

Variation in Macronuclear Genome Content of Three Ciliates with Extensive Chromosomal Fragmentation: A Preliminary Analysis

CASEY L. McGRATH,^{a,1} REBECCA A. ZUFALL^b and LAURA A. KATZ^a

^aDepartment of Biological Sciences, Smith College, Northampton, Massachusetts 01063, and

^bDepartment of Biology and Biochemistry, University of Houston, Houston, Texas 77204

ABSTRACT. The genome architecture of ciliates, including features such as nuclear dualism and large-scale genome rearrangements, impacts gene and genome evolution in these organisms. To better understand the structure of macronuclear chromosomes in ciliates with extensively processed chromosomes, a sample of complete macronuclear chromosomes was sequenced from three ciliate species: *Metopus es* (Class [CI]: Armophorea), *Nyctotherus ovalis* (CI: Armophorea), and *Chilodonella uncinata* (CI: Phyllopharyngea). By cloning whole macronuclear chromosomes into a plasmid vector, we generated nine clones from each of *M. es* and *C. uncinata*, and 37 clones from *N. ovalis*. Analysis of these macronuclear chromosomes provides insight into the evolution of genome features such as chromosome content, gene structure, and genetic code. Phylogenetic patterns can be found in telomere structure and codon usage, which are both more similar in *M. es* and *N. ovalis* than *C. uncinata*. In addition, we provide evidence of lateral transfer of a bacterial endo- β -mannanase gene onto a *M. es* chromosome and report the discovery of a 42-bp conserved sequence motif within *N. ovalis* untranslated regions.

Key Words. Chromosome evolution, ciliates, codon usage, genome processing, lateral gene transfer, sequence motifs, telomere.

CILIAES are characterized by nuclear dualism: each cell contains a somatic macronucleus (MAC) and a germline micronucleus (MIC) (reviewed in Klobutcher and Herrick 1997; Prescott 1994; Riley and Katz 2001). Following conjugation, a new MAC develops from a zygotic nucleus. During this process, a mitotic copy of the zygotic genome undergoes large-scale chromosomal rearrangements, including fragmentation of chromosomes, DNA elimination, and chromosome amplification.

The extent to which genome rearrangements occur varies widely among ciliates. In ciliates of the class Oligohymenophorea (e.g. *Tetrahymena* and *Paramecium*), the rearrangements are comparatively limited. For example, during MAC development, *Tetrahymena* fragments its five large zygotic chromosomes into approximately 200 distinct MAC chromosomes, ranging in size from 100 to 1,500 kb and totaling approximately 110 Mb. Each of these chromosomes contains several hundred genes and is amplified an average of 45 times (Chalker and Yao 2001; Eisen et al. 2006; Orias and Higashinakagawa 1990; Prescott 1994).

By contrast, ciliates from three diverse classes, termed “extensive fragmenters,” undergo chromosomal processing that is much more extreme, generating a MAC with as many as 25,000,000 chromosomes. In ciliates of the classes Spirotrichea, Phyllopharyngea, and Armophorea, zygotic chromosomes are fragmented extensively such that macronuclear “chromosomes” contain often only a single gene (Akhmanova et al. 1998; Riley and Katz 2001; Steinbrück et al. 1995). These chromosomes are generally small (only a few kb) and contain little non-coding DNA (Boxma et al. 2005; Cavalcanti et al. 2004a). For example, some species of the class Spirotrichea fragment their approximately 120 zygotic chromosomes into as many as 24,000 macronuclear chromosomes, each less than 15 kb in length; each chromosome may be amplified 950–15,000 times (Klobutcher and Herrick 1997; Prescott 1994). This extensive processing appears to have arisen more than once within ciliates (Katz 2001). The multiple origins of extensive fragmentation have been argued to be the result of relatively plastic epigenetic mechanisms in ciliates that allow for the diversification of macronuclear structure (Katz 2001; McGrath, Zufall, and Katz 2006).

Corresponding Author: L. A. Katz, Department of Biological Sciences, Smith College, Northampton, Massachusetts 01063—Telephone number: 413 585 3825; FAX number: 413 585 3786; e-mail: Lkatz@smith.edu

¹Current address: Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142.

Ciliates also appear to be relatively plastic in their use of the genetic code. For example, some members of the Spirotrichea (*Oxytricha*, *Stylonychia*) and Oligohymenophorea (*Paramecium*, *Tetrahymena*) have independently converged upon translating UAA and UAG (stop codons in the universal code) into glutamine (Lozupone, Knight, and Landweber 2001; Meyer et al. 1991; Tourancheau et al. 1995). *Vorticella* and *Opisthocyta* (CI: Oligohymenophorea) translate UAA as glutamate (Sanchez-Silva et al. 2003). UGA is translated into cysteine in *Euplotes* (CI: Spirotrichea), and into tryptophan in *Blepharisma* (CI: Heterotrichea) and *Colpoda* (CI: Colpodea). The observation of frequent code changes in ciliates is particularly striking given how many ciliate lineages remain unexplored.

To better understand the relationship between genome architecture and molecular evolution, we have sequenced macronuclear chromosomes from three extensively fragmenting ciliates: *Metopus es* (CI: Armophorea), *Nyctotherus ovalis* (CI: Armophorea), and *Chilodonella uncinata* (CI: Phyllopharyngea). Previously, MAC chromosome sequencing efforts have focused on more well-studied ciliates in the classes Oligohymenophorea: *Paramecium tetraurelia* (Dessen et al. 2001; Sperling et al. 2002) and *Tetrahymena thermophila* (Orias 2000; Turkewitz, Orias, and Kapler 2002), and Spirotrichea: *Oxytricha trifallax* (Cavalcanti et al. 2004a, b; Doak et al. 2003). A limited number of MAC chromosomes and untranslated regions (UTRs) have also been characterized from *N. ovalis* (Boxma et al. 2005; Destables et al. 2005), one of the target species in this analysis.

MATERIALS AND METHODS

Cloning and sequencing of macronuclear chromosomes. We characterized macronuclear chromosomes from three ciliates: *M. es* (CI: Armophorea; CCAP 1653/2), *N. ovalis* (CI: Armophorea, previously isolated from cockroaches (Riley and Katz 2001)), and *C. uncinata* (CI: Phyllopharyngea; ATCC 50194). DNA was isolated as described previously (Zufall et al. 2006). Total genomic DNA, including gene-sized macronuclear chromosomes, was cloned using Novagen’s pSTBlue-1 Perfectly Blunt Cloning Kit (70191) and protocols. This procedure first generates blunt ends on all the chromosomes and then clones chromosomes less than ~6 kb (size selected by vector) into a blunt plasmid. Plasmid DNA was isolated from transformant colonies using the Qiagen (Valencia, CA) Miniprep Kit (27106), and sequenced with the BigDye terminator Kit (Applied Biosystems, Foster City, CA) on an ABI 3100 automated sequencer.

Complete MAC chromosomes were identified by the presence of telomere repeats on both the ends. When necessary, we designed internal primers to complete sequences of larger chromosomes.

Analysis of macronuclear chromosomes. We used a combination of BLAST analysis, tRNA scans, and inspection by eye to identify putative open-reading frames (ORFs) within completed macronuclear chromosomes. We used a low BLAST score of only $1E-04$ as an initial criterion for identifying ORFs, and we recognize that further experimental work is required to evaluate the significance of these identifications (Table 2). To identify putative introns and assess genetic codes, significant BLAST hits were compared by eye with translations in corresponding MAC chromosomes using the Seqbuilder program (DNAsar, Madison, WI). Chromosomes were also searched for putative tRNAs using tRNAscan-SE (available at <http://www.genetics.wustl.edu/eddy/tRNAscan-SE/>). If neither of these approaches yielded a convincing ORF, we searched for putative ORFs by eye using a minimum requirement of 50 amino acids. In the end, there were some complete macronuclear chromosomes on which we could not find any putative ORF; the function of these chromosomes remains unknown.

Putative UTRs, defined here as the region between the end of the telomere and the start or stop of a predicted gene, were searched for conserved motifs using the MEME (Multiple Em for Motif Elicitation) software (available at <http://meme.ncr.net/meme/website/meme.html>). We considered a motif to be significant if its score was sufficiently different from that produced with shuffled characters. Correspondence Analysis of Codon Usage (COA) was performed using the CodonW software (Correspondence Analysis of Codon Usage; available at <http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html>). Analysis of Codon Usage searches for correspondence among codon tables along a user-defined number of orthogonal axes (two, in our case). It is analogous to a principal component analysis in that both enable comparisons of discrete characters between taxa. The results enable comparison between codon tables (codon usage) in different lineages.

RESULTS

***Nyctotherus ovalis*.** We characterized 32 complete MAC chromosomes of *N. ovalis* (Table 1; GenBank Accession No. EF125715–EF125746). The average length of the complete *N. ovalis* chromosomes represented in the clones is 703 bp (range: 316–1,565 bp). Of the complete MAC chromosomes sequenced, we identified putative genes in 21 clones (Table 1). Of these, 12 have strong BLAST hits, which include a 3-methyl-2-oxobutanoate dehydrogenase and a fibrillarlin (Table 2). The average gene length is 582 bp (range: 270–1,116 bp). Twelve genes contain between 1 and 4 introns per gene, averaging 27 bp in length (range:

21–66 bp). Putative tRNAs were identified on eight chromosomes: one each of alanine, arginine, asparagine, and glutamine, and two different matches each to lysine and proline. The chromosomes containing tRNAs averaged 371 bp (range: 316–474 bp). We were unable to identify putative ORFs or tRNAs on three chromosomes. Two *N. ovalis* chromosomes (EF125733 and EF125734) were identical to each other except for a 9-bp insertion/deletion in the 3'-UTR.

The telomere repeat in *N. ovalis* is C_4A_4 . Each end of a *N. ovalis* chromosome is capped by predominantly four repeats, always followed by a final CCCC; six clones, however, have five repeats on one end and one clone has seven (Fig. 1). In addition to the 32 chromosomes analyzed here, five clones were sequenced and found to have telomeres on one end of the chromosome (data not shown). These single-telomere clones may either be experimental artifacts or degenerated macronuclear chromosomes.

The average predicted 5'-UTR length is 81 bp and the average 3'-UTR length is 65 bp. Using MEME software, a significant sequence motif (42 bp) was discovered in the 5'- and 3'-UTRs of *Nyctotherus* chromosomes that contained predicted genes. The consensus of this motif is: AAGGAATGAGTATTAGATAGT-GATGGTTTGATGATATATAAA (E value = 3.5×10^{-17} vs. 9.1×10^2 with shuffled characters). The motif occurs on 10 chromosomes at 16 sites, in both the 5'- and 3'-UTRs and on both strands (Fig. 2). Almost all of the 5' motifs are on one strand, while almost all the 3' motifs are on the other strand. The same motif is also found at an additional site on one of the chromosomes for which we were unable to identify an ORF.

To ask whether there is a single motif found at each 5'- and 3'-end, we constrained MEME to identify one motif per UTR sequence. Using this approach, we find a motif whose consensus (TGAGTGAATAAGAATA) is similar to positions 3–11 in the longer motif identified above. However, support for this shorter motif that appears once per sequence is lower (E value 2.8×10^{-2}).

Patterns of codon usage by *N. ovalis* demonstrate greater overlap in correspondence with that of *M. es*, another armophorean, than with *C. uncinata* (Phyllopharyngea) (Fig. 3).

***Metopus es*.** We sequenced nine full-length chromosomes from *M. es*, with an average length of 973 bp (Table 1; GenBank Accession No. EF125706–EF125714). We identified predicted genes for eight of these chromosomes with an average gene length of 669 bp. One larger chromosome was found to have two predicted genes that are non-overlapping and on opposite strands (GenBank Accession No. EF125714). Five predicted genes contained one to three introns (average length: 32 bp; range: 23–45). Although these introns may seem small compared with most eukaryotic introns, short (20–33 bp) introns have been found through sequencing of the *T. thermophila* macronuclear genome (Eisen et al. 2006) and in a variety of protein coding genes in

Table 1. Summary of the features of macronuclear (MAC) chromosomes sequenced from *Chilodonella uncinata*, *Metopus es*, and *Nyctotherus ovalis*.

Taxon	Clones		Content			Average lengths				
	Total	Telos ^a	ORF ^b	tRNAs ^c	No ORF ^d	Chromo ^e (bp)	ORFs ^f (bp)	Introns (bp)	5' UTR (bp)	3' UTR (bp)
<i>Chilodonella uncinata</i>	9	9	8	1	0	1235	1005	85	65	92
<i>Metopus es</i>	9	9	8	0	1	973	669	32	67	111
<i>Nyctotherus ovalis</i>	37	32	21	8	3	703	582	27	81	65

^aNumber of clones with telomeres on both ends.

^bNumber of clones that contain a putative open reading frame.

^cNumber of clones that contain a putative tRNA.

^dNumber of clones with no putative open reading frame or tRNA.

^eAverage lengths (in base pairs) of chromosomes with telomeres on both ends.

^fAverage lengths (in base pairs) of putative open reading frames.

Table 2. Significant BLAST hits of predicted genes identified in macronuclear (MAC) chromosomes of *Chilodonella uncinata*, *Metopus es*, and *Nyctotherus ovalis*.

Taxon	GenBank#	BLAST hits (<i>E</i> value ^a)
<i>Chilodonella uncinata</i>	EF125697	Serine/threonine protein phosphatase (2E-119)
	EF125698	Serine/threonine protein kinase (1E-09)
	EF125699	Thioredoxin (7E-04)
	EF125701	Amidase family protein (7E-36)
	EF125705	Caltractin-like (2E-06)
<i>Metopus es</i>	EF125712	RanBP1 domain containing protein (2e-24)
	EF125713	Endo- β -mannanase (2E-68)
<i>Nyctotherus ovalis</i>	EF125724	Cyclin-dependent kinase regulatory subunit family protein (6E-36)
	EF125728	Hypothetical protein (4E-34)
	EF125730	Calmodulin (4E-45)
	EF125733	Unknown protein (2E-21); contains DUF1000 domain (3E-29) ^b
	EF125738	Rab_A61 (7E-59)
	EF125740	Hypothetical protein (1E-06)
	EF125741	DNAJ domain containing protein (3E-37)
	EF125742	Expressed protein (2E-07)
	EF125743	B9 protein (6E-34)
	EF125744	Protein transport protein sec13 (1E-34)
	EF125745	Fibrillarlin (9E-95)
	EF125746	3-methyl-2-oxobutanoate dehydrogenase (2E-121)

^a*E* value from blastp (protein–protein BLAST) search.

^bConserved domains identified with blastp are noted for hypothetical and unknown proteins.

P. tetraurelia (Russell, Fraga, and Hinrichsen 1994). Two predicted genes had strong BLAST hits, which include an endo- β -mannanase and a zinc-finger protein (Table 2). As in *N. ovalis*, the telomere repeat we observed in *M. es* is also C₄A₄. Although the range in number of repeats on each chromosome end is also similar to that of *N. ovalis* (4–7 repeats), the number of repeats is more variable (Fig. 1).

We found one gene among the nine cloned chromosomes from *M. es* that was likely acquired by lateral gene transfer (GenBank Accession No. EF125713). The best BLAST hit (*E* value = 2×10^{-68}) to the translated ORF is to a bacterial endo- β -mannanase from a genome project on *Yersinia frederiksenii* (ZP_00828118). The closest eukaryotic BLAST hit (*E* value = 7×10^{-17}) is to *Arabidopsis thaliana* (NP_201447). There are no significant BLAST hits to the nucleotide sequence. In addition, this ORF contains no in-frame TAA or TAG codons, which

would code for glutamine in *Metopus*, but for stop codons in a bacterial gene.

The average predicted lengths of 5'- and 3'-UTRs within *M. es* were 67 and 111 bp, respectively. Although we did not identify a significant sequence motif in the UTRs of *M. es* chromosomes, the 5'-UTR of one chromosome is composed almost completely of the repeated sequence ATTGAT. This repeat begins immediately after the telomere, is tandemly repeated 7 times, and is followed only by four nucleotides before the start codon of the ORF.

***Chilodonella uncinata*.** We sequenced nine complete MAC chromosomes from *C. uncinata*, with an average length of 1,235 bp (Table 1; GenBank Accession No. EF125697–EF125705). We identified eight putative genes, with an average size of 1,005 bp, and one putative tRNA for a glutamine. Four predicted genes contain 1–2 introns, with an average intron size of 85 bp (range: 22–187). Five genes have significant BLAST hits, which include a serine/threonine protein phosphatase and an

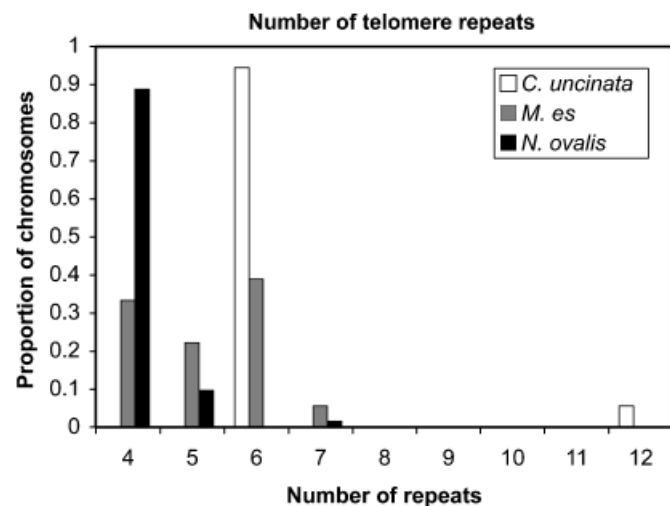


Fig. 1. Number of telomeric repeats observed within telomeres from macronuclear (MAC) chromosomes of *Chilodonella uncinata*, *Metopus es*, and *Nyctotherus ovalis*.

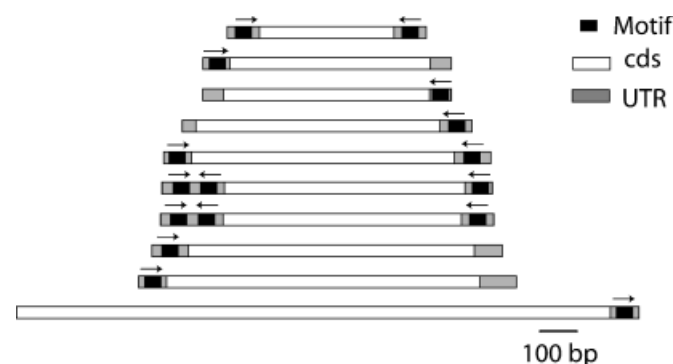


Fig. 2. Distribution of conserved 2-bp motifs in *Nyctotherus ovalis* chromosomes. Open rectangles indicate size and position of coding regions (exons plus introns), shaded rectangles predicted untranslated regions (UTR), and black rectangles the 2-bp motifs. Arrows indicate the orientation of the motif sequence. Images represent GenBank Accession No. EF125724, EF125727, EF125728, EF125731, EF125733, EF125734, EF125735, EF125737, EF125741, EF125746, ordered from top to bottom.

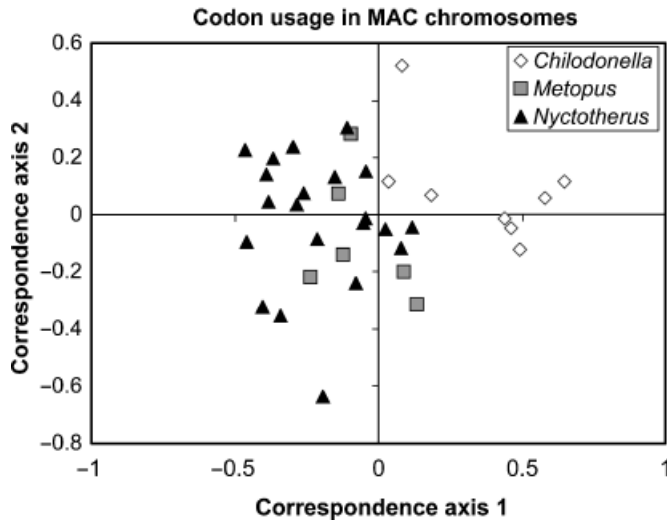


Fig. 3. Correspondence analysis of codon usage in *Chilodonella uncinata*, *Metopus es*, and *Nyctotherus ovalis* across two axes (see "Materials and Methods"). This analysis allows comparisons of codon usage tables between taxa. Codon usage is more similar in *M. es* and *N. ovalis* (Cl: Armophorea) than *C. uncinata* (Cl: Phyllopharyngea).

amidase family protein (Table 2). The average lengths of 5'- and 3'-UTRs are 65 and 92 bp, respectively.

The *C. uncinata* telomere repeat is C_4A_3 , which is unique among ciliates studied to date. The number of repeats on each end is conserved compared with the other two taxa, with almost all *C. uncinata* chromosome ends having six repeats (one telomere has 12 repeats; Fig. 1).

DISCUSSION

Chromosome structure. Our results confirmed the presence of extensively fragmented macronuclear chromosomes in the ciliates *N. ovalis* (Cl: Armophorea), *M. es* (Cl: Armophorea), and *C. uncinata* (Cl: Phyllopharyngea). The *C. uncinata* chromosomes

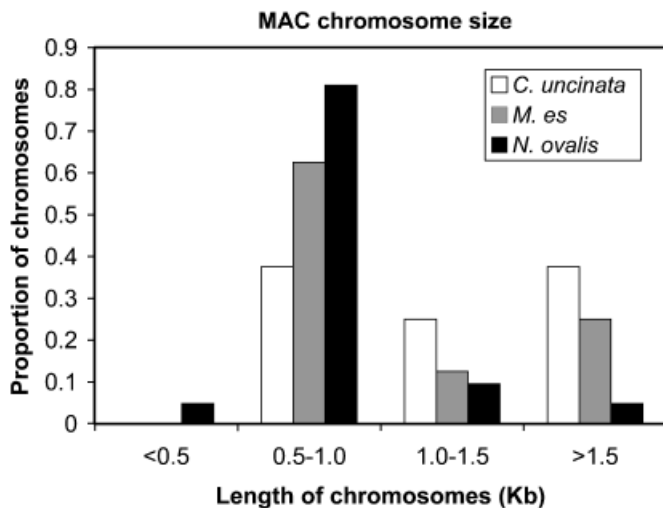


Fig. 4. Length (in kb) of macronuclear (MAC) chromosomes containing putative genes from *Chilodonella uncinata* (average length 1.3 kb), *Metopus es* (average length 1.1 kb), and *Nyctotherus ovalis* (average length 0.86 kb).

we cloned tend to be longer than those of the two Armophorea: about 63% of the chromosomes are longer than 1 kb in *C. uncinata*, compared with 33% in *M. es* and 14% in *N. ovalis* (Fig. 4). In addition, *C. uncinata* tends to have longer predicted genes, introns, and telomere repeats than the armophoreans. The difference in length may be due to either bias in chromosome sampling or may represent a difference in the genome structures of these species. Further investigations are needed to distinguish between these hypotheses.

Telomeres. We report a unique telomere sequence in *C. uncinata*, CCCCAA, and demonstrate that two members of the class Armophorea (*M. es* and *N. ovalis*) share a previously described telomere sequence, CCCCAA (Akhmanova et al. 1998). The observed Armophorea repeat is also found in members of the class Spirotrichea (Hoffman et al. 1995; Klobutcher et al. 1981), lending support to the hypothesized sister relationship of Spirotrichea and Armophorea (Katz 2001). These telomere sequences differ from that of the class Oligohymenophorea, the other class for which there are telomere data, which has the repeat [C/A]CCCAA (reviewed in Prescott 1994).

The consistency in number of telomere repeats across the three species varies with *C. uncinata* being most consistent, *M. es* being highly variable, and *N. ovalis* falling between. Although it is possible that this observation is the result of a loss of chromosome ends during the cloning procedure, it seems unlikely that chromosomes from the three species would be affected in different ways by the cloning protocol. The variation in telomere repeat number may also be an indication of differential regulation of telomerase or even different mechanisms in each of the species for extending and maintaining telomeres.

Genetic code. We found evidence of deviations from the universal genetic code in two of the three ciliate species. *Nyctotherus ovalis* genes use TAA and TGA as stop codons, supporting the hypothesis that *N. ovalis* utilizes the Standard Genetic Code (Destables et al. 2005). All *M. es* predicted genes have TGA as a stop codon, and some contain in-frame TAA and TAG codons, suggesting that *M. es* uses the ciliate MAC genetic code, where TAA and TAG are translated as glutamine (Lozupone et al. 2001; Meyer et al. 1991; Tourancheau et al. 1995). All the putative genes in *C. uncinata* use TAA as a stop codon, and one sequence characterized from this taxon for a different study has an in-frame TAG (Zufall et al. 2006). This indicates either a strong bias toward one stop codon if *C. uncinata* is using the standard genetic code or the use of an undescribed variant of the genetic code.

Codon usage. Patterns of codon usage in the predicted genes show a phylogenetic pattern: codon usage is more similar in *M. es* and *N. ovalis* (both Armophorea) than *C. uncinata* (Phyllopharyngea). This similarity is consistent with the closer relationship between these ciliates; however, the overlap is somewhat surprising given that the chromosomes we sequenced contain diverse genes with very different functions. This suggests that there may be phylogenetic conservation in features of translation such as tRNA abundances. However, codon usage tables from more members of these classes are needed to assess this possibility.

Sequence motifs. We report a conserved sequence motif (42 bp) that occurs at 16 sites across the UTR of 10 *N. ovalis* chromosomes. The frequency and location of this motif varies across chromosomes. When constrained to find one motif per UTR, as would be expected if the conserved sequence was involved in chromosomal processing, we find a shorter motif (16 bp) that has less support based on MEME analysis. The position of the shorter motif varies within the UTRs (data not shown). Such a conserved motif could provide any of a number of functions, including being a site for transcriptional regulation, chromosomal breakage, or replication. Sampling of additional chromosomes, coupled with experimental manipulation of this motif, is needed to

distinguish whether the motif is truly conserved and if so, what function it serves.

We found no sequence motifs shared by macronuclear chromosomes of the three ciliates. This lack of conservation suggests that regulatory regions in UTRs vary across this phylogenetic scale.

Lateral gene transfer. The genome structure of these ciliate macronuclei (the presence of gene-sized chromosomes bounded by telomeres) makes it easy to distinguish lateral gene transfer from bacterial contamination. We found one *M. es* macronuclear chromosome containing an endo- β -mannanase gene that appears to have been transferred from a bacterium. This chromosome has telomeres on both ends, contains an ORF that is most similar to a gene from *Y. frederiksenii*, contains no introns, and contains no in-frame bacterial stop codons. As more macronuclear chromosomes are characterized from ciliates with extensively processed genomes, it will be interesting to see the proportion of the genomes that contain laterally transferred genes.

ACKNOWLEDGMENTS

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

LITERATURE CITED

- Akhmanova, A., Voncken, F., van Alen, T., van Hoek, A., Boxma, B., Vogels, G., Veenhuiss, M. & Hackstein, J. H. P. 1998. A hydrogenosome with a genome. *Nature*, **396**:527–528.
- Boxma, B., de Graaf, R. M., van der Staay, G. W. M., van Alen, T. A., Ricard, G., Gabaldon, T., van Hoek, A. H. A. M., Moon-van der Staay, S. Y., Koopman, W. J. H., van Hellemond, J. J., Tielens, A. G. M., Friedrich, T., Veenhuis, M., Huynen, M. A. & Hackstein, J. H. P. 2005. An anaerobic mitochondrion that produces hydrogen. *Nature*, **434**:74–79.
- Cavalcanti, A. R. O., Dunn, D. M., Weiss, R., Herrick, G., Landweber, L. F. & Doak, T. G. 2004a. Sequence features of *Oxytricha trifallax* (class Spirotrichea) macronuclear telomeric and subtelomeric sequences. *Protist*, **155**:311–322.
- Cavalcanti, A. R. O., Stover, N. A., Orecchia, L., Doak, T. G. & Landweber, L. F. 2004b. Coding properties of *Oxytricha trifallax* (*Sterkiella histriomuscorum*) macronuclear chromosomes: analysis of a pilot genome project. *Chromosoma*, **113**:69–76.
- Chalker, D. L. & Yao, M. C. 2001. Non-genic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev.*, **15**:1287–1298.
- Dessen, P., Zagulski, M., Gromadka, R., Plattner, H., Kissmehl, R., Meyer, E., Betermier, M., Schultz, J. E., Linder, J. U., Pearlman, R. E., Kung, C., Forney, J., Satir, B. H., Van Houten, J. L., Keller, A. M., Froissard, M., Sperling, L. & Cohen, J. 2001. *Paramecium* genome survey: a pilot project. *Trends Genet.*, **17**:306–308.
- Destables, E., Thomas, N. A., Boxma, B., Van Alen, T. A., Van der Staw, W. M., Hackstein, J. H. P. & McEwan, N. R. 2005. The 3' untranslated region of mRNAs from the ciliate *Nyctotherus ovalis*. *Acta Protozool.*, **44**:231–236.
- Doak, T. G., Cavalcanti, A. R. O., Stover, N. A., Dunn, D. M., Weiss, R., Herrick, G. & Landweber, L. F. 2003. Sequencing the *Oxytricha trifallax* macronuclear genome: a pilot project. *Trends Genet.*, **19**:603–607.
- Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., 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., Cai, H., Collins, K., Stewart, B. 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., Patron, 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**:e286.
- Hoffman, D. C., Anderson, R. C., Dubois, M. L. & Prescott, D. M. 1995. Macronuclear gene-sized molecules of hypotrichs. *Nucleic Acids Res.*, **23**:1279–1283.
- Katz, L. A. 2001. Evolution of nuclear dualism in ciliates: a re-analysis in light of recent molecular data. *Int. J. Syst. Evol. Microbiol.*, **51**:1587–1592.
- Klobutcher, L. A. & Herrick, G. 1997. Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.*, **56**:1–62.
- Klobutcher, L. A., Swanton, M. T., Donini, P. & Prescott, D. M. 1981. All gene-sized DNA molecules in 4 species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proc. Natl. Acad. Sci. USA*, **78**:3015–3019.
- Lozupone, C. A., Knight, R. D. & Landweber, L. F. 2001. The molecular basis of nuclear genetic code change in ciliates. *Curr. Biol.*, **11**:65–74.
- McGrath, C., Zufall, R. A. & Katz, L. A. 2006. Genome evolution in ciliates. In: Katz, L. A. & Bhattacharya, D. (ed.), *Genomics and Evolution of Eukaryotic Microbes*. Oxford University Press, New York. p. 64–77.
- Meyer, F., Schmidt, H. J., Plumper, E., Hasilik, A., Mersmann, G., Meyer, H. E., Engstrom, A. & Heckmann, K. 1991. UGA is translated as cysteine in pheromone-3 of *Euplotes octocarinatus*. *Proc. Natl. Acad. Sci. USA*, **88**:3758–3761.
- Orias, E. 2000. Toward sequencing the *Tetrahymena* genome: exploiting the gift of nuclear dimorphism. *J. Eukaryot. Microbiol.*, **47**:328–333.
- Orias, E. & Higashinakagawa, T. 1990. Genome organization and reorganization in ciliated protozoa. *Zool. Sci.*, **7**:59–69.
- 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.
- Russell, C. B., Fraga, D. & Hinrichsen, R. D. 1994. Extremely short 20–33 nucleotide introns are the standard length in *Paramecium tetraurelia*. *Nucleic Acids Res.*, **22**:1221–1225.
- Sanchez-Silva, R., Villalobo, E., Morin, L. & Torres, A. 2003. A new non-canonical nuclear genetic code: translation of UAA into glutamate. *Curr. Biol.*, **13**:442–447.
- Sperling, L., Dessen, P., Zagulski, M., Pearlman, R. E., Migdalski, A., Gromadka, R., Froissard, M., Keller, A. M. & Cohen, J. 2002. Random sequencing of *Paramecium* somatic DNA. *Eukaryot. Cell*, **1**:341–352.
- Steinbrück, G., Raszikowski, S., Golembiewska-Skoczylas, M. & Sapetto-Rebow, B. 1995. Characterization of low and high molecular weight DNA in the macronucleus of the ciliate *Chilodonella steini*. *Acta Protozool.*, **34**:125–134.
- Tourancheau, A. B., Tsao, N., Klobutcher, L. A., Pearlman, R. E. & Adoutte, A. 1995. Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J.*, **14**:3262–3267.
- Turkewitz, A. P., Orias, E. & Kapler, G. 2002. Functional genomics: the coming of age for *Tetrahymena thermophila*. *Trends Genet.*, **18**:35–40.
- Zufall, R. A., McGrath, C., Muse, S. V. & Katz, L. A. 2006. Genome architecture drives protein evolution in ciliates. *Mol. Biol. Evol.*, **23**:1681–1687.

Received: 08/17/06, 11/22/06, 12/15/06, 01/20/07; accepted: 01/26/07