

Evolution of Programmed Ribosomal Frameshifting in the *TERT* Genes of *Euplotes*

Matthias Möllenbeck,¹ Michael C. Gavin,² Lawrence A. Klobutcher²

¹ Institute of Cell Biology, University Witten/Herdecke, Stockumer Str. 10, 58453 Witten, Germany

² Department of Biochemistry, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, CT 06032, USA

Received: 2 June 2002 / Accepted: 5 December 2003

Abstract. A number of recent studies indicate that programmed +1 ribosomal frameshifting is frequently required for the expression of genes in species of the genus *Euplotes*. In *E. crassus*, three genes encoding the telomerase reverse transcriptase (TERT) subunit have been previously found to possess one or two +1 frameshift sites. To examine the origin of frameshift sites within the *Euplotes* group, we have isolated segments of the *TERT* gene from five *Euplotes* species. Coupled with phylogenetic analysis, the results indicate that one frameshift site in the *TERT* gene arose late in the evolution of the group. In addition, a novel frameshift site was identified in the *TERT* gene of *E. minuta*, a species where frameshifting has not been previously reported. Coupled with other studies, the results indicate that frameshift sites have arisen during the diversification of the euplotids. The results also are discussed in regard to the mutations necessary to generate frameshift sites, and the specialization of TERT protein function that has apparently occurred in *E. crassus*.

Key words: Hypotrichous ciliates — *Euplotes* — Phylogenetic analysis — Programmed ribosomal frameshift — +1 frameshifting — Reverse transcriptase — Spirotrichea — Telomerase

Introduction

Translation typically proceeds with ribosomes reading and decoding successive three base codons of the mRNA. However, some genes contain the information necessary to encode a protein in separate reading frames, so that the ribosome must undergo a shift in frame to produce a full-length, functional protein. Such programmed translational frameshifting is relatively rare but is phylogenetically widespread, as examples are known in prokaryotes, eukaryotes, and a number of mobile genetic elements (reviewed by Baranov et al. 2002; Chandler and Fayet 1993; Ivanov et al. 2000). Sequence elements within the mRNA facilitate programmed frameshifting, and these typically include elements at the site of the frameshift, as well as more distally located sequences.

Over the past few years, a number of genes in ciliated protozoa of the genus *Euplotes* (class Spirotrichea) have been identified that appear to require a +1 translational frameshift to produce their protein products. Like other ciliated protozoa, *Euplotes* are binucleate single-cell organisms (reviewed by Jahn and Klobutcher 2002; Yao et al. 2002). Each cell contains a genetically silent micronucleus with conventional chromosomes, and a transcriptionally active polyploid macronucleus. The macronuclear genome is unusual in that it is composed exclusively of short linear, minichromosomes that typically contain single coding regions. The macronuclear DNA molecules are derived from a copy of the micronuclear genome during sexual reproduction (conjugation) through an extensive DNA reorganization process

that includes fragmentation of the micronuclear chromosomes and the *de novo* addition of telomeres to the ends of the resulting macronuclear minichromosomes.

The macronuclear chromosomes containing the genes encoding the regulatory subunit of cAMP-dependent protein kinase and a nuclear protein kinase of *E. octocarinatus* (Tan et al. 2001a, b), a gene encoding a La motif protein (p43) in *E. aediculatus* (Aigner et al. 2000), the ORF2 genes of the Tec2 transposons of *E. crassus* (Doak et al. 2003; Jahn et al. 1993), and one of the *E. crassus* genes encoding the reverse transcriptase subunit of telomerase (*TERT*) (Karamysheva et al. 2003; Wang et al. 2002) all appear to require a single +1 frameshift to produce a full-length protein product. Two additional *E. crassus TERT* genes appear to require two +1 translational frameshifts for expression (Karamysheva et al. 2003; Wang et al. 2002). The complete coding regions of only ~75 genes from various *Euplotes* species are known. Thus, although the sample size is currently small, the available data suggest that programmed translational frameshifting is unusually common in euplotids, with perhaps >5% of genes requiring a frameshift for expression (reviewed by Klobutcher and Farabaugh 2002).

The sequence elements necessary for the +1 frameshift in *Euplotes* genes have not been defined. However, all of the genes have a lysine codon (AAA) followed by a termination codon (TAA, except for ORF2 of the Tec transposons, in which it is TAG) and usually an additional A residue in the likely vicinity of the translational frameshift. This motif (AAA–TAA–A) bears similarities to sequence elements required for +1 frameshifting in some other noneuplotid genes (reviewed by Baranov et al. 2002; Stahl et al. 2002). These consist of (1) a “slippery codon,” which in this case would be the AAA lysine codon, that allows the ribosome to undergo a +1 shift in frame with the peptidyl-tRNA^{lys} still able to basepair with two nucleotides in the mRNA, and (2) a poorly recognized termination tetranucleotide (i.e., the termination codon plus the following nucleotide), which is thought to slow or stall the ribosome, allowing an opportunity for the ribosome to shift its reading frame. In these other systems, the frameshift appears to occur with the “slippery codon” occupying the P site of the ribosome. In the case of the *Euplotes* genes, there is no direct evidence concerning the precise position of the frameshift, but a number of lines of evidence suggest that it occurs when the ribosome encounters the AAA–TAA–A element (reviewed by Klobutcher and Farabaugh 2002). One aspect of the above model that appears inconsistent with *Euplotes* genes is that TAA–A appears to be used frequently at natural termination sites. However, a solution to this conundrum has been suggested

(Klobutcher and Farabaugh 2002) based on the fact that euplotids have also undergone stop codon reassignment, with the UGA stop codon of the standard genetic code now encoding cysteine (Lozupone et al. 2001; Tourancheau et al. 1995). Stop codon reassignment is thought to require changes in eukaryotic release factor 1 (eRF1), such that it no longer interacts with one or more of the standard stop codons. In euplotids, the changes in eRF1 that resulted in loss of the ability to recognize UGA as a stop codon may have also resulted in impaired recognition of the remaining two stop codons, producing a situation where termination is slow and normally rare shifts in reading frame are enhanced. It is possible that +1 ribosomal frameshifting is quite efficient in euplotids, so that the generation of frameshift sites in genes has no, or minimal, effects on the levels of the encoded protein product.

In this study we sought to determine if frameshift sites have arisen in *Euplotes* genes during the evolution of this group. An initial indication that this is the case has come from the analysis of *Euplotes TERT* genes. The single *TERT* gene isolated from *E. aediculatus* does not require frameshifting for expression (Lingner et al. 1997). In contrast, three *TERT* genes have been identified in *E. crassus* (Karamysheva et al. 2003; Wang et al. 2002). Two of these *E. crassus* genes (*EcTERT-1* and *EcTERT-3*) are present in the macronucleus and would require two +1 frameshifts for expression. The third gene (*EcTERT-2*) requires a single +1 frameshift for expression, but is present only in the micronucleus. It has been suggested that *EcTERT-2* functions in forming a telomerase for *de novo* telomere addition during macronuclear development, while the other two *TERT* genes encode proteins involved in maintaining telomere length during asexual reproduction (Karamysheva et al. 2003). We have used the polymerase chain reaction (PCR) to isolate portions of *TERT* genes corresponding to the site of the second frameshift in *EcTERT-1* and *EcTERT-3* from five other euplotids species. Coupled with phylogenetic analyses, the results indicate that the second frameshift site is present only in genes of late diverging species. In addition, we identified a novel frameshift site in the *TERT* gene of *E. minuta*, providing a further indication that frameshift sites have arisen during the evolution of the group. The data also suggest that the three *E. crassus TERT* genes have arisen by gene duplication relatively recently.

Materials and Methods

Cells and DNA Isolation

The DNA preparations for the various *Euplotes* species employed in this study were kindly provided by Dr. Carolyn L. Jahn

(Northwestern University Medical School, Chicago, IL). The particular strains used to prepare the DNA, and the sources of the cells, were as follows: *E. raikovi* strain 13 (CCAP 1624/19), *E. minuta* strain Lb-9 (CCAP 1624/18), and *E. rariseta* strain GES-3 were obtained from Dr. Pierangelo Luporini (University of Camerino, Camerino, Italy); *E. vannus* strain TM2 was obtained from Dr. Michael Gates (Cleveland State University, Cleveland, OH); and *E. eurytomus* GL from Dr. David Prescott (University of Colorado, Boulder, CO).

Oligonucleotides and PCR

All oligonucleotides used in PCR and sequencing were purchased from GIBCO-BRL Life Technologies (Gaithersburg, MD). The names and sequences of the oligonucleotides used in these studies (Fig. 1) are listed below using IUPAC nomenclature (M = A/C; N = A/C/G/T; W = A/T; S = C/G; R = A/G; Y = C/T; D = T/G/A; H = A/T/C). The following degenerate primers were used to amplify the regions surrounding frameshift site 2 of the *TERT* gene: Telo3, 5'-CCNATHATGCANTTYAAYAARAA RATH-3'; Telo4, 5'-CATRTCDATNSWDATNCCDATCCA-3'; and Telo5, 5'-ATGGAYATHGARAARTGHTAYGA-3'.

PCR, followed by direct sequencing of the bulk PCR product, was also performed for each of the newly isolated *TERT* gene segments to verify the presence/absence of frameshift sites. For *E. eurytomus*, *E. vannus*, and *E. minuta* the following primers designed from the cloned sequences were used in combination with Telo4 for this purpose: Stelo18 (*E. eurytomus*), 5'-ATGACAGCACAAATTCT AAAAAGAAAGA-3'; Stelo19 (*E. vannus*), 5'-ATGACTGTCCA AGCTATGAAGAGAAACA-3'; and Stelo20 (*E. minuta*), 5'-ATG AAAGTACAAATGTAAAGAGAAATA-3'. For the other species, the following pairs of primers designed from the cloned sequences were employed: Aaed1 (*E. aediculatus*), 5'-TACTTTCT TCAGATTTCTGG-3'; Aaed2 (*E. aediculatus*), 5'-TAATGACT GGTTGAAGTAAG-3'; Erar1 (*E. rariseta*), 5'-ATAACATTGTG AGTGACTCC-3'; Erar2 (*E. rariseta*), 5'-TTTGAAGTAGTTG TACTGCC-3'; Eraik1 (*E. raikovi*), 5'-GACCACTCAGTTGT TATCAC-3'; and Eraik2 (*E. raikovi*), 5'-CTGGTTTAAAGAGG GTGTCC-3'.

PCR was carried out using 50–100 ng of total genomic DNA and KlenTaq DNA polymerase (Sigma, St. Louis, MO) under conditions specified by the manufacturer. Thirty cycles of PCR were typically carried out, with a cycle consisting of a 95°C denaturation step for 1 min, an annealing step for 1 min, and a 72°C elongation step for 30 s to 1 min, depending on the length of the expected product. The temperature for the annealing step was adjusted based on the GC content of the primers. In cases in which degenerate primers were used, appropriate annealing temperatures were determined empirically. The degenerate primers Telo3 and Telo4 were used in the initial PCR reactions to isolate *Euplotes TERT* genes. In some cases a second “step-in” amplification was necessary in order to obtain a product. In these instances, five microliters of the initial PCR reaction products was used as the substrate in a subsequent PCR reaction (30 cycles) with the Telo5 degenerate oligonucleotide in combination with Telo4. The PCR products were analyzed on agarose or low-melting point (LMP) agarose gels. For cloning or direct sequencing, PCR reactions were run on a LMP agarose gel (GIBCO-BRL Life Technologies), and the appropriate DNA fragments excised and purified (Qian and Wilkinson 1991).

DNA Cloning and Sequencing

PCR products were cloned into the pCR 4-TOPO vector using the TOPO TA Cloning Kit for Sequencing (Invitrogen, Carlsbad, CA). Sequencing of cloned PCR products, as well as direct sequencing of

total PCR products, was carried out by the University of Connecticut Health Center Molecular Core facility using the *Taq* Dyedeoxy Termination Cycle Sequencing Kit (Perkin Elmer Cetus, Norwalk, CT). The T3 and T7 sequencing primers were used for sequencing clones, while the species-specific oligonucleotides were used for bulk sequencing of PCR products.

The DNA sequences of the two clones obtained from each organism are deposited in GenBank with the following accession numbers: *E. vannus* (AY303933/AY445830), *E. minuta* (AY303934/AY445831), *E. eurytomus* (AY303935/AY445832), *E. rariseta* (AY303936/AY445833), and *E. raikovi* (AY303932/AY445829).

Sequence and Phylogenetic Analyses

Similarity searches with newly obtained sequences were performed using BLAST (Altschul et al. 1997). Multiple sequence alignments of DNA and protein sequences were carried out using Clustal X with default parameters unless noted otherwise (Thompson et al. 1997). The alignments were subsequently used to create phylogenetic trees by the maximum parsimony, neighbor-joining, and minimum evolution methods using the Mega version 2.1 program package (Molecular Evolutionary Genetics Analysis [Kumar et al. 2001, 1994]). Confidence levels were evaluated by using 1000 bootstrap replications in each case. The following *Euplotes* sequences were used in the phylogenetic analyses: *E. vannus* (AY303933), *E. minuta* (AY303934), *E. eurytomus* (AY303935), *E. rariseta* (AY303936), *E. raikovi* (AY303932), *E. crassus* (*EcTERT-1*, AF528527; *EcTERT-2*, AY267543; *EcTERT-3*, AY267544), and *E. aediculatus* (U95964). In cases where a frameshift site was present, the predicted protein sequence was generated by removing the “T” nucleotide in the 5'-AAA-TAA-3' frameshift motif, as the frameshift is likely to occur within or near this motif (see Klobutcher and Farabaugh 2002). Some or all of the following noneuplotid ciliate and eukaryotic *TERT* gene and protein sequences were included in the analyses as outgroups: *Oxytricha trifallax* (AF060230), *Tetrahymena thermophila* (AF061284), *Paramecium tetraurelia* (AF515460), *Paramecium caudatum* (AB035309), *Saccharomyces cerevisiae* (Q06163), and *Homo sapiens* (AF015950). Additional trees were constructed using the sequences of the second cloned euplotid *TERT* genes/proteins, which contain a small number of sequence substitutions, but these showed no significant differences from the trees presented (data not shown).

A sliding window analysis of DNA sequence identity among the three *E. crassus TERT* gene coding regions was carried out using the Web-based version of mVISTA (<http://www-gsd.lbl.gov/vista/> [Bray et al. 2003; Dubchak et al. 2000; Mayor et al. 2000]) and employed a 25-base window.

Results and Discussion

Isolation and Sequencing of Segments of *TERT* Genes from Additional *Euplotes*

Segments of *TERT* genes corresponding to the second frameshift sites (FS2) in the *E. crassus TERT* genes were obtained from additional *Euplotes* species using a PCR procedure (Fig. 1). Degenerate oligonucleotide primers (Telo3, Telo4, and Telo5) were designed to regions of the gene that correspond to conserved amino acid residues in ciliate *TERT* proteins and used in PCR reactions with total genomic DNA from species of *Euplotes* where *TERT* genes had not previously been isolated. PCR products of

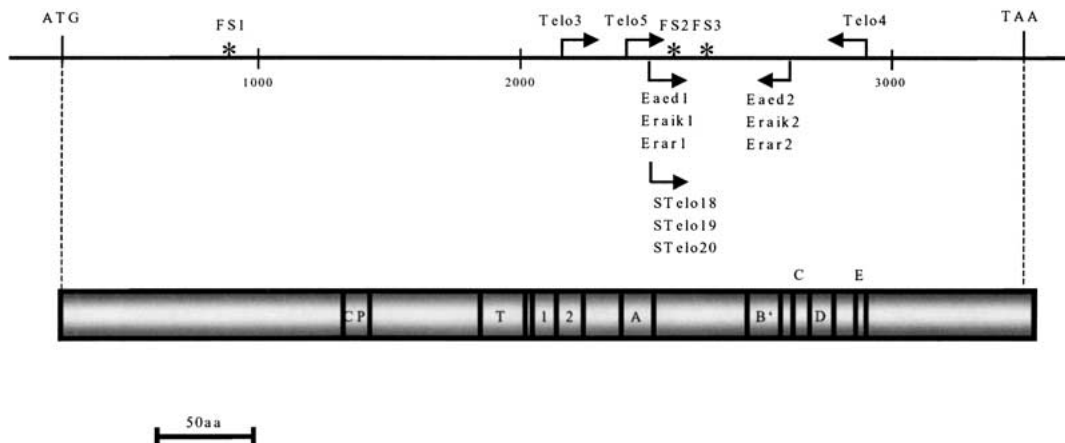


Fig. 1. Schematic illustration of the *Euplotes TERT* gene and encoded protein. The gene (top) as well as the protein (bottom) is shown to scale, with numbers referring to nucleotide positions and the bar representing 50 amino acids (aa). The start (ATG) and stop (TAA) codons are indicated, and the positions of the three known frameshift sites to (FS1–3) are marked with asterisks. Degenerate primers used for the initial PCRs (Telo3–5) are shown above the

gene, while species-specific primers (STelo18–20, Eaed1, Eaed2, etc.) used in subsequent PCR and bulk sequencing analyses are shown below the gene, with the arrows indicating the direction of synthesis. The locations of RT (reverse transcriptase) motifs 1, 2, and A to E (Xiong and Eickbush 1990), telomerase-specific motif T (Nakamura et al. 1997), and the motif specific for ciliated protozoa (CP) (Bryan et al. 1998) are specified.

the expected size were then cloned into a plasmid vector, and two clones isolated for sequence analysis from each species. In this manner, we were able to isolate the FS2 regions from five species (Fig. 2): *E. vannus*, *E. minuta*, *E. eurystomus*, *E. rariseta*, and *E. raikovi*. In each case, the sequences obtained from the two independent recombinant clones were extremely similar, differing by only 0.28–0.75%. Some of these sequence differences may be the result of PCR errors, but ~40% of the changes are synonymous, indicating that at least some may be the result of allelic variation. The latter explanation is not unreasonable, as none of the *Euplotes* strains employed were inbred. Although the number of clones analyzed was small, there were no indications of multiple *TERT* genes in any of the species. Finally, since the PCR procedure employed total genomic DNA as the template, where macronuclear DNA is much more abundant than micronuclear DNA (e.g., ~500:1 macronuclear to micronuclear copies in *E. crassus* [Baird and Klobutcher 1991]), it is likely that these sequences are derived from the macronuclear *TERT* genes of these *Euplotes* species, and not from any possible micronuclear copies that might be equivalent to the developmentally expressed and micronuclear-limited *EcTERT-2* gene of *E. crassus* (Karamysheva et al. 2003).

Similar PCR strategies were used in an attempt to obtain the frameshift site 1 (FS1) region of *TERT* genes from other euplotids. However, we were not successful in obtaining this segment of the *TERT* gene from other species despite using a number of PCR primer combinations directed against regions of the *TERT* gene corresponding to conserved amino acid residues.

The DNA sequences and predicted amino acid sequences of the *Euplotes TERT* genes around the FS2 site are shown in Fig. 2A. Like the previously analyzed *E. aediculatus TERT* gene (Lingner et al. 1997), the sequences from *E. minuta*, *E. eurystomus*, *E. rariseta*, and *E. raikovi* predict a continuous open reading frame through this region. In contrast, the *E. vannus* sequence contains an in-frame TAA termination codon. This termination codon is at the same position as those in the previously analyzed *E. crassus EcTERT-1* and *EcTERT-3* genes (Karamysheva et al. 2003; Wang et al. 2002) and occurs in the context of the AAA–TAA motif that is thought to promote frameshifting (Fig. 2A). In addition, 108 bp downstream of the FS2 site, an in-frame TAA termination codon was identified in the *E. minuta TERT* gene segment (Fig. 2B). The TAA codon begins at what corresponds to position 2171 in the *E. aediculatus TERT* gene and, again, occurs within a context (AAA–TAA) that is expected to promote a frameshift in *Euplotes* genes (Klobutcher and Farabaugh 2002). A +1 frameshift in this region would be required to produce a full-length *TERT* of protein with an amino acid sequence similar to that of other euplotid *TERT*s (Fig. 2B). To confirm the presence/absence of the FS2 and FS3 sites in the newly obtained *TERT* gene sequences, as well as in the *E. aediculatus TERT* gene, the frameshift regions were amplified from total genomic DNA using one or two primers designed using the sequences from the clones (Fig. 1; see Materials and Methods), and the resulting PCR products were directly sequenced. In each case, the bulk DNA sequences confirmed the presence/absence of the frameshift sites (data not shown).

A.

```

+
EcTERT-1 GTAGTTGATTGGATAAATGTGCAAGAAATAAGATGAAAGAGTGCTTCAGACAAAAATGGCAAAGATT
V V D L D K C A R K *
K M K E C F R Q K W Q K I
EcTERT-2 GTTATTGATTCTCA-AAATTTCAAAAAGAAATGAGAGATTATTTTAGGAAGAAATGGCAAAGATT
V I D S Q N F K K K K M R D Y F R K K W Q K I
EcTERT-3 GTAGTAGATCAAGAAAAATGTGAGAGAAATAAGATGAAGGATTGCTTCAGGCAGAAATGGCACAGATT
V V D Q E K C E R K *
K M K D C F R Q K W H K I
Evan GTTGTGATTATGATAAGTGTGAGAGAAATAAGATGAGAGAGTGCTTCAGACAGAGGTGGCAGAAAT
V V D Y D K C Q R K *
K M R E C F R Q R W Q K I
Emin CACATCGAAACGAA-GAACCTTAAGATCGAAAAATGAGCGAGCACTACCGCCAAAAGTGGCAGAGAATT
H I E T K N L K I E K M S E H Y R Q K W Q R I
Eaed GTTATCGATTGCAA-AAACTTTAGAAAAGAAAGAAATGAAAGATTATTTTAGACAGAAATCCAGAAGATT
V I D S K N F R K K E M K D Q F R Q K F Q K I
Eeury GTTATTGATTCTAA-GAACTTTAGAAAAGAAAGAAATGAAAGACTATTTTACAGACAGAAATCCAAAAAATT
V I D S K N F R K K E M K D Y F R Q K F Q K I
Erari GTGGTTGACTCCAA-GAACTTCCAAAAGAAACCTTTGAAAGACTACTTTAGACAGAAATTTCCAGAACTC
V V D S K N F Q K K P L K D Y F R Q K F Q K L
Eraik GTAATCGACTCTAA-CAACCATAGAAAACGTAAGATGAGTTACTACTTCAAGCAAAGTGGAGAGATATC
V I D S N N H R K R K M S Y Y F K T K W R D I

```

B.

```

+
EcTERT-1 GACGATCAGAACGACCTGAATGCCAAG-GAGACTCTCCTGGTTGAGAATCAGAAGAGAGAT
D D Q N D L N A K E T L L V E N Q K R D
EcTERT-2 AATGATCAAATGAATTGAATGCCAAG-AAGACTCTCTGGTTGAGAATCAGAAGAGAGAT
N D Q N E L N A K K T L L V E N Q K R D
EcTERT-3 AATGACCAAAATGACCTGAATGCTAAG-AAGACTCTCTGGTTGAGAACCAGAAGAGAGAT
N D Q N E L N A K K T L L V E N Q K R D
Eaed AATGAACAAAATGACTTAAATGCAAAG-AAAACATTAATTGTTGAAGCAAAGCAAAGAAAT
N E Q N D L N A K K T L I V E A K Q R N
Evan AATGATCAGAATGAGTTGAATGCCAAG-AAGACTTTACTTGTGAAAACCAAGAGAGAGAT
N D Q N E L N A K K T L L V E N Q K R D
Eeury AATGAACAAAACGATTTAAATGCCAAG-AAAACCTTAATAGTTGAGGCAAACAAAGAAAT
N E Q N D L N A K K T L I V E A K Q R N
Emin CAAGATCAAATGAACTGAATGCTAAATAAACACTAATAGTTGAGAACCAGAAGAGAGAT
Q D Q N E L N A K *
K T L I V E N Q K R D
Eraik CAAGAACAAAACGAAATCAATGCCAAA-AAGACGATAATTGTTGAATCTAAAAAGCGAGAC
Q E Q N E I N A K K T I I V E S K K R D
Erari AATGATCAAATGATTTAAATGCCAAG-AAAACATTAGTTGTTGAAAACAAAAGAGAATG
N E Q N D L N A K K T L V V E N K K R M

```

Fig. 2. *Euplotes TERT* DNA and protein sequences in the vicinity of putative programmed +1 frameshift sites. **A** The DNA and predicted protein sequences of the three *E. crassus TERT* genes in the vicinity of the FS2 site are shown (EcTERT-1, EcTERT-2, and EcTERT-3), along with the corresponding regions from the *TERT* genes of *E. vanus* (Evan), *E. minuta* (Emin), *E. aediculatus* (Eaed), *E. eurytomus* (Eeury), *E. rariveta* (Erari), and *E. raikovi* (Eraik). Asterisks indicate stop codons, putative frameshift signals in the DNA sequences are highlighted, and the + signs denote the likely positions of the base insertions associated with the generation of the shift in reading frame. DNA sequences were aligned with ClustalW using default parameters, except that the gap opening

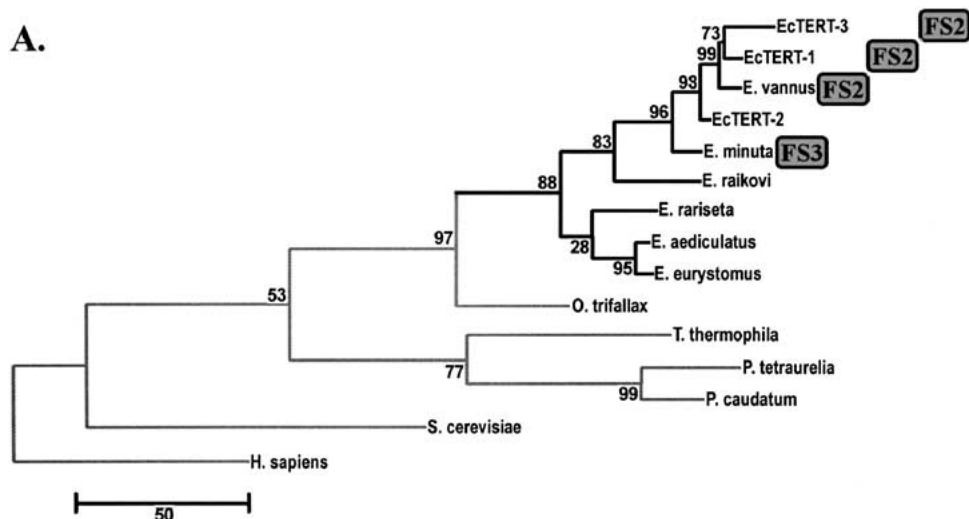
penalty was increased to 30. For the frameshifted genes, the predicted amino acid sequence in the initial reading frame is shown up to the stop codon, followed by the predicted amino acid sequence in the +1 frame. The region of DNA and amino acid sequences shown correspond to residues 2033–2101 and 645–667 of the DNA and protein sequences, respectively, of the *E. aediculatus TERT* (GenBank accession U95964). **B** The DNA and predicted protein sequences of the *Euplotes TERTs* in the vicinity of the FS3 site. The regions shown correspond to residues 2147–2206 and 683–702 of the DNA and protein sequences, respectively, of the *E. aediculatus TERT*. Other aspects of the figure as described above.

Phylogenetic Analyses

Phylogenetic analyses were performed to investigate the origin of frameshift sites within the *TERT* genes of *Euplotes* species. Maximum parsimony, minimum

evolution, and neighbor-joining analyses were carried out on both the *Euplotes* frameshift site 2 region inferred amino acid sequences (residues 610–846 of the *E. aediculatus TERT* protein) and the corresponding DNA sequences, along with

A.



B.

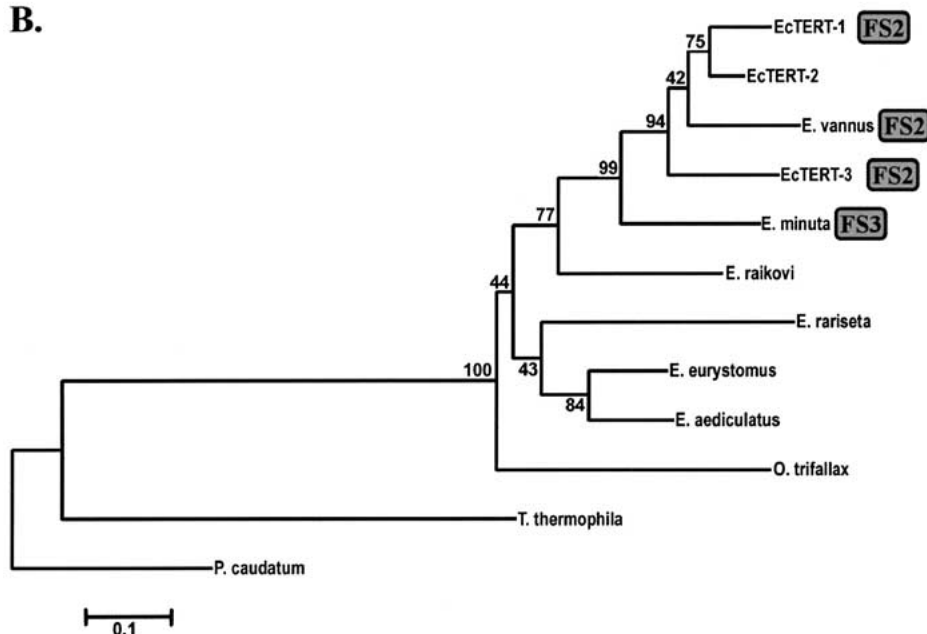
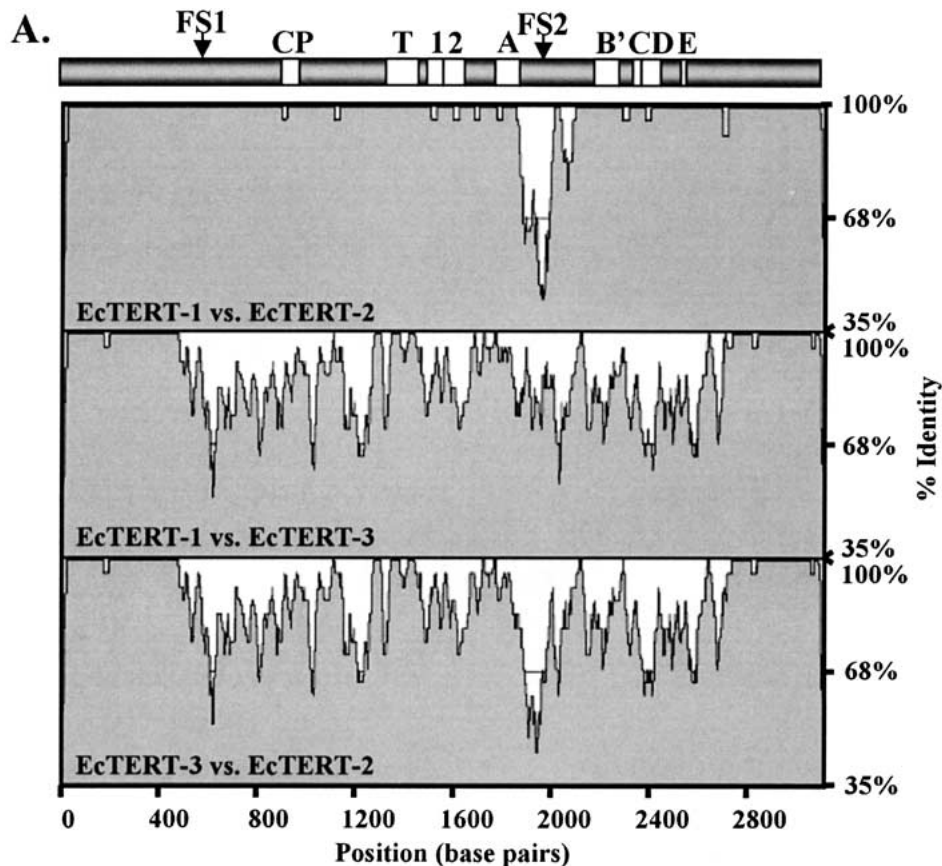


Fig. 3. Phylogenetic analyses of TERT protein and DNA sequences. **A** A tree generated by maximum parsimony analysis of the amino acid sequences in the vicinity of frameshift site 2 (corresponding to amino acids 609–847 of the *E. aediculatus* TERT protein) is shown, with the genes containing either FS2 or FS3 indicated. Euplotid branches are shown in black, with those of other ciliate and eukaryotic species in gray. The numbers at nodes represent bootstrap support values. Horizontal branch lengths are drawn to scale, with the bar indicating the substitutions per interval analyzed. In addition to the newly obtained sequences, regions of

TERT proteins from the following organisms were included in the phylogenetic analyses: *E. crassus* (EcTERT-1, AF528527; EcTERT-2, AY267543; EcTERT-3, AY267544), *E. aediculatus* (U95964), *Oxytricha trifallax* (AF060230), *Tetrahymena thermophila* (AF061284), *Paramecium tetraurelia* (AF515460), *Paramecium caudatum* (AB035309), *Saccharomyces cerevisiae* (Q06163), and *Homo sapiens* (AF015950). **B** A neighbor-joining tree produced from the TERT DNA sequences corresponding to amino acids 609–847 of the *E. aediculatus* TERT protein is shown. The scale bar denotes changes per residue.

sequences from other ciliates and eukaryotes as outgroups. Similar results were obtained with each method, and a maximum parsimony tree of the protein sequences is shown along with a neighbor-joining tree generated from the DNA sequences in Fig. 3. In all approaches, the ciliate species formed a clade, and there was strong bootstrap support for the major divisions of ciliates (Lynn and Small 1997). Specifically, the spichotrich ciliates (*Oxytricha*

trifallax and the *Euplotes* species) and the oligohymenophoran ciliates (*Tetrahymena thermophila* and *Paramecium* species) each formed well-supported monophyletic groups. The *Euplotes* species (hypotrich ciliates) also consistently formed a clade, although bootstrap support for the group was weak in the neighbor-joining protein and DNA trees (42 and 44%, respectively), as well as the minimum evolution DNA tree (52%).



B.

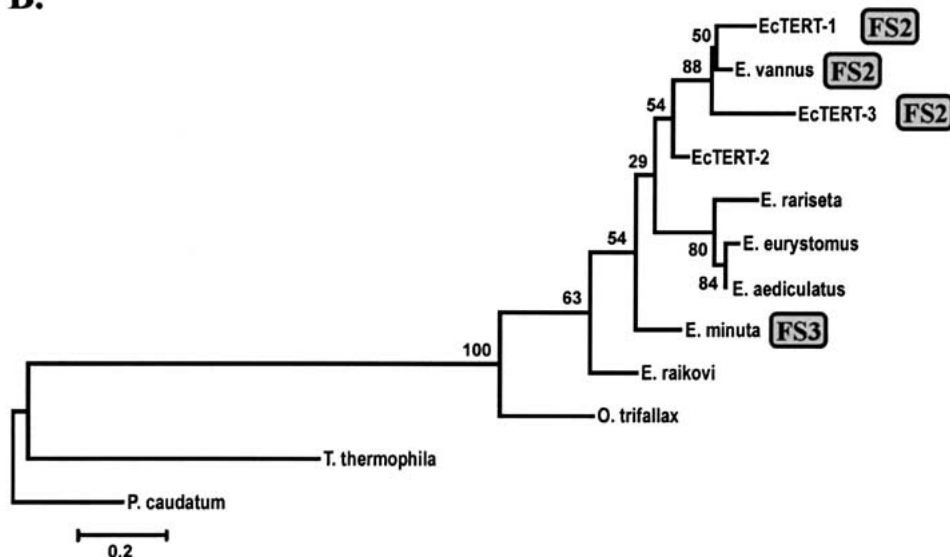


Fig. 4. **A** Sequence similarity of the three *E. crassus* TERT gene coding regions. Plots of sequence identity versus position in the coding region are shown for all pairwise combinations of the three *E. crassus* TERT gene coding regions. The analysis was performed with mVISTA (Bray et al. 2003; Dubchak et al. 2000; Mayor et al. 2000) using a 25-bp sliding window. A diagram of the *EcTERT-1*

coding region, with conserved domains and frameshift sites indicated, is also shown aligned with the *x*-axis. **B** Neighbor-joining tree constructed from the amino acid sequences corresponding to the *EcTERT-2*-specific region (corresponding to amino acids 621–694 of the *E. aediculatus* TERT protein). Bootstrap values are indicated and the scale bar denotes changes per residue.

Phylogenetic relationships among the euplotids are incompletely resolved. Classification schemes for the group have been based primarily on morphological characters (e.g., Borrer and Hill 1995; Gates and

Curds 1979), but more recent phylogenetic analyses based on 18S rRNA sequences are not entirely consistent with existing taxonomic schemes (Bernhard et al. 2001; Petroni et al. 2002). Nonetheless, our

analyses support a number of relationships inferred from the morphological and/or molecular data (Fig. 3). First, there is strong support in all trees for a close relationship between *E. eurytomus* and *E. aediculatus* (bootstrap values >84%). Second, there is a well-supported clade consisting of the *E. minuta*, *E. vannus*, and *E. crassus* TERT genes/proteins (bootstrap values all >85%; the relationship of the three *E. crassus* TERT genes is discussed further below). Third, for the most part, the *Euplotes* marine (*E. crassus*, *E. vannus*, *E. minuta*, and *E. raikovi*) and freshwater species (*E. eurytomus* and *E. aediculatus*) form separate clades within the *Euplotes* group. The single exception is the marine species *E. rariseta*, which grouped with the freshwater species in all but the minimum evolution DNA tree (Fig. 3 and data not shown). However, the bootstrap values associated with the placement of *E. rariseta* were uniformly low (<52%), so that the true position of this group is uncertain. The other problematical species is *E. raikovi*. In the Borrer and Hill (1995) classification scheme, *E. raikovi* is placed in a group separate from the other *Euplotes* species analyzed in the current study. The 18S rRNA gene analysis by Petroni et al. (2002) was consistent with this placement, as there was strong support for *E. raikovi* diverging the earliest among the *Euplotes* species analyzed. In contrast, in most of our analyses there is modest bootstrap support (66–83%) for *E. raikovi* forming a sister group with the lineage leading to the other marine *Euplotes* species (Fig. 3); only in the neighbor-joining tree produced from the TERT protein sequences did we observe *E. raikovi* as the earliest-branching euplotid, and even here there was low bootstrap support (42%).

In light of the phylogenetic analyses, the frameshift sites within the TERT genes appear to have arisen relatively late within the group. The newly identified FS3 site occurs only within the *E. minuta* TERT gene (Fig. 3), indicating that the site arose after this species diverged from the other euplotids. The situation for the FS2 site is somewhat more complex. The three genes containing FS2 (*E. crassus* macronuclear *EcTERT-1* and *EcTERT-3* and that of *E. vannus*), along with the *E. crassus* micronuclear-limited *EcTERT-2* gene, form a well-supported clade in all of the trees (Fig. 3; data not shown). Moreover, in both the protein-based maximum parsimony and minimum evolution trees, the three genes with FS2 group together with strong bootstrap support (Fig. 3; 99 and 76% bootstrap values, respectively), indicating that FS2 arose after the divergence of *E. minuta* and the *EcTERT-2* gene. In the other analyses either there is low support for the grouping of the three genes with FS2 or the *EcTERT-2* gene branches within this group, usually as a sister of *EcTERT-1* (e.g., Fig. 3B). As we discuss in the next section, it is likely that the

latter association is partially driven by gene conversion between the two loci. As a result, the most reasonable interpretation of the analyses is that FS2 arose in the lineage leading to the *EcTERT-1*, *EcTERT-3* and *E. vannus* genes. The close association of the *E. vannus* gene with those of *E. crassus* is not unexpected. While *E. crassus* and *E. vannus* are typically considered separate species that can be distinguished by both morphological and molecular criteria (e.g., Petroni et al. 2003; Valbonesi et al. 1992), a complex pattern of mating interactions has been observed between different strains of the two species (Caprette and Gates 1994), leading some to consider these organisms as a single species. Finally, we emphasize that although frameshift sites were limited to the TERT genes of the *E. crassus*–*E. vannus*–*E. minuta* species group, the phenomenon of genes requiring this particular form of +1 frameshifting for expression appears to be much more widely distributed among euplotids, as other genes with frameshift sites have been identified in *E. octocarinatus* (Caprette and Gates 1994) and *E. aediculatus* (Aigner et al. 2000).

Recent Origin of the Three *E. crassus* TERT Genes

Our interpretation of the origin of frameshift sites depends heavily on the conclusion that the three *E. crassus* TERT genes are of relatively recent origin. This aspect of the phylogenetic analyses was also surprising in light of the proposed functional differentiation of the TERT proteins (Karamysheva et al. 2003). The macronuclear *EcTERT-1* and *EcTERT-3* genes have been suggested to encode proteins that are involved in forming a telomerase that functions in maintaining telomere length during vegetative growth, while the micronuclear-limited *EcTERT-2* gene is thought to encode a protein that produces a telomerase with the ability to carry out *de novo* telomere synthesis, a process linked to the chromosome fragmentation process that occurs during macronuclear development. Chromosome fragmentation and *de novo* telomere formation are known to occur in a number of ciliates besides *Euplotes* (e.g., *Tetrahymena*, *Paramecium*, *Oxytricha*, *Stylonychia* [see Jahn and Klobutcher 2002; Yao et al. 2002]), so one might expect that this specialized form of telomerase arose early in ciliate evolution. Such a gene/protein would be expected to form a separate branch early in the TERT gene/protein trees, rather than the observed late branch for *EcTERT-2* that is closely associated with the other two *E. crassus* TERTs (Fig. 3).

Considering the significance of the placement of *EcTERT-2*, additional analyses were performed. Karamysheva et al. (2003) have noted that the *Ec*-

TERT-1 and *EcTERT-2* DNA sequences are 98.2% identical overall, with the majority of sequence differences located between the regions coding for conserved motifs A and B' (see Fig. 1). An mVISTA analysis (Bray and Pachter 2003; Dubchak et al. 2000; Mayor et al. 2000), comparing DNA sequence identity over a 25-bp sliding window, was performed for all pairwise combinations of the *E. crassus* *TERT* gene coding regions (Fig. 4A). With the exception of their termini, the comparisons of *EcTERT-1* with *EcTERT-3* and *Ec-TERT-2* with *EcTERT-3* show significant sequence variation throughout the coding regions. In contrast, the comparison of *EcTERT-2* with *EcTERT-1* reveals that the sequences are nearly identical throughout their length with the exception of an internal region corresponding to amino acids 619–692 of the EcTERT-1 protein (nucleotides 1955–2177 of the *EcTERT-1* macronuclear chromosome). Outside of this region there is a total of 10 nucleotide differences, and of these, 6 represent synonymous changes. Thus, these two genes appear to have participated in recent genetic exchanges, such as gene conversion. Moreover, if the EcTERT-2 protein does indeed form a telomerase with the capability of *de novo* telomere synthesis, the amino acid residues in the “variable” interval are likely to be the primary determinants of this specialized function.

With regard to the phylogenetic reconstructions, the previous analyses (Fig. 3) included regions of the genes/proteins that have likely undergone genetic exchange. As these regions of high sequence identity may have been responsible for the observed association of *EcTERT-2* with the other two *E. crassus* *TERT* genes, particularly *EcTERT-1*, additional trees were constructed using only the sequences corresponding to the *EcTERT-2*-specific region (1961–2182 and 621–694 of the *E. aediculatus* *TERT* macronuclear chromosome and protein, respectively). The DNA-based trees had few statistically supported lineages, likely owing to the shorter stretches of sequence analysed. The protein-based trees were also less robust, but there was support for a number of groupings (Fig. 4B and data not shown). All three tree-construction methods placed EcTERT-1, EcTERT-3, and *E. vannus* in a strongly supported clade (bootstrap values of 86–88%), which is consistent with a unique origin for FS2 late in the euplotid lineage. In addition, EcTERT-2 was consistently placed as a sister group to the FS2-containing genes, albeit with more modest bootstrap support (54, 59, and 39% in the neighbor-joining, maximum parsimony, and minimum evolution trees, respectively). The results thus suggest that the three *E. crassus* *TERT* genes arose late in the evolution of the euplotids, with *EcTERT-2* probably arising by gene duplication just prior to the divergence of *E. crassus* and *E. vannus*. Moreover, if EcTERT-2 is indeed involved in forming a telomerase specialized

for *de novo* telomere addition, the results suggest that this innovation might be limited to only some *Euplotes* species. We note that aside from the work in *E. crassus* (Karamysheva et al. 2003), there have been no reported systematic attempts in other ciliates to isolate micronuclear-limited *TERT* genes. Work of this type, along with biochemical analyses, will be necessary to determine if specialized *TERT* subunits of telomerase are indeed responsible for *de novo* telomere addition in a broad range of ciliates.

Origin of Frameshift Sites

For two of the previously identified *Euplotes* genes harboring frameshift sites, it has been possible to compare their DNA sequences with nonframeshifted *Euplotes* homologue to infer the nature of the DNA change responsible for producing the frameshift site (Jahn et al. 1993; Tan et al. 2001b). In each of these instances, it appears that insertion of a single base into a sequence that resembled the *Euplotes* frameshift sequence was responsible for generating the site (e.g., a change from AAA–AAA to AAA–TAA–A). Based on the alignment of the euplotid *TERT* gene sequences, a simple insertion of this type appears to have created the new frameshift site (FS3) that we identified in the *E. minuta* *TERT* gene (Fig. 2B). Under the assumption that the +1 frameshifting process in *Euplotes* is efficient, these types of single base mutations can easily be rationalized. While the nucleotide insertion results in a shift in reading frame that would typically produce a protein with altered sequence and impaired function, the same insertion creates the frameshift signal that allows the ribosome to shift the reading frame, restoring the original amino acid sequence.

The FS2 site of the *TERT* genes presents a somewhat different scenario. The alignment of *TERT* gene sequences indicates that an insertion occurred 13 bases upstream of the AAA–TAA motif (Fig. 2A). In this case, the gene likely harbored an out-of-frame AAA–TAA motif, which was brought into frame by the upstream one base insertion. Assuming that the translational frameshift occurs at the AAA–TAA motif, this situation results in a change in the amino acid sequence of the *TERT* protein between the insertion site and the frameshift motif (Fig. 2A). This significant change in amino acid sequence is likely tolerated, as it involves a poorly conserved region of the *TERT* protein (see Fig 1; Lingner et al. 1997). Alternatively, the presence of multiple *TERT* genes in *E. crassus* may have allowed for a multistep process. That is, *TERT* gene duplication may have initially generated a copy that was nonessential. The nonessential copy would have begun to accumulate mutations, including a base insertion that resulted in a frameshifted and nonfunctional protein. Then sub-

sequent base substitutions created the AAA–TAA motif that restores the original reading frame.

In summary, the *TERT* gene data indicate that programmed +1 frameshift sites have originated within euplotid genes during the recent evolution of the group. This may be an indication that this type of frameshift is particularly efficient, such that protein levels are unaffected. If so, random base insertions that generate an AAA–TAA motif, or that bring such a motif into the proper reading frame, may be selectively neutral. Studies of additional frameshift genes in multiple *Euplotes* species will be useful in assessing if this is indeed the case. Such studies may also prove informative in regard to whether frameshifting serves a regulatory role in the euplotids. In this regard, it will be interesting to identify frameshift genes in early diverging *Euplotes* species and then determine if the frameshift is conserved in more recently branching species, which would be expected if the frameshift was involved in a beneficial regulatory function.

Acknowledgments. We would like to thank Jan Postberg for help with the phylogenetic analyses, Dorothy Shippen for helpful discussions and for providing unpublished data, and Sara Avatapalli for technical assistance. This work was supported by grants from the National Science Foundation (MCB-9816765) and The Robert Leet and Clara Guthrie Patterson Trust to L.A.K. and a University of Connecticut Health undergraduate fellowship to M.C.G.

References

- Aigner S, Lingner J, Goodrich KJ, Grosshans CA, Shevchenko A, Mann M, Cech TR (2000) *Euplotes* telomerase contains an La motif protein produced by apparent translational frameshifting. *EMBO J* 19:6230–6239
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Baird SE, Klobutcher LA (1991) Differential DNA amplification and copy number control in the hypotrichous ciliate *Euplotes crassus*. *J Protozool* 38:136–140
- Baranov PV, Gesteland RF, Atkins JF (2002) Release factor 2 frameshifting sites in different bacteria. *EMBO Rep* 3:373–377
- Bernhard D, Stechmann A, Foissner W, Ammermann D, Hehn M, Schlegel M (2001) Phylogenetic relationships within the class Spirotrichea (Ciliophora) inferred from small subunit rRNA gene sequences. *Mol Phylogenet Evol* 21:86–92
- Borrer AC, Hill BF (1995) The order Euplotida (Ciliophora): Taxonomy with division of *Euplotes* into several genera. *J Eukaryot Micro* 42:457–466
- Bray N, Dubchak I, Pachter L (2003) AVID: A global alignment program. *Genome Res* 13:97–102
- Bryan TM, Sperger JM, Chapman KB, Cech TR (1998) Telomerase reverse transcriptase genes identified in *Tetrahymena thermophila* and *Oxytricha trifallax*. *Proc Natl Acad Sci USA* 95:8479–8484
- Caprette CL, Gates MA (1994) Quantitative analysis of interbreeding in populations of *vannus*-morphotype *Euplotes*, with special attention to the nominal species *E. vannus* and *E. crassus*. *J Eukaryot Micro* 41:316–324
- Chandler M, Fayet O (1993) Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* 7:497–503
- Doak TG, Witherspoon DJ, Jahn CL, Herrick G (2003) Selection on the genes of *Euplotes crassus* *Tec1* and *Tec2* transposons: Evolutionary appearance of a programmed frameshift in a *Tec2* gene encoding a tyrosine family site-specific recombinase. *Eukaryot Cell* 2:95–102
- Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA (2000) Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res* 10:1304–1306
- Gates MA, Curds CR (1979) The dargyrome of the genus *Euplotes*. *Bull Br Mus Nat Hist (Zool)* 35:127–200
- Ivanov IP, Gesteland RF, Atkins JF (2000) Antizyme expression: A subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. *Nucleic Acids Res* 28:3185–3196
- Jahn CL, Klobutcher LA (2002) Genome remodeling in ciliated protozoa. *Annu Rev Microbiol* 56:489–520
- Jahn CL, Doktor SZ, Frels JS, Jaraczewski JW, Krikau MF (1993) Structures of the *Euplotes crassus* *Tec1* and *Tec2* elements: Identification of putative transposase coding regions. *Gene* 133:71–78
- Karamysheva Z, Wang L, Shrode T, Bednenko J, Hurley LA, Welch J, Shippen DE (2003) A family of *TERT* genes in *Euplotes crassus* whose expression is controlled by ribosomal frameshifting and developmentally programmed gene elimination. *Cell* 113:565–576
- Klobutcher LA, Farabaugh PJ (2002) Shifty ciliates: Frequent programmed translational frameshifting in euplotids. *Cell* 111:763–766
- Kumar S, Tamura K, Nei M (1994) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci* 10:189–191
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17:1244–1245
- Lingner J, Hughes TR, Shevchenko A, Mann M, Lundblad V, Cech TR (1997) Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276:561–567
- Lozupone CA, Knight RD, Landweber LF (2001) The molecular basis of nuclear genetic code change in ciliates. *Curr Biol* 11:65–74
- Lynn DH, Small E (1997) A revised classification of the phylum ciliophora Doflein, 1901. *Rev Soc Mex Hist Nat* 47:65–78
- Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I (2000) VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16:1046–1047
- Nakamura TM, Morin GB, Chapman KB, Weinrich SL, Andrews WH, Lingner J, Harley CB, Cech TR (1997) Telomerase catalytic subunit homologs from fission yeast and human. *Science* 277:955–959
- Petroni G, Dini F, Verni F, Rosati G (2002) A molecular approach to the tangled intrageneric relationships underlying phylogeny in *Euplotes* (Ciliophora, Spirotrichea). *Mol Phylogenet Evol* 22:118–130
- Petroni G, Rosati G, Vannini C, Modeo L, Dini F, Verni F (2003) *In situ* identification by fluorescently labeled oligonucleotide probes of morphologically similar, closely related ciliate species. *Microb Ecol* 45:156–162
- Qian L, Wilkinson M (1991) DNA fragment purification: Removal of agarose 10 minutes after electrophoresis. *Biotechniques* 10:736–738
- Stahl G, McCarty GP, Farabaugh PJ (2002) Ribosome structure: Revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci* 27:178–183
- Tan M, Heckmann K, Brunen-Nieweler C (2001a) Analysis of micronuclear, macronuclear and cDNA sequences encoding the regulatory subunit of cAMP-dependent protein kinase of

- Euplotes octocarinatus*: Evidence for a ribosomal frameshift. *J Eukaryot Micro* 48:80–87
- Tan M, Liang A, Brunen-Nieweler C, Heckmann K (2001b) Programmed translational frameshifting is likely required for expressions of genes encoding putative nuclear protein kinases of the ciliate *Euplotes octocarinatus*. *J Eukaryot Micro* 48:575–582
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Tourancheau AB, Tsao N, Klobutcher LA, Pearlman RE, Adoutte A (1995) Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J* 14:3262–3267
- Valbonesi A, Ortenzi C, Luporini P (1992) The species problem in a ciliate with a high multiple mating type system, *Euplotes crassus*. *J Eukaryot Micro* 39:45–54
- Wang L, Dean SR, Shippen DE (2002) Oligomerization of the telomerase reverse transcriptase from *Euplotes crassus*. *Nucleic Acids Res* 30:4032–4039
- Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* 9:3353–3362
- Yao MC, Duharcourt S, Chalker DL (2002) Genome-wide rearrangements of DNA in ciliates. In: Craig N, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington, DC