. Y., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q. H., Amedeo, P., Jones, K. M., Tallon, L. J., Delcher, A. L., Salzberg, S. L., Silva, J. C., Haas, B. J., Majoros, W. H., Farzad, M., Carlton, J. M., Smith, R. K., Garg, J., Pearlman, R. E., Karrer, K. M., Sun, L., Manning, G., Elde, N. C., Turkewitz, A. P., Asai, D. J., Wilkes, D. E., Wang, Y. F., Cai, H., Collins, K., Stewart, A., Lee, S. R., Wilamowska, K., Weinberg, Z., Ruzzo, W. L., Wloga, D., Gertig, J., Frankel, J., Tsao, C. C., Gorovsky, M. A., Keeling, P. J., Waller, R. F., Paton, N. J., Cherry, J. M., Stover, N. A., Krieger, C. J., del Toro, C., Ryder, H. F., Williamson, S. C., Barbeau, R. A., Hamilton, E. P. & Orias, E. 2006. Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.*, **4**:1620–1642.
- Fast, N. M., Xue, L. R., Bingham, S. & Keeling, P. J. 2002. Re-examining alveolate evolution using multiple protein molecular phylogenies. *J. Eukaryot. Microbiol.*, **49**:30–37.
- Jahn, C. L. & Klobutcher, L. A. 2002. Genome remodeling in ciliated protozoa. *Annu. Rev. Microbiol.*, **56**:489–520.
- Katz, L. A. 2001. Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. *Int. J. Syst. Evol. Microbiol.*, **51**:1587–1592.
- Katz, L. A., Lasek-Nesselquist, E. & Snoeyenbos-West, O. L. O. 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.
- Katz, L. A., Bornstein, J., Lasek-Nesselquist, E. & Muse, S. V. 2004. Dramatic diversity of ciliate histone H4 genes revealed by comparisons of patterns of substitutions and paralog divergences among eukaryotes. *Mol. Biol. Evol.*, **21**:555–562.
- Klobutcher, L. A. & Herrick, G. 1997. Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.*, **56**:1–62.
- Ku, M., Mayer, K. & Forney, J. D. 2000. Developmentally regulated excision of a 28-base-pair sequence from the *Paramecium* genome requires flanking DNA. *Mol. Cell Biol.*, **20**:8390–8396.
- Leander, B. S., Clopton, R. & Keeling, P. 2003. Phylogeny of gregarines (Apicomplexa) as inferred from small-subunit rDNA and beta-tubulin. *Int. J. Syst. Evol. Microbiol.*, **53**:345–354.
- Lowe, J., Li, H., Downing, K. H. & Nogales, E. 2001. Refined structure of alpha beta-tubulin at 3.5 Å resolution. *J. Mol. Biol.*, **313**:1045–1057.
- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWweese-Scott, C., Geer, L. Y., Gwadz, M., He, S. Q., Hurwitz, D. I., Jackson, J. D., Ke, Z. X., Lanczycki, C. J., Liebert, C. A., Liu, C. L., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D. C. & Bryant, S. H. 2005. CDD: a conserved domain database for protein classification. *Nucleic Acids Res.*, **33**:D192–D196.

- McGrath, C. L., Zufall, R. A. & Katz, L. A. 2006. Genome evolution in ciliates. In: Katz, L. A. & Bhattacharya, D. (ed.), *Genomics and Evolution of Microbial Eukaryotes*. Oxford University Press, New York. p. 64–77.
- McGrath, C. L., Zufall, R. A. & Katz, L. A. 2007. Variation in macronuclear genome content of three ciliates with extensive chromosomal fragmentation: a preliminary analysis. *J. Eukaryot. Microbiol.*, **53**, (in press).
- Mochizuki, K. & Gorovsky, M. A. 2004. Small RNAs in genome rearrangement in *Tetrahymena*. *Curr. Opin. Genet. Dev.*, **14**:181–187.
- Posada, D. & Crandall, K. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**:817–818.
- Prescott, D. M. 1994. The DNA of ciliated protozoa. *Microbiol. Rev.*, **58**:233–267.
- Riley, J. L. & Katz, L. A. 2001. Widespread distribution of extensive genome fragmentation in ciliates. *Mol. Biol. Evol.*, **18**:1372–1377.
- Ronquist, F. & Huelsenbeck, J. P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**:1572–1574.
- Saldarriaga, J. F., McEwan, M. L., Fast, N. M., Taylor, F. J. R. & Keeling, P. J. 2003. Multiple protein phylogenies show that *Oxyrrhis marina* and *Perkinsus marinus* are early branches of the dinoflagellate lineage. *Int. J. Syst. Evol. Microbiol.*, **53**:355–365.
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, **18**:502–504.
- Schneider, T. D. & Stephens, R. M. 1990. Sequence logos—a new way to display consensus sequences. *Nucleic Acids Res.*, **18**:6097–6100.
- Seegmiller, A. & Herrick, G. 1998. A short internal eliminated sequence with central conserved sequences interrupting the LA-MSC gene of the 81 locus in the hypotrichous ciliates *Oxytricha fallax* and *O. trifallax*. *J. Eukaryot. Microbiol.*, **45**:55–58.
- Swofford, D. 2002. PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods). Sinauer Associates, Sunderland, MA.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**:4673–4680.
- Thompson, W., Rouchka, E. C. & Lawrence, C. E. 2003. Gibbs recursive sampler: finding transcription factor binding sites. *Nucleic Acids Res.*, **31**:3580–3585.
- Thompson, W., Palumbo, M. J., Wasserman, W. W., Liu, J. S. & Lawrence, C. E. 2004. Decoding human regulatory circuits. *Genome Res.*, **14**:1967–1974.
- Van de Peer, Y., Baldauf, S. L., Doolittle, W. F. & Meyer, A. 2000. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. *J. Mol. Evol.*, **51**: 565–76.
- Yao, M. C., Duharcourt, S. & Chalker, D. L. 2002. Genome-wide rearrangements of DNA in ciliates. In: Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. (ed.), *Mobile DNA II*. ASM Press, Washington DC. p. 730–758.
- Zufall, R. A., McGrath, C. L., Muse, S. V. & Katz, L. A. 2006. Genome architecture drives protein evolution in ciliates. *Mol. Biol. Evol.*, **23**:1681–1687.

Received: 08/28/06, 01/02/07, 02/21/07; accepted: 03/07/07