Phylogenetic relationships amongst tetrahymenine ciliates inferred by a comparison of telomerase RNAs

Amanda J. Ye and Daniel P. Romero

**INTRODUCTION**

Generally accepted classification schemes of distantly related taxa within the phylum *Ciliophora* have been based upon the considerable morphological, physiological and behavioural data compiled in the 100 years between 1880 and 1980 (Corliss, 1979). Phylogenetic relationships were often derived by the comparison of detailed characteristics, such as the ciliary membrane ultrastructural detail revealed by silver impregnation and freeze-fracture techniques (Bardele, 1981; Corliss, 1973). By the 1980s, comparative analyses of evolutionarily conserved molecules often confirmed relationships that had been deduced from the morphological data (Greenwood et al., 1991a). In cases where the molecular and morphological data have led to different conclusions, comparisons of multiple molecular characters have often prevailed in establishing generally accepted phylogenetic relationships (Baroin-Tourancheau et al., 1992).

It has been extremely difficult to distinguish species in the suborder *Tetrahymenina*, order *Hymenostomatida*, based on their morphology. These organisms have apparently been constrained to maintain their overall ‘design’ for millions of years, being particularly well suited to an ecological niche that has changed little since they first evolved. The relatively narrow limits of scale, form, nutrition and karyotype for the tetrahymenine ciliates are responsible for the morphological constancy exhibited by the species in this suborder, following their divergence from a common ancestor 30–40 million years ago (Nanney et al., 1998; Van Bell, 1985a).

 Whereas tetrahymenine species are virtually indistinguishable at the morphological level, members of this suborder exhibit a high degree of divergence at the molecular level, reflecting their ancient origin. This is evident in the high frequency of intraspecific variation revealed by isoenzyme analysis (Jerome & Lynn, 1996). The comparison of rRNA sequences has proven to be particularly useful in inferring phylogenetic relationships, due to their high percentage of invariant sequences and conserved secondary structures that make it possible to accurately align truly homologous nucleotides (James et al., 1989; Lane et al., 1985; Sogin et al., 1971). Comparative analyses of SS rRNA, 5.8S rRNA, self-splicing introns and portions of both large- and small-subunit rRNAs have formed the basis for tetrahymenine molecular phylogenies (Nanney et al., 1989, 1998; Preparata et al., 1989; Sogin et al., 1986; Van Bell, 1985b). The validity of these phylogenies has been largely substantiated by similar comparative analyses of mitochondrial DNA (Morin & Cech, 1988) and the histone H3II/H4II intergenic region (Brunk et al., 1990).

**Keywords:** tetrahymenine ciliates, telomerase, *Tetrahymena*, *Colpidium*, *Glaucoma*
Although the comparison of rRNAs has been particularly effective in determining phylogenetic relationships, the tetrahymenine rRNA genes have not diverged to the extent necessary to position some of the more closely related species within the suborder (Jerome & Lynn, 1996). Comparative analysis of an alternative molecular model that has evolved at a much faster rate than the relatively conservative rRNAs is required to resolve these taxonomic issues. The sequences being compared must differ sufficiently to provide many instances of sequence variation, but not to the extent that homologous residues cannot be aligned with confidence (Pace et al., 1989).

Telomeric DNA is synthesized by telomerase, an unusual ribonucleoprotein reverse transcriptase whose integral RNA molecule dictates the synthesis of telomeric repeats (reviewed by Collins, 1999). We have used telomerase RNA (TER) sequences as a model molecule to establish phylogenetic relationships within the suborder *Tetrahymenina*. The rate of change for TERs is approximately an order of magnitude greater than that of rRNAs (Romero & Blackburn, 1991). Despite their considerable primary sequence divergence, a parsimonious alignment of tetrahymenine TERs is possible due to their relatively short length (approx. 160 nt) and a conserved secondary structure (Romero & Blackburn, 1991) that has been validated by chemical and enzymic structure probing both in vitro and in vivo (Bhattacharyya & Blackburn, 1994; Zaug & Cech, 1995). An overall secondary structure model aids in establishing an accurate alignment (Pace et al., 1989), especially in light of the identification of many functional domains for this molecule (Tu et al., 1995, 1998; Gilley & Blackburn, 1999; Yu et al., 1990). A phylogeny of 35 vertebrate species based on TER sequences (ranging between 388 and 559 nt in length) is topologically consistent with that derived from the analysis of haemoglobin sequences (Chen et al., 2000), demonstrating the suitability of TERs in deducing phylogenetic relationships.

**METHODS**

**General methods.** Samples of *Colpidium campylum* and *Colpidium colpoda* were kindly provided by Dr Denis Lynn (University of Guelph, Canada). *Colpidium striatum* was purchased from Carolina Biological Supply and *Glaucoma* sp. (ME60q) was kindly provided by Dr Ellen Simon (University of Illinois, USA). *Glaucoma* sp. (ME60q) and *C. campylum* cultures were maintained in 1–2% PPYS (proteose peptone, yeast extract and sequestrene) at 30 °C in the presence of 1× PSF (penicillin, streptomycin, fungizone; Life Technologies) to prevent bacterial and fungal growth, as described previously (Yu & Blackburn, 1990). *C. colpoda* and *C. striatum* were grown in bacterized media as described for *Paramecium tetraurelia* cultures (McCormick-Graham & Romero, 1996). Total DNA was isolated by detergent lysis (Yu & Blackburn, 1990) and molecular techniques were as described by Sambrook et al. (1989). Total RNA was isolated by the guanidinium HCl lysis of whole cells as described by McCormick-Graham & Romero (1995). Oligonucleotides were radiolabelled at the 5’ terminus with T4 polynucleotide kinase and [γ-32P]ATP (specific activity 7000 Ci mmol⁻¹) as described by Sambrook et al. (1989).

**TER gene detection by Southern blot hybridization.** Total DNA isolated from the tetrahymenine species was digested with a variety of restriction enzymes prior to Southern blot analyses. Radiolabelled probes were synthesized by the incorporation of [γ-32P]dATP (specific activity 3000 Ci mmol⁻¹) during PCR amplification of cloned TER genes from either *Tetrahymena thermophila* (Greider & Blackburn, 1989) or *Glaucoma chattoni* (Romero & Blackburn, 1991) as described previously (McCormick-Graham & Romero, 1996). Southern blots, hybridized to the radiolabelled probes under low stringency conditions (McCormick-Graham & Romero, 1995), revealed cross-hybridization to single restriction fragments for all four species surveyed, independent of the restriction digest (data not shown). The *G. chattoni* TER probe was used to detect the *Glaucoma* sp. (ME60q) holome; the *T. thermophila* TER probe cross-hybridized to Southern blots of the three *Colpidium* species.

**Cloning and sequencing TER genes.** Preparative restriction digests of total DNA (30–40 µg) were resolved in 0.8% low-melting-point agarose gels. Regions corresponding to cross-hybridizing restriction fragments (as revealed by genomic Southern blot analyses) were excised and DNA was recovered from gel slices as described by McCormick-Graham & Romero (1995). The size-enriched DNA was ligated into the appropriate cloning site of the pUC118 polyclinker prior to transformation of a competent *Escherichia coli* strain as described by Sambrook et al. (1989).

Bacterial colonies were transferred onto Nytran filters (Schleicher & Schuell) and the size-selected libraries were screened with the identical radiolabelled probes and hybridization conditions used for Southern blot analyses. The restriction digests and Southern blot analyses of cross-hybridizing clones were culminated by subcloning the appropriate restriction fragments into the poly cloning sites of *E. coli* and the size-selected libraries from the four species were as follows: *C. striatum*, 2.0 kb *BglII* fragment; *C. campylum*, 1.2 kb *BclI* fragment; *C. colpoda*, 1.3 kb *Dral* fragment; *Glaucoma* sp. (ME60q), 2.7 kb *BclI* fragment. The complete sequences (both strands) of subclones were determined using the dideoxynucleotide termination method with Sequenase (USB) and [γ-32P]dATP (Amersham; specific activity 1000 Ci mmol⁻¹).

**RNA primer extensions.** Definition of TER 5’ termini by primer extension was as described by McCormick-Graham & Romero (1996). Approximately 0.2 pmol of a radiolabelled primer complementary to the templating region of tetrahymenine TER (GATTTTTGGGT; specific activity 4 mCi mmol⁻¹) was annealed to 20 µg total RNA prior to extension by AMV reverse transcriptase. Primer extension products were analysed by autoradiography following their resolution in a polyacrylamide sequencing gel.

**Phylogenetic analysis.** Functional and structural elements conserved for ciliate TERs facilitated a parsimonious alignment of 17 tetrahymenine TER sequences. [The online version of this paper (http://ijis.sgmjournals.org) contains a supplementary figure showing the sequence alignment.] This alignment was used for phylogenetic analysis by maximum-parsimony (MP) methods. PAUP* [Version 4.0b8] (Swofford, 2000) was used to calculate the evolutionary distances between pairs of nucleotide sequences, using the model derived by Hasegawa et al. (1985).
RESULTS AND DISCUSSION

TER genes from four tetrahymenine species [C. striatum, C. campylum, C. colpod and Glaucoma sp. (ME60q)] were identified and isolated from size-selected genomic libraries by virtue of cross-hybridization to either the T. thermophila or the G. chattoni homologue. Northern blot analysis of total RNA from these species indicated an approximate length of 160 nt for all four RNAs (data not shown). The 5’ termi were mapped by primer extension of an oligonucleotide complementary to the template region. The predicted 3’ termini coincide precisely with 5–8 consecutive thymidine residues, a typical Pol III transcription termination signal (Geiduschek & Tocchini-Valenti, 1988) and a feature shared by all known ciliate TERs (Greider & Blackburn, 1989; Lingner et al., 1994; McCormick-Graham & Romero, 1995, 1996; Romero & Blackburn, 1991; Shippen-Lentz & Blackburn, 1990).

In the absence of mating incompatibility data, the classification of C. campylum and C. striatum as distinct species has been based on morphological characteristics. The TER sequences for these two Colpidium species are identical, suggesting that their divergence from a common ancestor is a relatively recent event. The presence of sequence variations within the non-transcribed regions flanking the TER transcript (GenBank accession numbers AF417609 and AF417611) and genomic DNA restriction digest polymorphisms revealed by Southern blot analyses with a TER probe (data not shown) supports their classification as two distinct species. This situation is analogous to that of two closely related Tetrahymena species, T. thermophila and Tetrahymena malaccensis, whose TER sequences differ at a single nucleotide position (Romero & Blackburn, 1991). Given the identity of C. striatum and C. campylum TER sequences, only the C. striatum homologue was included in the phylogenetic analysis. [The online version of this paper (at http://ijs.sgmjournals.org) contains a supplementary figure showing the sequence alignment.]

The parsimonious alignment of homologous nucleotides of the four new tetrahymenine TER sequences with those of 13 previously published sequences (McCormick-Graham & Romero, 1995) was achieved by minimizing the number of gapped positions. TER secondary structural elements are well conserved amongst species from distantly related ciliates (Lingner et al., 1994; McCormick-Graham & Romero, 1995, 1996; Romero & Blackburn, 1991) and were helpful in the identification of homologous nucleotides (Fig. 1). The functionally important template and its boundary domains (Autexier & Greider, 1995, 1998; Yu et al., 1990) are absolutely conserved for all 17 species. Conserved nucleotides of unknown function are also evident, particularly in the loops and bulges of helix IV in the consensus secondary structure. There are also ambiguities where truly homologous nucleotides could not be identified with any degree of certainty. In particular, a portion of the sequence between helices I and II, and the unstructured nucleotides between helices III and I (Fig. 1), were excluded from the phylogenetic analysis. Similarity values derived from a pairwise comparison of homologous nucleotides from 10 of the 17 TER sequences are shown in Table 1. Similarities range between 60 and 94%, with the majority at less than 80% (Table 1), indicative of the relatively rapid divergence rate for TERs.

Although phylogenetic analyses of morphological and molecular characters indicate that Glaucoma and Colpidium are the closest relatives to the Tetrahymena lineage, the relatedness of these three genera to each other has been somewhat controversial. Both morphological and molecular data have led to classification schemes depicting the three as a monophyletic group. Corliss (1979) and Van Bell (1985b) propose that Colpidium is more closely related to Tetrahymena than to Glaucoma, whereas both Small & Lynn (1981) and Greenwood et al. (1991b) suggest that Colpidium is more closely allied with Glaucoma. Paraphyletic trees based on the large subunit rRNA D2 variable region from nearly 40 tetrahymenine species indicate that Colpidium is more closely related to some Tetrahyiena species than to Glaucoma (Nanney et al., 1998; Preparata et al., 1989). Greenwood et al. (1991b) contend that the relatively short lengths and small number of independently variable nucleotide positions in 5S and 5.8S rRNAs limit the ability to differentiate the placement of closely related taxa with any confidence.

http://ijs.sgmjournals.org
recapitulates those derived from similar analyses of generated a phylogenetic tree (Fig. 2) that largely of TER sequences based on the parsimonious alignment hymenine genera. Distance analysis of tetrahymenine might help to resolve the relatedness of the tetra-

hope was that a phylogenetic analysis of this molecule of similar length (Romero & Blackburn, 1991), our might not resolve the relative positions of Colpidium and Glaucoma to the Tetrahymena species (Fig. 2), an inspection of helix I (Fig. 1) may shed some light on the relatedness of these genera.

TER helix I represents a long-range base-pairing interaction, which effectively divides the folded RNA into two distinct domains. Because they define the overall structure of the molecule, long-range base-pairing interactions tend to be more conservative, evolving at a slower rate than stem–loop helices (Pace et al., 1989). A direct comparison of helix I from representative species of the three tetrahymenine genera is highlighted in Fig. 3. Whereas the Tetrahymena and Colpidium helix I consists of 5 consecutive and invariant G–C base pairs [calc. $\Delta G = -\Delta G$ kcal mol$^{-1}$] (32.6 kcal mol$^{-1}$)], the homologous structure in the Glaucoma TER consists of 7 base pairs [calc. $\Delta G = -\Delta G$ kcal mol$^{-1}$ (41.4 kcal mol$^{-1}$)], with co-variation at 4 of 14 nt. Based upon the composition of helix I for these tetrahymenine species, we speculate that Tetra-
hymena and Colpidium may prove to be more closely related to each other than either are to Glaucoma.

T. paravorax TER is approximately 11 nt shorter than TERs isolated from all other Tetrahymena species. One consequence of this gap is the absence of stem–loop II (Fig. 1) from the T. paravorax TER secondary structure (McCormick-Graham & Romero, 1995). To test whether the absence of these nucleotides unduly influences the position of T. paravorax in the phylogenetic tree, the 16 nt that participate in stem–loop II were excluded for all 17 species prior to analysis.

### Table 1. Similarities and nucleotide distances among tetrahymenine TERs

The upper right half of the table gives similarity values $H$ for pairwise comparisons of 10 of the 17 TER sequences included in the analysis. $H$ is defined according to Sogin et al. (1986) as $m/(n + u + g/2)$, where $m$ is the number of sequence positions with matching nucleotides in the two sequences, $u$ is the number of sequence positions with non-matching nucleotides and $g$ is the number of sequence positions that have a gap in one sequence opposite a nucleotide in the other sequence. The absolute numbers of base substitutions and gapped positions (in parentheses) are shown in the lower half of the Table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrahymena thermophila</td>
<td>–</td>
<td>0.894</td>
<td>0.915</td>
<td>0.763</td>
<td>0.847</td>
<td>0.690</td>
<td>0.720</td>
<td>0.706</td>
<td>0.653</td>
<td>0.671</td>
</tr>
<tr>
<td>Tetrahymena pyriformis</td>
<td>15 (1)</td>
<td>–</td>
<td>0.932</td>
<td>0.793</td>
<td>0.851</td>
<td>0.694</td>
<td>0.711</td>
<td>0.704</td>
<td>0.653</td>
<td>0.669</td>
</tr>
<tr>
<td>Tetrahymena borealis</td>
<td>12 (1)</td>
<td>10 (0)</td>
<td>–</td>
<td>0.796</td>
<td>0.837</td>
<td>0.729</td>
<td>0.739</td>
<td>0.746</td>
<td>0.655</td>
<td>0.681</td>
</tr>
<tr>
<td>Tetrahymena pigmentosa</td>
<td>28 (11)</td>
<td>25 (9)</td>
<td>24 (10)</td>
<td>–</td>
<td>0.851</td>
<td>0.640</td>
<td>0.676</td>
<td>0.690</td>
<td>0.626</td>
<td>0.645</td>
</tr>
<tr>
<td>Tetrahymena hegewischi</td>
<td>19 (6)</td>
<td>19 (5)</td>
<td>21 (5)</td>
<td>19 (5)</td>
<td>–</td>
<td>0.657</td>
<td>0.706</td>
<td>0.683</td>
<td>0.626</td>
<td>0.638</td>
</tr>
<tr>
<td>Tetrahymena paravorax</td>
<td>41 (7)</td>
<td>41 (6)</td>
<td>36 (6)</td>
<td>42 (16)</td>
<td>43 (11)</td>
<td>–</td>
<td>0.644</td>
<td>0.678</td>
<td>0.598</td>
<td>0.617</td>
</tr>
<tr>
<td>Colpidium striatum</td>
<td>36 (8)</td>
<td>38 (7)</td>
<td>34 (7)</td>
<td>39 (13)</td>
<td>38 (8)</td>
<td>49 (5)</td>
<td>–</td>
<td>0.758</td>
<td>0.648</td>
<td>0.660</td>
</tr>
<tr>
<td>Colpidium colpoda</td>
<td>38 (8)</td>
<td>39 (7)</td>
<td>33 (7)</td>
<td>37 (13)</td>
<td>40 (10)</td>
<td>44 (5)</td>
<td>33 (4)</td>
<td>–</td>
<td>0.691</td>
<td>0.724</td>
</tr>
<tr>
<td>Glaucoma chattoni</td>
<td>43 (12)</td>
<td>45 (9)</td>
<td>42 (13)</td>
<td>42 (19)</td>
<td>44 (16)</td>
<td>50 (13)</td>
<td>41 (16)</td>
<td>35 (16)</td>
<td>–</td>
<td>0.942</td>
</tr>
<tr>
<td>Glaucoma sp. (ME60q)</td>
<td>41 (11)</td>
<td>42 (10)</td>
<td>39 (12)</td>
<td>40 (18)</td>
<td>43 (15)</td>
<td>48 (12)</td>
<td>40 (15)</td>
<td>31 (15)</td>
<td>8 (1)</td>
<td>–</td>
</tr>
</tbody>
</table>

Since TERs diverge at rates approximately one order of magnitude greater than those for rRNA molecules of similar length (Romero & Blackburn, 1991), our hope was that a phylogenetic analysis of this molecule might help to resolve the relatedness of the tetra-
hymenine genera. Distance analysis of tetrahymenine TER sequences based on the parsimonious alignment generated a phylogenetic tree (Fig. 2) that largely recapitulates those derived from similar analyses of rRNAs, mitochondrial DNA and the histone H3II/H4II intergenic region (Morin & Cech, 1988; Nanney et al., 1998; Sadler & Brunk, 1992; Sogin et al., 1986). Twelve Tetrahymena species are clustered into two distinct subtrees, represented by T. thermophila and Tetrahymena pigmentosa, while the singular Tetra-
hymena paravorax falls outside these two main branches. Although our phylogeny based on TER sequences does not resolve the relative positions of Colpidium and Glaucoma to the Tetrahymena species (Fig. 2), an inspection of helix I (Fig. 1) may shed some light on the relatedness of these genera.

[Fig. 2. A distance matrix phylogenetic tree based on TER sequences from the tetrahymenine ciliates. The consensus tree for 1000 bootstrap resamplings is shown, with the bootstrap supports indicated as a percentage at the base of each bifurcation. This is an unrooted tree.]
Despite the removal of stem–loop II from the formulation, there were no topological differences in the phylogenies shown in Fig. 2, consistently positioning _T. paravorax_ outside the main _Tetrahymena_ branch.

The distance between _T. paravorax_ and the rest of the _Tetrahymena_ species in the current study is consistent with other molecular phylogenetic analyses that have included this unusual species (Brunk et al., 1990; Nanney et al., 1998; Preparata et al., 1989; Sadler & Brunk, 1992). Despite its apparent distance, comparisons of macronuclear DNA size distributions clearly support the inclusion of _T. paravorax_ in the genus _Tetrahymena_, as it shares a size profile more similar to that of other _Tetrahymena_ species than to _Glaucoma_ (Brunk et al., 1990).

The rather extreme distance of _T. paravorax_ from 39 other tetrahymenine species in a phylogeny based on the large subunit rRNA D2 variable region was attributed to an anomaly, most likely due to violations of underlying assumptions of the analysis (Nanney et al., 1998). In light of phylogenies based on the histone H3II/H4II intergenic sequence and TERs, the apparent distance described by Nanney et al. (1998) is more likely to be an accurate reflection of the relatedness of _T. paravorax_ to the other 39 members of the suborder, which included two _Colpidium_ species. It would be interesting to see if the relative positions of _Colpidium_ to _Glaucoma_ and _T. paravorax_ can be resolved by a similar analysis of other conserved molecules. In particular, the inclusion of _Colpidium_ species in a re-evaluation of phylogenies derived from the histone H3II/H4II intergenic sequence data (Brunk et al., 1990; Sadler & Brunk, 1992) could be quite revealing.

**ACKNOWLEDGEMENTS**

This research was supported by Public Health Service grant GM-50861 from the National Institutes of Health (D.P.R.).

**REFERENCES**


Sogin, M. L., Ingold, A., Karlok, M., Nielsen, H. & Engberg, J. (1986). Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major Tetrahymena groups. EMBO J 5, 3625–3630.


