

Published in final edited form as:

J Mol Evol. 2011 December ; 73(5-6): 266–272. doi:10.1007/s00239-011-9468-x.

Heterogeneous rates of molecular evolution among cryptic species of the ciliate morphospecies *Chilodonella uncinata*

Laura A. Katz^{1,2}, Jennifer DeBerardinis¹, Meaghan S. Hall¹, Alexandra M. Kovner¹, Micah Dunthorn², and Spencer V. Muse³

¹Department of Biological Sciences, Smith College, Northampton MA 01063, USA

²Program in Organismic and Evolutionary Biology, UMass-Amherst, Amherst MA 01003, USA

³Department of Statistics, North Carolina State University, Raleigh NC 27695-7566, USA

Abstract

While molecular analyses have provided insight into the phylogeny of ciliates, the few studies assessing intraspecific variation have largely relied on just a single locus (e.g. nuclear small subunit rDNA (nSSU-rDNA) or mitochondrial cytochrome oxidase I). In this study, we characterize the diversity of several nuclear protein-coding genes plus both nSSU-rDNA and mitochondrial small subunit rDNA (mtSSU-rDNA) of five isolates of the ciliate morphospecies *Chilodonella uncinata*. Although these isolates have nearly identical nSSU-rDNA sequences, they differ by up to 8.0% in mtSSU-rDNA. Comparative analyses of all loci, including β -tubulin paralogs, indicate a lack of recombination between strains, demonstrating that the morphospecies *C. uncinata* consists of multiple cryptic species. Further, there is considerable variation in substitution rates among loci as some protein-coding domains are nearly identical between isolates while others differ by up to 13.2% at the amino acid level. Combining insights on macronuclear variation among isolates, the focus of this study, with published data from the micronucleus of two of these isolates indicates that *C. uncinata* lineages are able to maintain both highly divergent and highly conserved genes within a rapidly evolving germline genome.

Keywords

Cryptic species; molecular evolution; genome architecture; ciliate

Introduction

Patterns of molecular evolution among ciliates are known to be shaped by their unusual genome architecture, namely the presence of both germline and somatic nuclei within each cell (McGrath et al. 2006; Zufall et al. 2006). To date, most studies of intraspecific genetic variation in ciliates typically rely on just a single locus, with the notable exception of the model ciliates *Paramecium tetraurelia* and *Tetrahymena thermophila* (e.g. Catania et al. 2009; Simon et al. 2008). Yet analyses of several genes from the somatic genome, coupled with data from the germline genome, are needed to assess intraspecific patterns of molecular evolution in ciliates. In light of previous data suggesting both high and low levels of gene conservation in two geographically isolated strains of *C. uncinata* (Robinson and Katz 2008) and high divergence among strains in the germline genome (Katz and Kovner 2010), we aimed to determine the level of intraspecific variation in protein-coding loci of additional

strains. This allowed us to test the hypotheses that the somatic macronucleus is marked by heterogeneous patterns of molecular evolution and that *C. uncinata* is composed of multiple genetically distinct cryptic species.

The genetics of ciliates must be interpreted in light of the presence of both a “germline” micronucleus and “somatic” macronucleus within each cell. The smaller micronucleus, which is not transcribed, resembles a typical eukaryotic genome in that it has few large chromosomes. The transcriptionally active macronucleus contains chromosomes modified from a zygotic nucleus (Juraneck and Lipps 2007; Zufall et al. 2006). Macronuclear chromosomes are generated through extensive processing that includes amplification, elimination of internal excised sequences, and amplification (McGrath et al. 2006). Extensive processing in ciliates—Spirotrichea, Armophorea, and Phyllopharyngea (including *Chilodonella uncinata*, the focus of this study)—produces approximately 20,000,000 gene-sized chromosomes (i.e. each macronuclear chromosome contains only a single gene) in the macronucleus (Katz and Kovner 2010; Zufall et al. 2006).

Maintenance of these two dimorphic genomes allows ciliates to explore protein space in novel ways compared to other eukaryotes (McGrath et al. 2006; Zufall et al. 2006). For example, ciliates have remarkably divergent histone H4 genes compared to other eukaryotes, with some lineages having multiple paralogs that differ by up to 20% of their amino acids (Katz et al. 2004). Moreover, ciliates with highly processed genomes have greater variation in protein-coding gene families across multiple loci than do ciliates that do not extensively process their somatic genomes (Zufall et al. 2006). These insights are based on comparisons between species and little is known about patterns of protein evolution within ciliate species.

The presence of cryptic species underlying conserved morphospecies – species identified by shared morphology – has been found across the ciliate tree of life. Cryptic species – defined here as morphospecies that contain multiple biological and/or genetic species – were discovered in *Tetrahymena* and *Paramecium* as strains within morphospecies that are reproductively isolated from one another (Nanney 1999; Sonneborn 1937; Sonneborn 1957). Some cryptic species have been shown to be ecologically distinct (e.g. differential prey and/or abiotic factors; Weisse and Lettner 2002; Weisse et al. 2008), indicating that elucidating cryptic diversity is important for studies focusing on topics such as food webs and species interactions.

Application of molecular tools has revealed extensive genetic variation underlying morphospecies (Barth et al. 2006; Gentekaki and Lynn 2009; Katz et al. 2005; Nanney et al. 1998; Simon et al. 2008). For example, individuals from multiple populations of both *Paramecium caudatum* and *Paramecium micronucleatum* have nearly identical internal transcribed spacer regions of the rDNA locus (i.e. low divergence within species for this locus), but reveal considerably more diversity based on mitochondrial encoded cytochrome oxidase 1 (Barth et al. 2006). There are a few examples of widely sampled ciliate morphospecies that appear to lack underlying genetic diversity; e.g., *Laboea strobila* and *Pelagostrobilidium neptuni* (Doherty et al. 2010; Katz et al. 2005).

Here we assess genetic data from five strains of the morphospecies *C. uncinata*: three recently isolated strains are morphologically identical (at least by light microscopy) to two strains, one from the American Type Culture Collection (USA-ATCC) and the other from Poland (POL), that were identified as *C. uncinata* by ciliate taxonomist Wilhelm Foissner, University of Salzburg (Robinson and Katz 2008). We were motivated by the observation that the strains USA-ATCC and Pol have nearly identical nSSU-rDNA, and actin sequences, and both conserved and divergent macronuclear β -tubulin sequences (Robinson and Katz 2008). Moreover, analyses of micronuclear sequences from these two strains reveal that

macronuclear-destined sequences (e.g. protein-coding domains) are fragmented (Zufall and Katz 2007) and embedded within rapidly evolving germline-specific sequences (Katz and Kovner 2010). To test whether the patterns of molecular evolution in macronuclear genes extend to other strains and loci, we analyzed macronuclear sequences of mtSSU-rDNA, nSSU-rDNA, actin, α -tubulin, and β -tubulin from all five isolates and assessed patterns of substitution across loci.

Methods

Ciliate culturing

Three new isolates of *C. uncinata* are from within the USA: USA-WH was acquired from Woods Hole, MA, and USA-SC1 and USA-SC2 were both collected from Lyman Pond on the Smith College campus in Northampton, MA (Table 1). The two isolates that had been previously characterized are both available at ATCC (USA = ATCC@50194, POL = ATCC@PRA-256), as reported in Robinson and Katz (2008). All isolates were cultured at room temperature in the dark, and maintained in filtered and autoclaved water from a local pond with a rice grain to feed bacterial populations. Clonal lines were generated for each isolate by passing a single cell through three rounds of isolation. Lines were transferred to fresh media every other week. To isolate DNA, cells were pelleted by spinning at 5,000 rpm for 20 minutes, and DNA was extracted using phenol-chloroform following Riley and Katz (2001).

Characterization of Mitochondrial and Nuclear Genes

All genes were amplified using Phusion Hot Start High Fidelity DNA Polymerase (Finnzymes F 540L). Nuclear SSU-rDNA (nSSU-rDNA), α -tubulin, β -tubulin, and actin were amplified using primers described previously (Robinson and Katz 2008) while mitochondrial SSU-rDNA (mtSSU-rDNA) was amplified using primers described in Dunthorn et al. (2011). Amplified products were cloned using Zero Blunt TOPO kits (Invitrogen 42-0245), and screened using the polymerase TaqGold. Clones were miniprepmed using Invitrogen's 96-well format harvested-cell method, protocol B. Several steps were taken to avoid contamination: only one strain was worked on at any one time; PCRs were set up in a cleaned hood; reported sequences were generally obtained from pooled PCR reactions; and we discarded any reactions for which a negative control turned up positive.

As we aimed to characterize paralogs from PCRs using degenerate primers, we sequenced multiple clones per reaction (~8 initially, with more added as paralogs were discovered) in just one direction and then fully sequenced at least two representatives of each unique sequence. Sequences were generated using the BigDye terminator RR mix from PE Applied Biosystems (Wellesley, MA, 4303152). Reactions were cleaned using gel filtration columns from Edge Biosystems (Gaithersburg, MD) and analyzed on a PerkinElmer ABI-3100 automated sequencer at the Center for Molecular Biology (Smith College, Northampton, MA). Additional sequencing reactions and sequencing were performed at the Penn State Genomics Core Facility (University Park, PA).

Sequence Analysis

Contigs were assembled in SeqMan (DNASar), and annotated using Seqbuilder (DNASar). MacClade (Maddison and Maddison 2005) and MegAlign (DNASar) were used to create alignments. Genealogies based on nucleotide alignments were estimated using PhyML (Guindon and Gascuel 2003) as implemented in Seaview v. 4.2.4 (Gouy et al. 2010), optimizing the number of invariant sites and using 6 rate classes. We analyzed codon bias by measuring the number of effective codons (N_c) (Wright 1990) using the program CodonW

(Peden 2005). For the β -tubulin gene family, we found several sequences that appeared to be PCR recombinants in that they occurred only in single PCR reactions. These recombinants were detected by eye and using GARD (Kosakovsky Pond et al. 2006), and the chimeric sequences were removed from subsequent analyses.

Synonymous and nonsynonymous substitution rates were estimated for the model of Muse and Gaut (1994) using HyPhy (Kosakovsky Pond et al. 2005). Estimates for the conserved paralogs SP1 and SP2 were compared to corresponding estimates for the divergent P3 paralogs across lines.

Results

Levels of conservation differ considerably between nSSU-rDNA and mtSSU-rDNA sequences among the five isolates. The isolates share nearly identical nSSU-rDNA sequences: the two haplotypes found (one shared by POL, USA-SC1, and USA-SC2; and the other shared by USA-ATCC and USA-WH) differ by only four single nucleotide polymorphisms (SNPs) across a 1 kb region (Table 3; Fig 1a). Unlike the nSSU-rDNA, the mtSSU-rDNA gene is highly divergent among the five isolates, with the sequence from USA-SC2 differing by up to 8.0% (nucleotide differences; Table 3, Fig 1b). POL and USA-SC1, which share an identical nSSU-rDNA locus, also have identical mtSSU-rDNA genes, and USA-ATCC and USA-WH are more similar to each other than either is to USA-SC2 (Fig. 1b).

Actin sequences are highly similar within all our five isolates of *C. uncinata*. We examined between three and six clones in each of the isolates characterized for this study (SC-1, SC-2 and WH) and found that each contained a single sequence differing from one another by less than 3% at the nucleotide level (Tables 2 and 3). The two lines POL and USA-SC1 share identical sequences and the two lines USA-ATCC and USA-WH are very similar (Fig. 1c). Actins from all five isolates are nearly identical at the amino acid level (Table 3). Again, the USA-SC2 sequence is the most divergent, differing from the other isolates by as much as 2.3% of its nucleotides and by three amino acid substitutions (Table 3).

Alpha-tubulin sequences are also highly conserved across all isolates. In each of USA-SC2, USA-SC1, and USA-WH, we identified only one sequence. These sequences are highly similar at the nucleotide level and identical at the amino acid level (Table 3). In a previous study (Israel et al. 2002), we identified two α -tubulin sequences in USA-ATCC. One of these two paralogs, found less frequently, was highly divergent in some regions from the conserved sequences analyzed in this study. Because we did not find any similar sequences in POL, USA-SC1, USA-SC2, or USA-WH, we did not include this previously published sequence in analyses here.

The β -tubulin sequences in our three new isolates are similar to those previously characterized by Robinson and Katz (2008). Each clonal line contains a subset of five β -tubulin paralogs: SP1, SP2, P1, P2, and P3 (Table 2; Fig 2), with the P3 paralogs being most divergent (up to 24% from the conserved SP1/SP2 amino acids sequences). Genealogical analyses indicate that all the haplotypes fall into several major clades and that the topology of each gene family member is consistent with that of the mtSSU-rDNA phylogeny (Fig. 2): USA-SC1 and POL share identical sequences, USA-ATCC and USA-WH are similar, and USA-SC2 is the most divergent at the nucleotide and amino acid levels (Fig 2).

Patterns of substitutions

We calculated the effective number of codons for protein-coding genes using codonW (Wright 1990). Codon bias is variable across protein-coding genes, with more highly

conserved haplotypes having a lower effective number of codons (Table 3). Actin, α -tubulin, and two β -tubulin paralogs (SP1 and SP2), use approximately 27–29 effective codons while the β -tubulin P3 paralogs use nearly double that with 52–54 codons.

To assess the patterns of substitutions further, we analyzed the ratios of dN/dS between the conserved SP1/SP2 β -tubulin paralogs and the more divergent P3 sequences. The clade-based estimate of dN/dS for the SP1/SP2 is 0 while that of the P3 clade is 0.050, suggesting a difference in selection intensity between the P3 and SP1/SP2 lineages. We also assessed patterns along individual branches of the β -tubulin topology, though this was not straightforward as not all paralogs were isolated from all lines (for example, we did not find a P3 sequence in USA-SC2). Although variances of the rate estimates are high, estimates of both synonymous and nonsynonymous rates for individual P3 branches are roughly an order of magnitude higher than those for their corresponding SP1 and SP2 branches (Table 4).

Discussion

Analyses of multiple molecular markers isolated from macronuclear and mitochondrial genomes lead to two complementary insights: 1) the morphospecies *C. uncinata* is comprised of multiple genetically distinct cryptic species; and 2) these species maintain both highly conserved and highly divergent loci in the macronucleus, while germline-restricted (micronuclear) sequences for two of the isolates have been shown to have so many substitutions that they are unalignable (Katz and Kovner 2010).

Cryptic species of the morphospecies *C. uncinata*

Several lines of evidence suggest that there are multiple cryptic species underlying the morphospecies *C. uncinata* including the high divergence among some loci between isolates and the lack of evidence of recombination within or between loci. Four of the five isolates of *C. uncinata* that we characterized are genetically distinct at mtSSU-rDNA and in protein-coding loci (e.g. the divergent β -tubulin P3 locus; Table 3, Figs 1–2). The level of divergence among isolates varies across loci, with the deepest divergence varying from 2.2% for actin to 8.0% for mtSSU to 13.2% for β -tubulin P3 (Table 3). Moreover, we see no evidence of recombination within loci or independent assortment between genes. The lack of recombination indicates that the four genetically distinct *C. uncinata* – USA-ATCC, USA-SC2, USA-WH and (POL + USA-SC1) – are reproductively isolated. Intriguingly, the most divergent line, USA-SC2, is sympatric with a second isolate, USA-SC1, which in turn is genetically identical to an isolate from Poland. While we took considerable care to eliminate contamination in PCR reactions, we recognize that the identity between the Pol and USA-SC1 lines may still be due to contamination of either molecular data or of the ciliate cultures. Hence, though more collections are needed to further assess biogeographical patterns, our data indicate that the geographic distribution of the cryptic species is complex, involving both sympatry and allopatry.

We propose mtSSU-rDNA as a tool for rapidly identifying cryptic species underlying ciliate morphospecies. Mitochondrial loci, primarily cytochrome *c* oxidase, have been shown to be effective in assessing intra-morphospecies variation in various ciliate clades (Barth et al. 2008; Catania et al. 2009; Chantangsi et al. 2007; Gentekaki and Lynn 2009; Snoke et al. 2006; Strüder-Kypke and Lynn 2010). To date, the mitochondrial locus of SSU-rDNA has only been used to assess phylogenetic relationships for deeper nodes in the Colpodea (Dunthorn et al. 2011). Yet mtSSU-rDNA primers can readily be used to assess the nature of diverse ciliate morphospecies (Dunthorn et al. 2011).

Divergence among loci

The cryptic species of *C. uncinata* maintain both highly conserved and highly divergent macronuclear coding domains. This pattern is particularly striking as micronuclear-specific sequences, at least between USA-ATCC and POL, are evolving at such a rapid rate that they can not be aligned. For example, internal eliminated sequences in the actin gene are more than 40% divergent (Katz and Kovner 2010) while the coding are 2.3% divergent (Table 3, Fig. 1C). The heterogeneity in levels of divergence among loci is reflected in patterns of codon usage, with more highly conserved genes showing more codon bias as measured by the effective numbers of codons (Wright 1990).

Further evidence for heterogeneity in patterns of evolution is seen in comparing the β -tubulin P3 paralogs to the conserved SP1/SP2 paralogs. These sequences arose by gene duplication within *C. uncinata*, yet the subsequent patterns of substitutions have varied such that the ratio of nonsynonymous to synonymous changes is higher for P3 sequences compared among strains ($dN/dS = 0.05$) as compared to the SP1/SP2 sequences ($dN/dS = 0$). Yet even the highly divergent P3 sequences (differ by ~24% amino acids from SP1/SP2) remain under some form of constraint, as the dN/dS for this clade is still much less than the value of 1.0 expected for pseudogenes. Moreover, the individual estimates of synonymous rates vary across lineages, with the P3 sequence having at least 10x higher rates of synonymous substitutions (Table 4).

Our data suggest that selection on coding domains in ciliates can be very strong over long evolutionary times. Even though sufficient time has elapsed for substitutions to render germline-specific sequence unalignable (Katz and Kovner 2010), coding domains at some loci (e.g. actin, β -tubulin SP1/SP2) remain conserved even to the level of eliminating most silent site mutations (Figs 1 and 2, Table 2). Within the same genomic context, some protein-coding genes evolve rapidly such as β -tubulin-P3 evolve and perhaps take on a new function. These observations are consistent with our hypothesis that the genome architecture of ciliates enables these lineages to explore protein space in novel manners (Katz and Kovner 2010; McGrath et al. 2006; Zufall et al. 2006).

Synthesis

The observation of elevated rates of evolution of a mitochondrial locus compared to most nuclear loci (the exception being the highly divergent β -tubulin paralog P3; Table 2) among lineages of the ciliate morphospecies *C. uncinata* is not surprising given elevated mutation rates in mitochondria in many taxa (Lynch 2007). More unusual, though, is the distinction in patterns of evolution between germline-limited (micronuclear) and somatic (macronuclear) sequences: here we document considerable conservation in coding domain, even among the silent sites within codons, while germline-limited sequences between two of the five cryptic species are so divergent that they are unalignable (Katz and Kovner 2010). These observations extend the comparisons of ciliate genome processing (i.e. between germline and somatic genomes) to the chromosomal rearrangements in B and T cells that generate diversity in the adaptive immune systems of some animals (Andersson et al. 2006; Herrick 1994; Kloc and Zagrodzinska 2001; Zufall et al. 2005). In contrast to adaptive immunity, where chromosomal rearrangements are associated with somatic hypermutation (Ginger et al. 2010; Zmasek and Godzik 2011), we observe conservation in coding domains (i.e. somatic sequences) embedded within rapidly evolving germline sequences.

Acknowledgments

This work was supported by two grants to L.A.K.: NSF DEB award (DEB 0816828) and an NIH AREA award 1R15GM081865-01.

Literature cited

- Andersson JO, Hirt RP, Foster PG, Roger AJ. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *Bmc Evolutionary Biology*. 2006; 6:18. [PubMed: 16512904]
- Barth D, Krenek S, Fokin SI, Berendonk TU. Intraspecific genetic variation in *Paramecium* revealed by mitochondrial cytochrome c oxidase I sequences. *Journal of Eukaryotic Microbiology*. 2006; 53:20–25. [PubMed: 16441579]
- Barth D, Tischer K, Berger H, Schlegel M, Berendonk TU. High mitochondrial haplotype diversity of *Coleps* sp. (Ciliophora: Prostomatida). *Environ. Microbiol.* 2008; 10:626–634. [PubMed: 18190514]
- Catania F, Wurmser F, Potekhin AA, Przybos E, Lynch M. Genetic Diversity in the *Paramecium aurelia* Species Complex. *Molecular Biology and Evolution*. 2009; 26:421–431. [PubMed: 19023087]
- Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N, Ikonomi P. Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. *Int. J. Syst. Evol. Microbiol.* 2007; 57:2412–2425. [PubMed: 17911319]
- Doherty M, Tamura M, Costas BA, Ritchie ME, McManus GB, Katz LA. Ciliate diversity and distribution across an environmental and depth gradient in Long Island Sound, USA. *Environmental Microbiology*. 2010; 12:886–898. [PubMed: 20113332]
- Dunthorn M, Foissner W, Katz LA. Expanding Character Sampling for Ciliate Phylogenetic Inference Using Mitochondrial SSU-rDNA as a Molecular Marker. *Protist*. 2011; 162:85–99. [PubMed: 20708960]
- Gentekaki E, Lynn DH. High-Level genetic diversity but no population structure Inferred from nuclear and mitochondrial markers of the peritrichous ciliate *Carchesium polypinum* in the Grand River Basin (North America). *Applied and Environmental Microbiology*. 2009; 75:3187–3195. [PubMed: 19304815]
- Ginger ML, Fritz-Laylin LK, Fulton C, Cande WZ, Dawson SC. Intermediary Metabolism in Protists: a Sequence-based View of Facultative Anaerobic Metabolism in Evolutionarily Diverse Eukaryotes. *Protist*. 2010; 161:642–671. [PubMed: 21036663]
- Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*. 2010; 27:221–224. [PubMed: 19854763]
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 2003; 52:696–704. [PubMed: 14530136]
- Herrick G. Germline-soma relationships in ciliates protozoa: the inception and evolution of nuclear dimorphism in one-celled animals. *Sem. Dev. Biol.* 1994; 5:3–12.
- Israel RL, Kosakovsky Pond SL, Muse SV, Katz LA. Evolution of duplicated alpha-tubulin genes in ciliates. *Evolution*. 2002; 56:1110–1122. [PubMed: 12144013]
- Juranek SA, Lipps HJ. New insights into the macronuclear development in ciliates. *International Review of Cytology*. 2007; 262:219–251. [PubMed: 17631190]
- Katz LA, Bornstein J, Lasek-Nesselquist E, Muse SV. Dramatic diversity of ciliate histone H4 genes revealed by comparisons of patterns of substitutions and paralog divergences among eukaryotes. *Mol Biol Evol*. 2004; 21:555–562. [PubMed: 14694079]
- Katz LA, Kovner AM. Alternative processing of scrambled genes generates protein diversity in the ciliate *Chilodonella uncinata*. *J. Exper. Zool. B*. 2010; 314:480–488.
- Katz, La; McManus, GB.; Snoeyenbos-West, OLO.; Griffin, A.; Pirog, K.; Costas, B.; Foissner, W. Reframing the 'Everything is everywhere' debate: evidence for high gene flow and diversity in ciliate morphospecies. *Aquatic Microbial Ecology*. 2005; 41:55–65.
- Kloc M, Zagrodzinska B. Chromatin elimination--an oddity or a common mechanism in differentiation and development? *Differentiation*. 2001; 68:84–91. [PubMed: 11686238]
- Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005; 21:676–679. [PubMed: 15509596]

- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. *Bioinformatics*. 2006; 22:3096–3098. [PubMed: 17110367]
- Lynch, M. *Origins of Genome Architecture*. Sunderland, MA: Sinauer Associates; 2007.
- Maddison, WP.; Maddison, DR. *MacClade v. 4.0.8*. Sinauer Assoc.; 2005.
- McGrath, C.; Zufall, RA.; Katz, LA. Genome evolution in ciliates. In: Katz, LA.; Bhattacharya, D., editors. *Genomics and Evolution of Eukaryotic Microbes*. Oxford University Press; 2006.
- Muse SV, Gaut BS. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates with application to the chloroplast genome. *Mol. Biol. Evol.* 1994; 11:715–724. [PubMed: 7968485]
- Nanney, DL. When is a rose?: the kinds of *Tetrahymena*. In: Wilson, RA., editor. *Species: new interdisciplinary essays*. Cambridge: The MIT Press; 1999. p. 93-118.
- Nanney DL, Park C, Preparata R, Simon EM. Comparison of sequence differences in a variable 23S rRNA domain among sets of cryptic species of ciliated protozoa. *J. Euk. Microbiol.* 1998; 45:91–100. [PubMed: 9495037]
- Peden, J. *CodonW version 1.4.2*. 2005.
- 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*. *Molecular Biology and Evolution*. 2008; 24:2495–2503. [PubMed: 17890762]
- Simon EM, Nanney DL, Doerder FP. The "*Tetrahymena pyformis*" complex of cryptic species. *Biodiv. Conserv.* 2008; 17:365–380.
- Snoke MS, Berendonk TU, Barth D, Lynch M. Large global effective population sizes in *Paramecium*. *Mol. Biol. Evol.* 2006; 23:2474–2479. [PubMed: 16987953]
- Sonneborn TM. Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl. Acad. Sci. USA*. 1937; 23:378–383. [PubMed: 16588168]
- Sonneborn, TM. Breeding systems, reproductive methods, and species problems in protozoa. In: Mayr, E., editor. *The species problem*. Washington DC: American Association for the Advancement of Science; 1957. p. 155-324.
- Strüder-Kypke MC, Lynn DH. Comparative analysis of the mitochondrial cytochrome *c* oxidase subunit 1 (CO1) gene in ciliates (Alveolata, Ciliophora) and evaluation of its suitability as a biodiversity marker. *Syst. Biodivers.* 2010; 8:131–148.
- Weisse T, Lettner S. The ecological significance of intraspecific variation among freshwater ciliates. *Verh. Internat. Verein. Limnol.* 2002; 28:1880–1884.
- Weisse T, Strüder-Kypke MC, Berger H, Foissner W. Genetic, morphological, and ecological diversity of spatially separated clones of *Meseres corlissi* Petz & Foissner, 1992 (Ciliophora, Spirotrichea). *Journal of Eukaryotic Microbiology*. 2008; 55:257–270. [PubMed: 18681840]
- Wright F. The 'effective number of codons' used in a gene. *Gene*. 1990; 87:23–29. [PubMed: 2110097]
- Zmasek CM, Godzik A. Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. *Genome Biology*. 2011; 12:13.
- 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]
- Zufall RA, Robinson T, Katz LA. Evolution of developmentally regulated genome rearrangements in eukaryotes. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*. 2005; 304B:448–455.

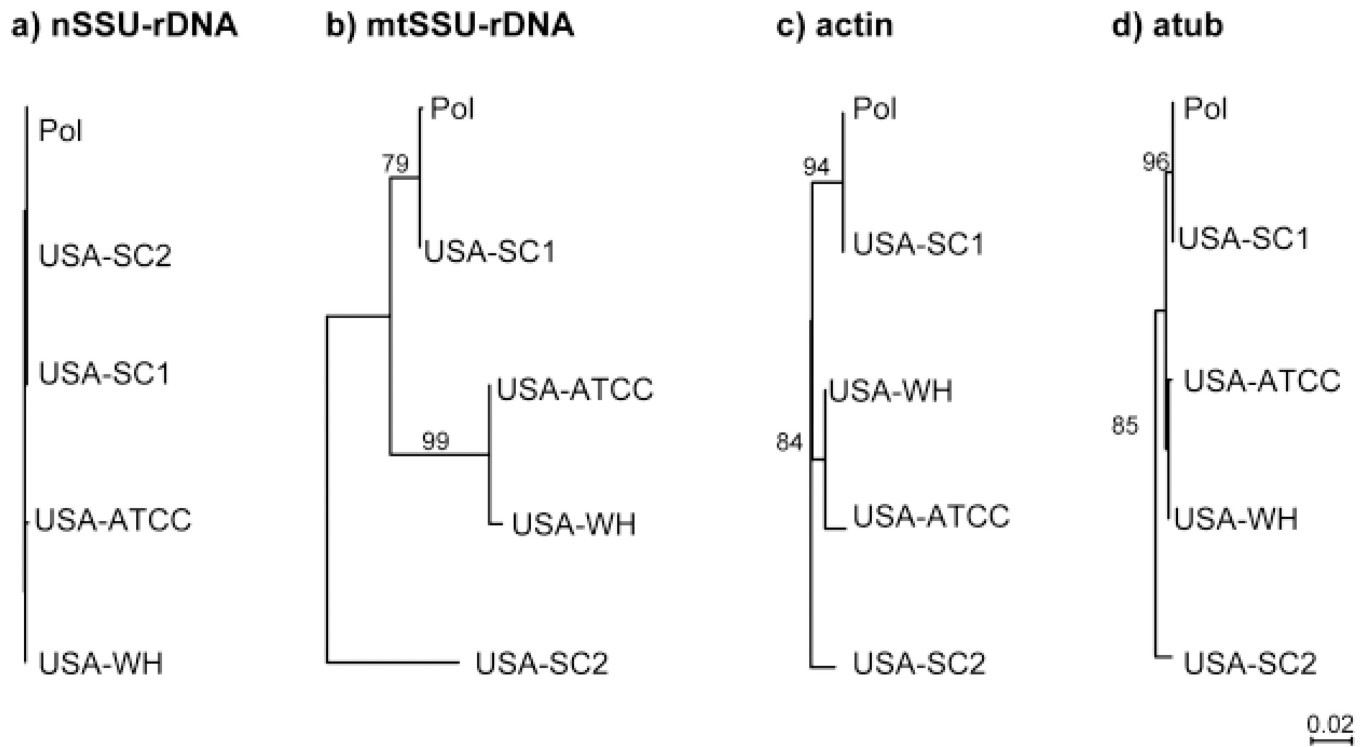


Figure 1. Genealogies of nSSU-rDNA, mtSSU-rDNA, actin and α -tubulin nucleotide sequences drawn to the same scale to highlight heterogeneity in rates of sequence evolution. All topologies estimated by PhyML (Guindon and Gascuel 2003) as implemented in Seaview (Gouy et al. 2010).

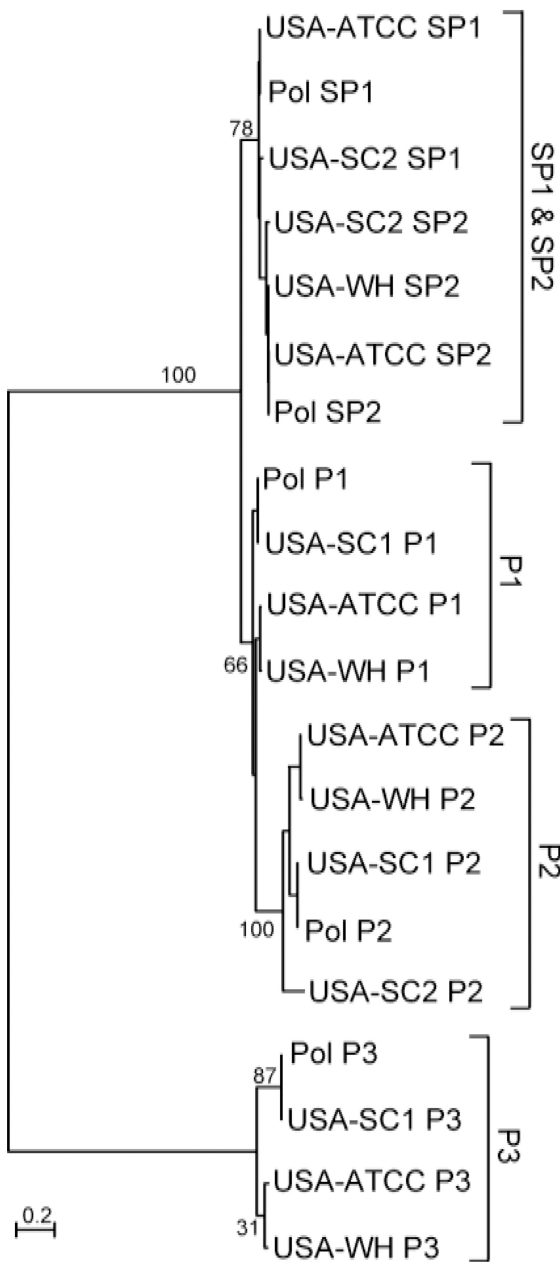
β -tubulin

Figure 2. Genealogies of β -tubulin gene family members demonstrating heterogeneity in rates of evolution among sequences as branch lengths for P3 are greater than for other members. See Figure 1 for additional notes.

Table 1Origin of *C. uncinata* clonal isolates compared in this study.

Abbreviation	Detail
POL	Originally derived as a contaminant in a culture of another ciliate sent from a lab in Poland. Deposited in 2009 as ATCC® PRA-256
USA-ATCC	Originally derived from a culture of another ciliate isolated from an undisclosed location in the USA, deposited in 1988 as ATCC® 50194
USA-SC1	Lyman Pond (Smith College): Northampton, MA, December, 2007: N42 19'07"; W72 38'24"
USA-SC2	Lyman Pond (Smith College): Northampton, MA September, 2007: N42 19'07"; W72 38'24"
USA-WH	Weeks Pond, Cape Cod MA, October 2007, Woods Hole MA: N41 33'21"; W70 36'52"

Table 2

The number of actin, α -tubulin, and β -tubulin clones characterized for each *C. uncinata* isolate.

	Actin		β -tubulin						Reference
	atub	SP1	SP2	P1	P2	P3			
POL	6	14	7	14	8	35	20	(Robinson and Katz, 2008)	
USA-ATCC	12	12	22	12	12	15	16	(Robinson and Katz, 2008)	
USA-SCI	2	21		21		1	2	This study	
USA-SC2	3	16	2	16	5			This study	
USA-WH	3	2	2	2	1	1	20	This study	

Table 3

Highly conserved gene family members are most biased in their codon usage (low number of effective codons (Nc)) while more divergent β -tubulin paralogs exhibit less bias.

	Nucleotide Divergence	AA Divergence	Nc (range)	Length (nuc)
nSSU-rDNA	0–0.4%	N/A	N/A	1060
mtSSU-rDNA	0.1–8.0%	N/A	N/A	905
actin	0–2.2%	0–0.5%	27–29	584
α -tubulin	0–2.3%	0	27–28	1033
β -tubulin SP1	0–2.3%	0	28–29	942
β -tubulin SP2	0–2.7%	0	29	942
β -tubulin P3	0–13.5%	1.0–6.2%	52–54	936

Divergence estimates represent the range of differences among strains. Effective number of codons, Nc, as estimated by Wright (1990). Beta-tubulin gene family members P1 and P2 are not included as they are alternatively processed from overlapping micronuclear regions (Katz and Kovner 2010).

Table 4

Heterogeneity in estimates of dN and dS along branches leading to SP1/SP2 paralogs compared to P3 paralogs

Strain	SP1		SP2		P3	
	dN	dS	dN	dS	dN	dS
Pol	6.25×10^{-18}	1.00×10^{-20}	0	4.48×10^{-3}	4.82×10^{-2}	9.13×10^{-1}
USA-ATCC	1.00×10^{-30}	6.61×10^{-18}	0	1.00×10^{-30}	2.18×10^{-3}	1.33×10^{-1}
USA-WH	-	-	0	1.80×10^{-2}	4.51×10^{-2}	5.82×10^{-2}