ARTICLE
Journal of Cellular Biochemistry 115:2077–2088 (2014)

Journal of Cellular Biochemistry

Variability in Secondary Structure of 18S Ribosomal RNA as Topological Marker for Identification of *Paramecium* Species

Farah R. Shakoori,[1] Fareeda Tasneem,[1] K. Al-Ghanim,[2] S. Mahboob,[2] F. Al-Misned,[2] Nusrat Jahan,[3] and Abdul Rauf Shakoori[2,4]*

[1]*Department of Zoology, University of the Punjab, Quaid-I-Azam Campus, Lahore 54590, Pakistan*

[2]*Department of Zoology, College of Science, King Saud University, P.O. Box 24555, Riaydh 11451, Kingdom of Saudi Arabia*

[3]*Department of Zoology, Government College University, Lahore, Pakistan*

[4]*School of Biological Sciences, University of the Punjab, Quaid-I-Azam Campus, Lahore 54590, Pakistan*

ABSTRACT

Besides cytological and molecular applications, *Paramecium* is being used in water quality assessment and for determination of saprobic levels. An unambiguous identification of these unicellular eukaryotes is not only essential, but its ecological diversity must also be explored in the local environment. 18SrRNA genes of all the strains of *Paramecium* species isolated from waste water were amplified, cloned and sequenced. Phylogenetic comparison of the nucleotide sequences of these strains with 23 closely related *Paramecium* species from GenBank Database enabled identification of *Paramecium multimicronucleatum* and *Paramecium jenningsi*. Some isolates did not show significant close association with other *Paramecium* species, and because of their unique position in the phylogenetic tree, they were considered new to the field. In the present report, these isolates are being designated as *Paramecium caudatum pakistanicus*. In this article, secondary structure of 18SrRNA has also been analyzed as an additional and perhaps more reliable topological marker for species discrimination and for determining possible phylogenetic relationship between the ciliate species. On the basis of comparison of secondary structure of 18SrRNA of various isolated *Paramacium* strains, and among *Paramecium caudatum pakistanicus*, *Tetrahymena thermophila*, *Drosophila melanogaster*, and *Homo sapiens*, it can be deduced that variable regions are more helpful in differentiating the species at interspecific level rather than at intraspecific level. It was concluded that V3 was the least variable region in all the organisms, V2 and V7 were the longest expansion segments of *D. melanogaster* and there was continuous mutational bias towards G.C base pairing in *H. sapiens*. J. Cell. Biochem. 115: 2077–2088, 2014.

KEY WORDS: RIBOTYPING; SSrRNA; SECONDARY STRUCTURE OF 18SrRNA; *Paramecium* spp; PHYLOGENETIC RELATIONSHIP OF CILIATES

Small subunit ribosomal rRNA are ancient structures which have been most frequently used as molecular clocks for the phylogenetic studies [Srivastava et al., 2011; Xia et al., 2011; Chen and Zhang, 2012; Yu et al., 2013]. Conservation of ribosomal RNA sequences in the course of evolution is very important to maintain three dimensional structure [Hofacker et al., 2002; Ueno et al., 2007; Letsch et al., 2010]. Ribosomal RNA molecule folds to form an extensive base-paired secondary structure as part of the formation and function of ribosome in cellular protein synthesis [Voigt et al., 2008]. Due to the slowly evolving property of 18S rRNA sequences, the construction of their secondary structures has proven to be a useful topographical marker. These markers are being used for inferring better understanding of phylogenetic relationships among organisms as well as their functional domains in ribosomes [Hwang et al., 2000].

Primary nucleotide sequences of ribosomal RNA can be folded into secondary structural models through Watson–Crick and wobble base-pairing interactions between the bases. Hydrogen bonding between the canonical (A.U; G.C), non canonical (G.U) and intermediate unstable pairings (A.C) result in the formation of helices whereas other

* Correspondence to: Prof. Dr. Abdul Rauf Shakoori, Distinguished National Professor & Director, School of Biological Sciences, University of the Punjab, Quaid-i-Azam Campus, Lahore 54590, Pakistan. E-mail: arshaksbs@yahoo.com; arshakoori.sbs@pu.edu.pk

Manuscript Received: 24 May 2014; Manuscript Accepted: 10 July 2014

Accepted manuscript online in Wiley Online Library (wileyonlinelibrary.com): 15 July 2014

DOI 10.1002/jcb.24885 • © 2014 Wiley Periodicals, Inc.

non pairing regions form the contiguous structures of hairpin loops, terminal and lateral bulges [Varriale et al., 2008; Weigand et al., 2012]. RNA secondary structures can be divided into structurally "conserved" and "variable" regions based on documenting the mutational patterns responsible for the variation in the double helical regions, loops or whole stems [Xia et al., 2003]. The large portion of the ribosomal RNA exhibit conserved base pairing patterns which provides an objective criterion to improve the alignment based on structural information and the tree reconstruction process of ribosomal RNA data sets. Both these criterions have been shown to be helpful in developing the accuracy of phylogenetic affiliations within various taxa [Marande et al., 2009].

Despite the fact that analysis of rRNA secondary structure models has proven to be a powerful tool for establishing accurate phylogenetic information, they are rarely being used in phylogenetic analysis. The major reason is that the formation of a secondary structure model is time consuming process, and there are very few software packages which allow the simultaneous analysis of paired and unpaired rRNA regions [Voigt et al., 2008]. However, ribosomal RNA secondary structure models have been reported in literature by a number of authors for various organisms [Strüder-Kypke et al., 2001; Yi et al., 2009; Lee and Gutell, 2012; Gao et al., 2012, 2013; Zhang et al., 2012] although no consensus model has been attained in the case of *Paramecium* species.

Paramecium is a genus of ciliate, belonging to the class Oligohymenophorea and subclass Peniculita, and is known since the beginning of microscopy that is, about 250 years ago. Its large size (80–400 μm), ease of cultivation, and cosmopolitan distribution has made it a "universal laboratory animal" [Fokin, 2011]. Its taxonomy has long history and is still changing. Based on the cells shape and size, Woodruff [1921] divided the genus *Paramecium* into two clearly defined groups; the *aurelia* group and *bursaria* group. The members of the *aurelia* group viz. *P. aurelia*, *P. caudatum*, and *P. multimicronucleata*, are with long spindle- or cigar-shaped bodies; those of the bursaria group viz. *P. bursaria*, *P. calkinsi*, *P. putrinum*, and *P. trichium* are with shorter, broader and dorsoventrally flattened bodies. Then Jankowski divided the genus into three subgenera [*putrinum*, *woodruffi*, and *aurelia*] based on their morphological features [Strüder-kypke et al., 2000].

However, the genus *Paramecium* has been classified into four subgenera [Fokin et al., 2004]. The first subgenus *Chloroparamecium*, named for its symbiotic relationship with the green alga *Chlorella*, consists only of *P. bursaria*. This appears as basal node in the tree presenting all the species of the genus. The second subgenus Helianter includes *P. putrinum* and *P. duboscqui*, both of which branch off in the genus after *P. bursaria*. The third subgenus, *Cypriostomum* encompasses *P. woodruffi*, *P. nephridiatum*, *P. calkinsi* and *P. polycaryum*. This subgenus is placed in the middle of the genus tree. Finally, the fourth subgenus, named *Paramecium* due to the fact that all of its species have a cigar-shape, contains the following species: the *P. aurelia* species complex, *P. jenningsi*, *P. schewiakoffi*, *P. caudatum*, *P. multimicronucleatum*, and some rare species, namely, *P. wichtermani*, *P. africanum*, *P. jankowski*, and *P. ugandae* [Fokin et al., 2004; Hoshina et al., 2006].

Among the 15 known species of *Paramecium aurelia*; four species (*P. biaurelia*, *P. primaurelia*, *P. sexaurelia*, and *P. tetraurelia*) are cosmopolitan in distribution, five species [*P. septaurelia*, *P. decaurelia*, *P. undecaurelia*, *P. dodecaurelia*, and *P. sonneborniare*] are known from US, while *P. quadecaurelia* has been reported only from Australia [Przybos and Fokin, 2000; Przybos et al., 2002, 2003, 2005, 2008a,b, 2009, 2012]. *P. novaurelia* was described as limited only to Europe, later on this species was also reported from Asia (Turkey, Anatolian Upland). *P. decaurelia* was also known from USA (Florida) alone but later on this was also recorded in Japan [Przybos et al., 2007]. Likewise *P. pentaurelia* that was originally reported from North America, Australia and Europe but afterwards it was also recorded from Poland [Przybos et al., 2011].

Based on the random distribution of these species, the genus *Paramecium* could be considered larger than assumed as many of territories have still to be checked. One of the major reasons is unbalanced *Paramecium* species sampling throughout the world [Li et al., 2009]. Therefore it is necessary to investigate the rarely studied areas such as South America, Africa, and tropical Asia or, on the opposite, Arctic and Antarctic areas. Although European countries are much more investigated as compared to Asian regions but even there is continuous discovery of new species which brings idea that the list of *Paramecium* species is still incomplete. Hence, the addition of new species by more sampling is needed to resolve the phylogeny within the *Paramecium* genus [Strüder-kypke et al., 2000].

The 1990's have been regarded as the beginning of a new age in ciliate systematics [Lynn and Struder-Kypke, 2002]. An update of earlier classification of ciliates has started with the rRNA molecular analysis that has recently become widely accepted to resolve the problems of systematics as well as to reconstruct the phylogeny at the species and subspecies levels [Shang et al., 2003; Breiner et al., 2008]. The nucleotide sequences are functionally equivalent in all organisms and their genes are highly conserved; therefore phylogenies based on these sequences reveal such features of the tree of life that can not be resolved by using simple morphological approaches [Lukashenko, 2009].

Little work has been done on this genus throughout the world particularly in Asian regions; therefore, the additions of new species by increased sampling are needed to resolve the phylogeny within the *Paramecium* genus. The present study was aimed at isolating and purifying *Paramecium* species from water samples originating from various regions of the Punjab province, Pakistan. The 18S rRNA gene was used as a molecular marker for the characterization of these isolated species. Inferred from sequence information topological tree along with related genera obtained from GenBank was constructed to achieve a picture of phylogenetic relationship of these species within the genus *Paramecium*. Secondary structures based on the primary nucleotide sequences of 18S rRNA sequences of all isolated *Paramecium* species have been compared to derive some common criteria of phylogenetic relationships.

MATERIALS AND METHODS

CULTURING AND IDENTIFICATION OF PARAMECIA

Paramecium species were sampled from water bodies of different localities [Lahore, Kasur, Mansehra, Sheikhupura] of the Punjab Province, Pakistan. Paramecia were isolated and raised in a bold

basal salt medium [0.25 g/l $NaNO_3$, 0.025 g/L $MgSO_4 \cdot 7H_2O$, 0.075 g/L K_2HPO_4, 0.175 g/L KH_2PO_4, 0.025 g/L NaCl, 0.05 g/L EDTA, 0.031 g/L KOH, 0.04 g/L $FeSO_4 \cdot 7H_2O$, 0.001 M H_2SO_4, 0.01142 g/L H_3BO_3, 0.00881 g/L $ZnSO_4 \cdot 7H_2O$, 0.00144 g/L $MnCl_2 \cdot 4H_2O$, 0.00071 g/L MoO_3, 0.00157 g/L $CuSO_4 \cdot 5H_2O$ and 0.00049 g/L $Co(NO_3)_2 \cdot 6H_2O$], diluted 1:1000 with distilled water. The medium was bacterized the day before use with wheat grains by adjusting pH at 7.4–7.6 [Shakoori et al., 2004]. Initial identification of paramecia was done by observing its body shape, morphological features and movement patterns [Curds, 1982; Curds et al., 1983].

ISOLATION OF GENOMIC DNA

Genomic DNA was isolated by incubating *Paramecium* cells with 10 mM Tris-HCl [pH 7.5] for 10–12 h at 27 °C before harvesting them by low speed centrifugation at 6741x g at 7 °C for 10 min. The cells were lysed by using lysis buffer (42% urea, 0.30 M NaCl, 10 mM Tris-HCl [pH 7.5], 10 mM EDTA, 10% Nonidet P40 and 1% SDS). The lysis mixture was extracted twice with phenol:chloroform (1:1) by following Sambrook and Russell [1989] DNA extraction method.

PCR AMPLIFICATION

A set of primers Euk For 5′-AATATGGTTGATCCTGCCAGT-3′ and Euk Rev 5′-TGATCCTTCTGCAGGTTCACCTAC-3′ were used to amplify the conserved region of 18S rDNA as used by Regensbogenova et al. [2004]. PCR was performed in a final volume of 50 μL comprising 10 mM dNTPs; 5 μL $MgCl_2$; 3μL of 10x buffer, 1 μL 10 mM of each primer and 1 μL of 2.5 U Taq-polymerase (Fermentas, USA). PCR amplification protocol comprised one cycle of 5 min at 94 °C, followed by five cycles, each of one min denaturation at 94 °C, 2 min annealing at 43 °C, and 2 min extension at 72 °C, and 30 cycles, each of one min at 94 °C, 2 min at 50 °C and 2 min at 72 °C; and finally ending the cycle at 72 °C extension for 20 min.

CLONING AND SEQUENCING OF 18S rRNA GENE

After confirmation of appropriate size, each PCR product was purified using Nucleospin Extract 11 (Macherey–Nagel Germany) and were inserted into a pTZ57 R/T cloning vector (Fermentas #K1214). Plasmids were harvested by using Quiagen Kit (QIA prep spin) and sent to Macrogen (South Korea) for sequencing. Subsequent sequencing was performed in both directions using primer walking.

PHYLOGENETIC ANALYSIS

The 18SrRNA gene sequences available from GenBank/EMBL databases [www.ncbi.nlm.nih.gov] with the following accession numbers: *Paramecium multimicronucleatum* [AB252006, AB252007, AB255361], *Paramecium tetraurelia* [EF502045, X03772, AB252009, AB252008, AF149979], *Paramecium jenningsi* [AF100311], *Paramecium schewiakoffi* [AJ54882], *Paramecium primaurelia* [AF100315], *Paramecium caudatum* [AF217655, AB252005, AB252004, AB252003], *Paramecium duboscqui* [HM140398, AM236094, AF100312], *Paramecium woodruffi* [AF255362], *Paramecium nephridiatum* [AF100317, AF100316], *Paramecium putrinum* [AF255360], *Paramecium bursaria* [AB252000, AB252001, AB252002, AB219526, AB206543, AB206542, AB206541, AB206540, AB206539, AB206538, AB206537], *Paramecium calkinsi* [AF100301, AF100310], *Paramecium bursaria* [AB206545, AF100314] and *Paramecium polycaryum* [AF100313], were compared with 18SrRNA gene sequences of isolates of present study. The closely related species from GenBank Database were used to align with all isolated *Paramecium* species by using Megaline alignment programme. The phylogenetic tree was constructed to ascertain the relationship of all isolated *Paramecium* species by using Clustal X.

ANALYSIS OF SECONDARY STRUCTURES OF 18S rRNA

The secondary structure of 18S rRNAs of all *Paramecium* species obtained through a RNA structure programme (http://rna.urmc.rochester.edu/RNAstructure.html) were used as a criterion for phylogenetic relationship among different species of *Paramecium*. Analysis of RNA models was based on documenting the mutational patterns responsible for variation in the double helical regions, loops or whole stems. Models were analyzed by comparing the structures of the same species (viz., FT1 with FT2, FT4, FT5, and FT6 with FT7, FT8) or closely related species (FT2 with FT3). However, FT6 was used to compare the secondary structures of *T. thermophila*, *D. melanogaster*, and *H. sapiens* to determine conserved homologous regions. The nucleotides were numbered after every 50 nucleotides. Terminologies of Ouvrard et al. [2000] were used to explain secondary structures. Briefly, the glossary adopted was as follows: Stem, right handed double helix composed of a succession of complementary hydrogen bonded nucleotides between paired strands; Single strand, unpaired nucleotides separating strands; Hairpin loop, stem closed distally by a loop of unpaired nucleotides; Terminal bulge, succession of unpaired nucleotides at the end of a hairpin loop; Lateral bulge, succession of unpaired nucleotides on one strand of a stem; Internal bulge, group of nucleotides from the two opposite strands unable to form canonical base pairs; Junction or multibranched loop, succession of group of unpaired nucleotides joining the last proximal pairing of several stems; Compensatory base to maintain structure, mutation on one strand of a stem following initial mutation of a complementary substitution; Semi-compensatory substitution, mutation of a base on one strand not affecting structure (A.U or G.C replaced by G.U). The mismatched nucleotides were highlighted by using different colors.

RESULTS AND DISCUSSION

MORPHOLOGICAL IDENTIFICATION

All strains were recognized as members of the *Paramecium aurelia* group, since they had slipper shaped body rounded at the front and pointed at the back (Supplementary Fig. 1). Slipper shaped body is one of the distinguishing characteristics between *aurelia* and *bursaria* group of the genus *Paramecium*.

MOLECULAR IDENTIFICATION/RIBOTYPING

PCR amplification reactions of all eight strains yielded a fragment of approximately 1.8 Kb. These 18SrDNA fragments were cloned in pTZ57R/T vector and were used for transformation of *E. coli* DH5α cells. 18SrDNA genes of all strains were nucleotide sequenced. Initial sequencing was based on M13 forward and M13 reverse primers. The

internal primers of small subunit rDNA were designed on the basis of initial results (primer walking) for the final sequencing. Sequences were verified for identity by comparing with the 18SrRNA sequences of *Paramecium* species available in GenBank. The percentage similarities of our results with GenBank sequences divided our sequences into two groups. FT1, FT2, FT4, and FT5 showed maximum similarity (97.7–98.4%) with *Paramecium multimicronucleatum*, and FT4 showed 99.7% similartity with *Paramecium jenningsi*. FT6, FT7, and FT8 showed weaker (96.3–96.5) resemblance with *Paramecium multimicronucleatum* (Table I). The sequences were submitted to GenBank Database to get their accession numbers (Table II).

PHYLOGENETIC ANALYSIS

The homology values and the absolute number of mismatching nucleotides of all isolated *Paramecium* species sequences were calculated. The absolute number of mismatched nucleotides ranged from 4 to 91. The homology values showed the extent of variation from 0.945 to 1.00 (Table III).

According to the subgeneric classification, all *Paramecium* species of this study fell under the subgenus *Paramecium* based on 18S rDNA sequence analysis. Phylogenetic tree based on all the eight *Paramecium* sequences, along with their GenBank sequence entries, is shown in Figure 1, where three distinct clades are obvious. This was contrary to the conclusion derived from the percentage similarity of 18srRNA sequences as shown in Table I. First clade of this study appeared as a group of four species of *P. multimicronucleatum* (FT1, FT2, FT4, and FT5), the second one comprised *P. jenningsi* (FT3) and the third clade was closer to *P. caudatum* (FT6, FT7, and FT8). In fact phylogenetic comparison does not support the close association of the third clade with any other species of the present isolates. This group was quite distant from other species showing its separate position in the tree potentially indicating that this species is new to the field. It is being designated as *Paramecium*

TABLE I. Comparison of Multiple Alignments of 18SrRNA Sequences of Eight *Paramecium* Species With Each Other and With Previously Reported *Paramecium* Species Sequences

Name of organisms	FT1	FT2	FT3	FT4	FT5	FT6	FT7	FT8
FT1	***	99.8	96.6	99.7	99.4	95.1	94.9	95.1
FT2	***	***	96.9	99.8	99.3	95.1	94.9	95.0
FT3	***	***	***	96.8	96.4	96.3	96.1	96.3
FT4	***	***	***	***	99.4	95.2	94.7	94.8
FT5	***	***	***	***	***	94.9	94.5	94.6
FT6	***	***	***	***	***	***	99.0	99.3
FT7	***	***	***	***	***	***	***	99.1
FT8	***	***	***	***	***	***	***	***
P. multi- micronucleatum (3)	98.0–98.4	98.1–98.4	96.7–96.8	98.0–98.4	97.7–98.1	95.0–95.1	95.0–95.1	95.0–95.1
P. tetraurelia (4)	96.3–96.6	96.2–96.7	99.0–99.3	96.3–96.6	95.9–96.3	95.9–96.3	95.9–96.4	95.9–96.3
P. primaurelia (1)	96.6	96.8	99.3	96.6	96.4	96.3	96.2	96.2
P. jenningsi (1)	96.7	96.6	99.7	96.7	96.3	96.5	96.3	96.5
P. schewakoffi (1)	96.2	96.5	99.0	96.2	96.1	96.1	96.0	96.0
P. caudatum (4)	94.2–94.5	94.3–94.5	95.2–95.4	94.2–94.5	94.0–94.2	94.8–95.1	94.8–95.1	94.7–95.0
P. woodruffi (1)	94.2	94.3	94.3	94.2	94.0	93.8	93.8	93.7
P. nephridiatum (2)	94.1	94.3	94.1	94.1	93.9	93.6	93.7	93.5
P. polycarium (1)	92.7	92.8	93.2	92.7	92.6	93.1	92.7	93.0
P. duboscqui (2)	93.3–93.9	93.5–94.0	93.5–93.8	93.3–93.9	93.4–93.8	93.4–93.8	93.2–93.7	93.3–93.8
P. calkinsi (2)	93.2–93.3	93.2–93.5	93.6–93.8	93.3–93.3	92.9–93.3	92.8–93.0	92.9–93.0	92.7–93.0
P. putrinum (1)	93.6	93.8	93.5	93.6	93.5	92.8	93.1	92.7
P. bursaria (7)	92.2–92.4	92.6–93.0	92.1–92.6	92.2–92.4	92.2–92.9	92.2–92.3	92.1–92.2	92.1–92.2

TABLE II. Identification of Isolated *Paramecium* Species Based on Homology Analysis of 18S rRNA With Their Accession Numbers

Sr #	Strain designation	Species name	Geographic origins	Accession numbers
1	FT1	*P. multimicronucleatum*	Kasur	HE650906
2	FT2	*P. multimicronucleatum*	Kasur	HE650907
3	FT3	*P. jenningsi*	Lahore	HE662760
4	FT4	*P. multimicronucleatum*	Sheikhupura	HE662761
5	FT5	*P. multimicronucleatum*	Sheikhupura	HE662762
6	FT6	*P. caudatum*_Pakistan isolate*	Mansehra	HE662763
7	FT7	*P. caudatum*_Pakistan isolate*	Mansehra	HE662764
8	FT8	*P. caudatum*_Pakistan isolate*	Mansehra	HE662765

Paramecium caudatum pakistanicus

TABLE III. Homology and nucleotide distance data between all eight isolated *Paramecium* 18SrRNA sequences. The upper right half of the table gives homology values H for all pair-wise comparisons of *Paramecium* small subunit rRNA sequences. H= m/(m + u + gl2) 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 number of base changes between nucleotide sequences is shown in the lower half of the table

Organism	FT1	FT2	FT3	FT4	FT5	FT6	FT7	FT8
FT1		0.996	0.967	0.997	0.993	0.995	0.953	0.952
FT2	10		0.995	0.992	0.952	0.952	0.951	1.002
FT3	57	*58*		0.967	0.962	0.966	0.965	0.966
FT4	4	*12*	60		0.993	0.967	0.949	0.949
FT5	13	*18*	72	8		0.945	0.947	0.945
FT6	82	*89*	61	82	91		0.995	0.995
FT7	83	90	62	84	88	7		0.990
FT8	85	*91*	61	87	92	7	8	

caudatum pakistanicus. Based on morphological features, all these three species (*P. multimicronucleatum*, *P. jenningsi*, and *P. caudatum*) belong to the aurelia subgroup of the *Paramecium* genus.

SECONDARY STRUCTURE ANALYSIS

The secondary structures of 18S rRNA sequences of all isolated *Paramecium* species have been drawn on the basis of comparative criteria in this study. Conserved and variable nucleotide regions were identified using multiple alignments of the present study sequences with available sequences from GenBank (Supplementary Fig. 2a–g). The predicted models of all secondary structures folded into 44 helices along with eight expansion segments (V1–V5 and V7–V9). The numbering system of Nelles et al. [1984] has been followed in this study for identification of helices. Nearly all of the variations are contributed by fast evolving regions (V2, V4, and V7) that have been reported to be farthest from the functional region. V1, V5, V8, and V9 regions are less variable in size and are supposed to be closer to the functional region. V3 and V6 with least variability are found to be quite close to the functional region [Gagnon et al., 1996; Milyutina et al., 2001].

INTRASPECIFIC VARIATION IN *Paramecium multimicronucleatum* (FT1,2,4,5)

The predicted secondary structure models of all *Paramecium multimicronucleatum* (FT1, FT2, FT4, and FT5) were compared to resolve the phylogeny at intraspecific level. Despite four substitutional differences between FT1 and FT4 in the primary sequence, no effect was observed in their secondary structure. FT4 model therefore is not being discussed here. The overall secondary structures of all species showed same configuration except the helices 21–23 (Figs. 2 and 3). Helix 21 comprises V4 region along with six sub helices (E21_2, E21_5, E21_6, E21_7, E21_8, and E21_9). Helix 21 is seen branching off from the internal loop of helix 14 in FT1 as against the more profound single stable helix in FT2. In FT5 (Fig. 3), however, this helix appeared as a junctional bridge between the two internal loops. E21_2, E21_5 and E21_6 are the most conserved sub helices in V4 region among all *Paramecium* species. E21_7 appeared as a short stem with terminal hairpin loop (GGAAUA) among FT1 and FT2 (Figs. 2 and 3) whereas in FT5 (Fig. 3) this helix appeared as a connecting bridge between the two junctional loops. E21_8 (40 bp) folded into similar structural pattern among all species, except for the absence of an internal loop (AGG, UAA) in FT1 (32 bp) which is a part of coaxial junction loop. E21_9 appeared as an independent sub helix in FT2, whereas in all other species, this sub helix is connected with helices 22 and 23. The area spanning helices 22 and 23 in FT2 and helices 23–26 in FT5 is the most variable in terms of number of helices and the primary sequence within the helices. The interesting thing about this region is that, it is in the core segment area, beyond the hyper-variable V4 zone. In the case of FT2, helix 22 is shown starting as separate helix and extends several bases beyond the junction. This extension is designated as 22_1, as opposed to FT1 and FT5, where this is single stable helix with internal and terminal hairpin loops. Addition of bases at seven different sites and two substitutional differences resulted in splitting of helix 23 into two helices in FT2 as opposed to FT1. However, in the case of FT5, this helix is in junction with E21_7 of V4 zone where it failed to form an independent sub helix. Addition of four nucleotides and three semi-compensatory substitutions in FT5 are documented beyond helix 23. These additions and substitutions caused the length variation of this region which resulted in different structural motifs of 24–26 helices from the proposed models of FT1 and FT2. It can be concluded here that although there are no differences at the primary sequence level in the variable regions, but the additions and substitutions at core segment resulted in different structural configurations at V4 region. Similar conclusions were drawn by Zimmermann et al. [2011] who showed that variations at intraspecific level are very low and V4 region is not powerful enough to discriminate the cryptic species.

COMPARISON OF *Paramecium jenningsi* (FT3) AND *P. multimicronucleatum* (FT4)

Figure 4 shows interspecific variation between *P. jenningsi* and *P. multimicronucleatum*. Zone V1 (20 bp long) is well conserved in both species. Four substitutional changes were observed within terminal loop of stem 10 in V2 region (109 bp) without disturbing their secondary structure. V3 zone (54 bp) also appeared undisturbed, despite three substitutional mutations: one in internal and two in terminal loops. These helices prove the assumption proposed by Ouvrard et al. [2000] that substitutions are less important in bulges, compared to those in stems. In V4 region (305 bp) E21 appeared as one of the more variable regions with seventeen mutational differences and the addition of two base pairs in FT3. This sub

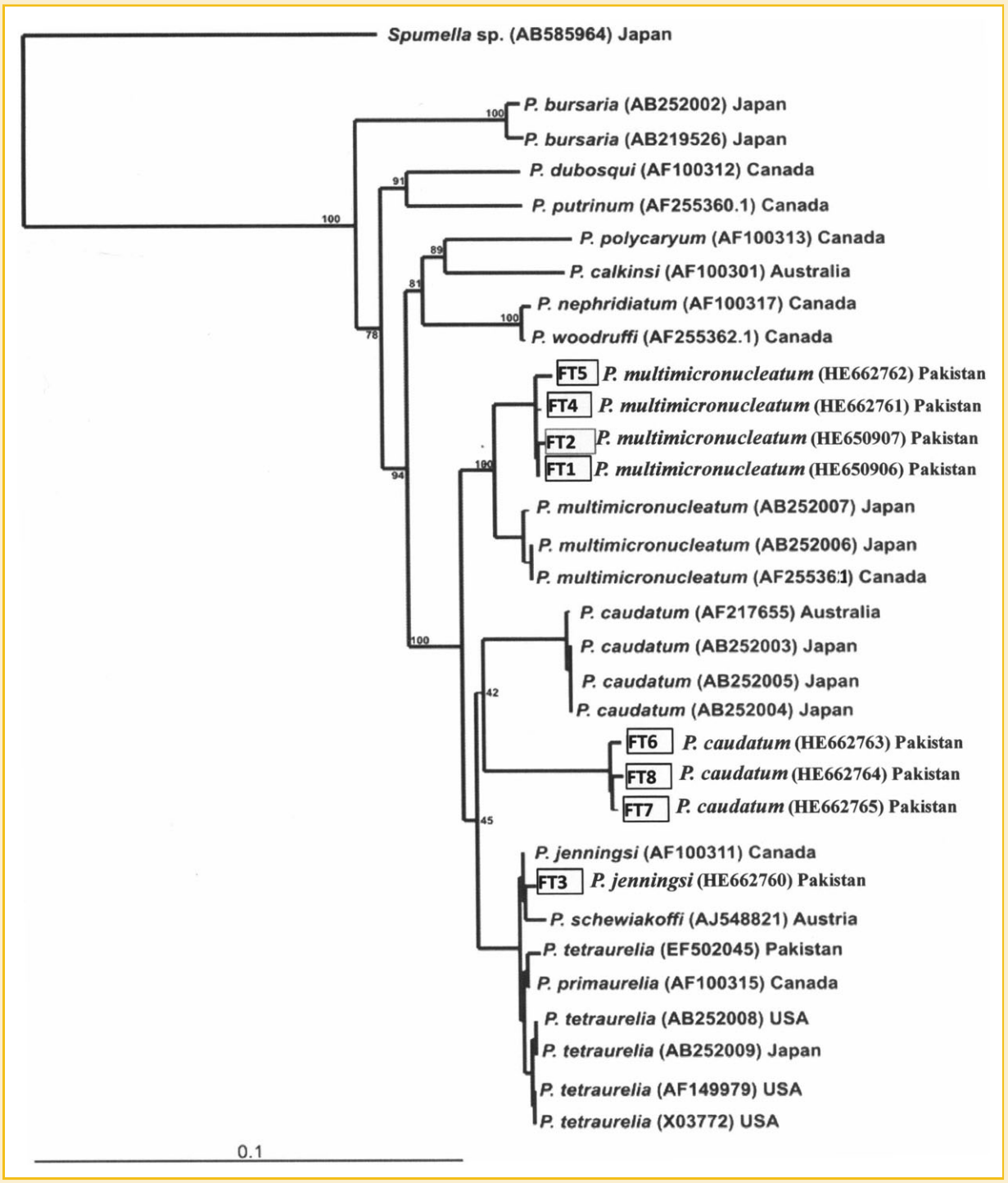

Fig. 1. Phylogenetic tree showing the relationship of eight *Paramecium* species with the previously reported sequences of different *Paramecium* species from GenBank Database. *Spumella* sp. is taken as an outgroup species.

helix is shorter in length (30 bp) in FT4, compared to its correspondence helix (36 bp) in FT5. This zone is hypervariable in terms of its primary structure but showed less variability in terms of its secondary structure. Four compensatory and twelve semi-compensatory changes have been observed in this helix. These substitutions resulted in the appearance of single nucleotide bulges at positions 624 (A), 629 (G), and 640 (A) in shorter stem of FT4. Addition of two nucleotides at positions 634 (C) and 635 (U) in this helix of FT3 moved the single nucleotide bulges of FT4 into the stem, facing their opposite nucleotides in such a manner that helped to form the canonical base pairs. It can be assumed that the formation of two Watson-Crick base pairs viz., C . G at 614 . 646 and G . C at

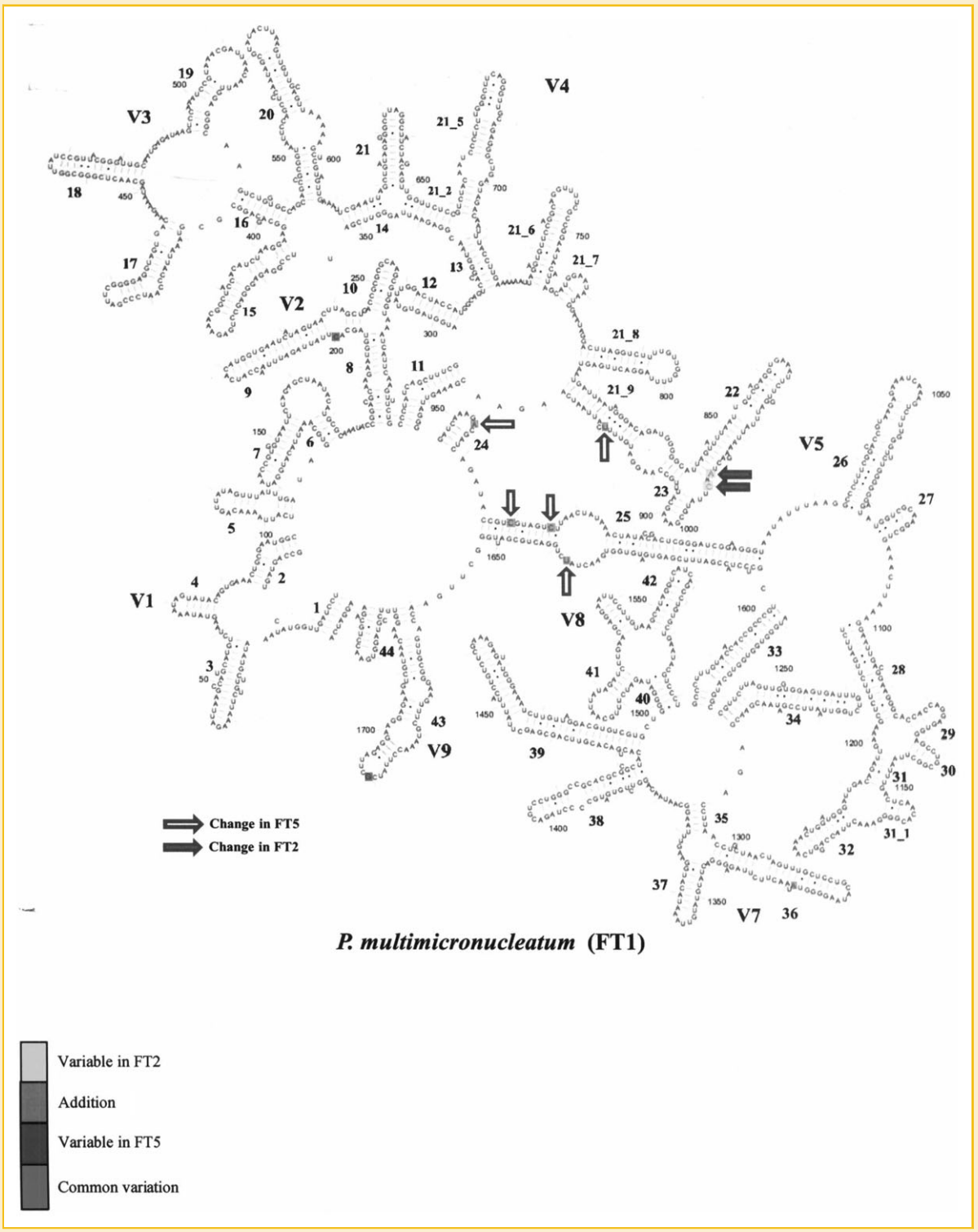

Fig. 2. Secondary structure of 18S rRNA of *P. multimicronucleatum* (FT1). This structure was used for comparison with that of FT2 and FT5 strains of *P. multimicronucleatum.* Structural differences between different isolates have been highlighted with different colors in the on-line-version of the article. In the print version this differentiation has been done by using different types of arrows (hollow and solid) and solid triangles.

630 . 640 in this helix makes it more stable than the model for FT4. Sub helix 21_5 showed substitutions at five different sites in primary sequence but without any change in structural configuration of both species. V5 (34 bp), V7 (72 bp) and V8 (151 bp) showed well conserved shapes of secondary structures in both species, although some substitutional changes appeared at different sites within these zones but without affecting the structural configuration of both models. The original configurations of secondary structures of both

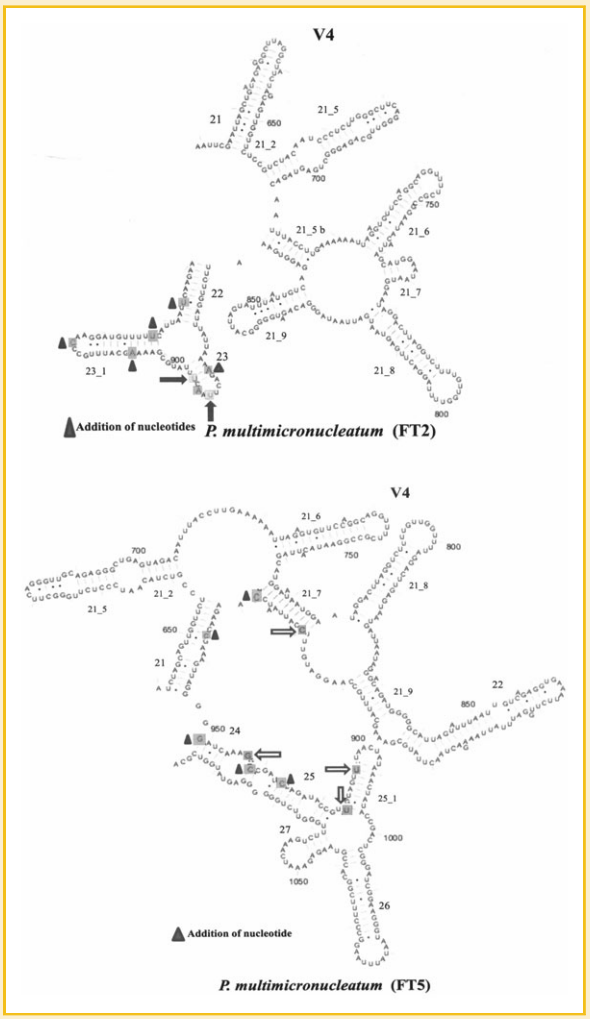

Fig. 3. Comparison of V4 region of secondary structures of 18SrRNA of FT2 and FT5 strains of *P. multimicronucleatum*. Structural differences between different isolates have been highlighted with different colors in the on-line-version of the article. In the print version this differentiation has been done by using different types of arrows (hollow and solid) and solid triangles.

models are retained because majority of the differences appeared within the bulges. Variable region V9 (104 bp) showed altered configuration of both models with four substitutional changes. Helix 43 of FT3 in this expansion zone emerged as an independent helix as opposed to FT4 where it is part of the core segment rather than bordering it as a separate helix.

COMPARISON OF *Paramecium caudatum pakistanicus* WITH ANOTHER CILIATE, FRUIT FLY AND HUMAN

The secondary structure of 18S rRNA of phylogenetically diverse taxa have been compared to appreciate the persistence of homologies and development of dissimilarities during the course of evolution. The sequence of 18S rRNA of a protozoan ciliate *Paramecium caudatum pakistanicus* was aligned with sequences of another ciliate *Tetrahymena thermophila*, a fruit fly *Drosophila melanogaster* and human *Homo sapiens* (Supplementary Fig. 2a–g). On analysis of core segments, 70 common variation sites were identified among the secondary structure models of all these species (Fig. 5; Supplementary Figures 3 and 4). Apart from the common variation sites, species specific mutational changes were also observed in core segments. In *T. thermophila,* when compared with *P. caudatum*, 38 mutational differences resulted in 11 G.C, 3 U.A, and 3 U.G base pairs, whereas the remaining variation sites were detected in hair pins or junctional loops forming no base pairs. *D. melanogaster* was found to be different at 33 sites including 9 A.U and 5 G.C. *H. sapiens* showed 32 differences with wholly or partially compensatory mutations including 10 G.C, 2 G.U and 2 A.U base pairs. In V1 expansion zone, helix 4 of *T. thermophila* folded into single stem by closing the terminal hair pin loop, similar to *P. caudatum*. However, one compensatory and one semi-compensatory mutations; U.U [60.81], A.G [61.80] of *P. caudatum* replaced by C.G [59.92] and C.G [60.91], respectively, in *T. thermophila* led to the formation of compact stem by removing the internal bulge of this helix in *P. caudatum*. The structural configuration of this region in *D. melanogaster* and *H. sapiens* appeared to be in conjunction with the 5th helix forming a long stem. However, a short stem of nine conserved residues appear to branch off at the junctional site of 4th and 5th helix in *H. sapiens*. V2 region varies in length from 106 in *P. caudatum*, to 124 in *T. thermophila*, 135 in *D. melanogaster* and to 173 in *H. sapiens*. This hyper variable region with 38 common variation sites among all four species folded into different structural shapes. An insertion of long nucleotides in this region increased the length of helix 8 in *H. sapiens* as opposed to its corresponding helix in other species. An interesting phenomenon of these insertions is their biased trend towards G.C base pairs which reduced the A.U percentage of this zone to 29% as opposed to 61% in *P. caudatum*, 63% in *T. thermophila* and 61% in *D. melanogaster* (Table IV). Despite mutational changes at 20 different sites, V3 region was identified as least variable in terms of its length and base pair composition. *P. caudatum* (65 bp) was recognized as 57%, *T. thermophila* (68 bp) as 61%, *D. melanogaster* (70 bp) as 60% and *H. sapiens* (68 bp) as 58% A.U rich. V4 zone emerged as second hyper variable region with 92 common mutational sites in all the four species. These wholly and partially compensatory mutations resulted in structural configuration of seven subhelices of helix 21. *D. melanogaster* was observed as longest in length (314 bp) with an insertion of 81 base pairs. However, despite the insertion of these nucleotides the A.U base compositions remain almost the same in *P. caudatum* (57%), *T. thermophila* (62%) and *D. melanogaster* (66%) with the exception of *H. sapiens* where it was reduced to 39%. Structural configuration of the V5 region folded into single stem with an internal and terminal hairpin loops in all three species except *H. sapiens* where the replacement of A.U with G.C pairs removed the internal bulge completely and terminal hairpin loop partially. V7 appeared as the second longest region of *D. melanogaster* (192 bp) with the insertion of 98 nucleotides. V8 expansion segment showed almost the same length size (140–144) in all the four species. However, the base substitutions reduced the A.U content to 44% in *H. sapiens* as opposed to 53–59% in other species which resulted in different structural models of V8 regions. Secondary structure of V9 length variable region predicts the same model for all the four species despite the presence of dozens of base substitutions. This structural conservation, despite substitutional differences, proves their functional role in structure formation as also proposed by Ouvrard et al. [2000].

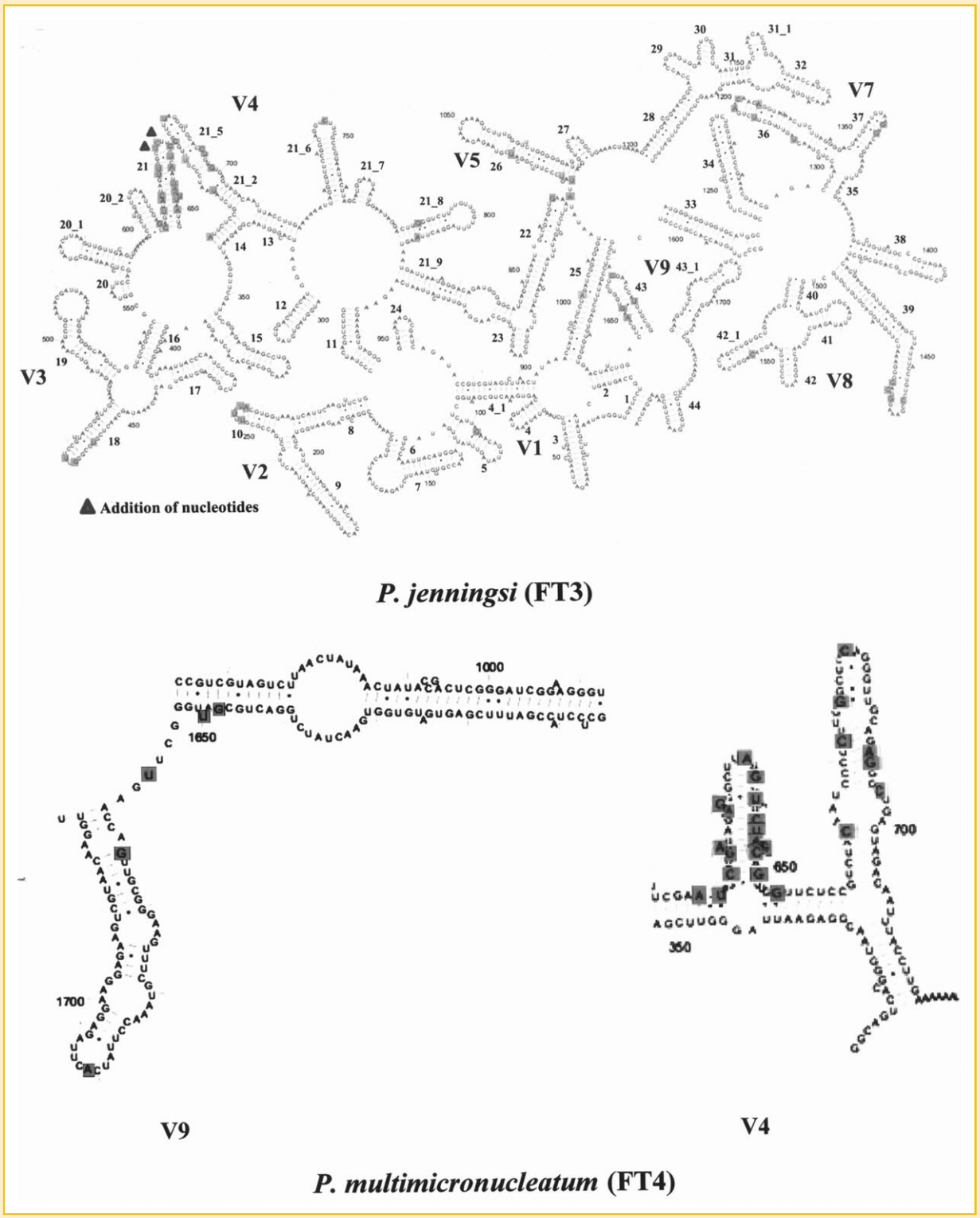

Fig. 4. Secondary structure of 18S rRNA of *Paramecium jennigsi* (FT3) and V4 and V9 regions of *P. multimicronucleatum* (FT4). The base pair substitutions of both the species have been highlighted with blue color in the on-line-version, and also with solid triangle in the printed version.

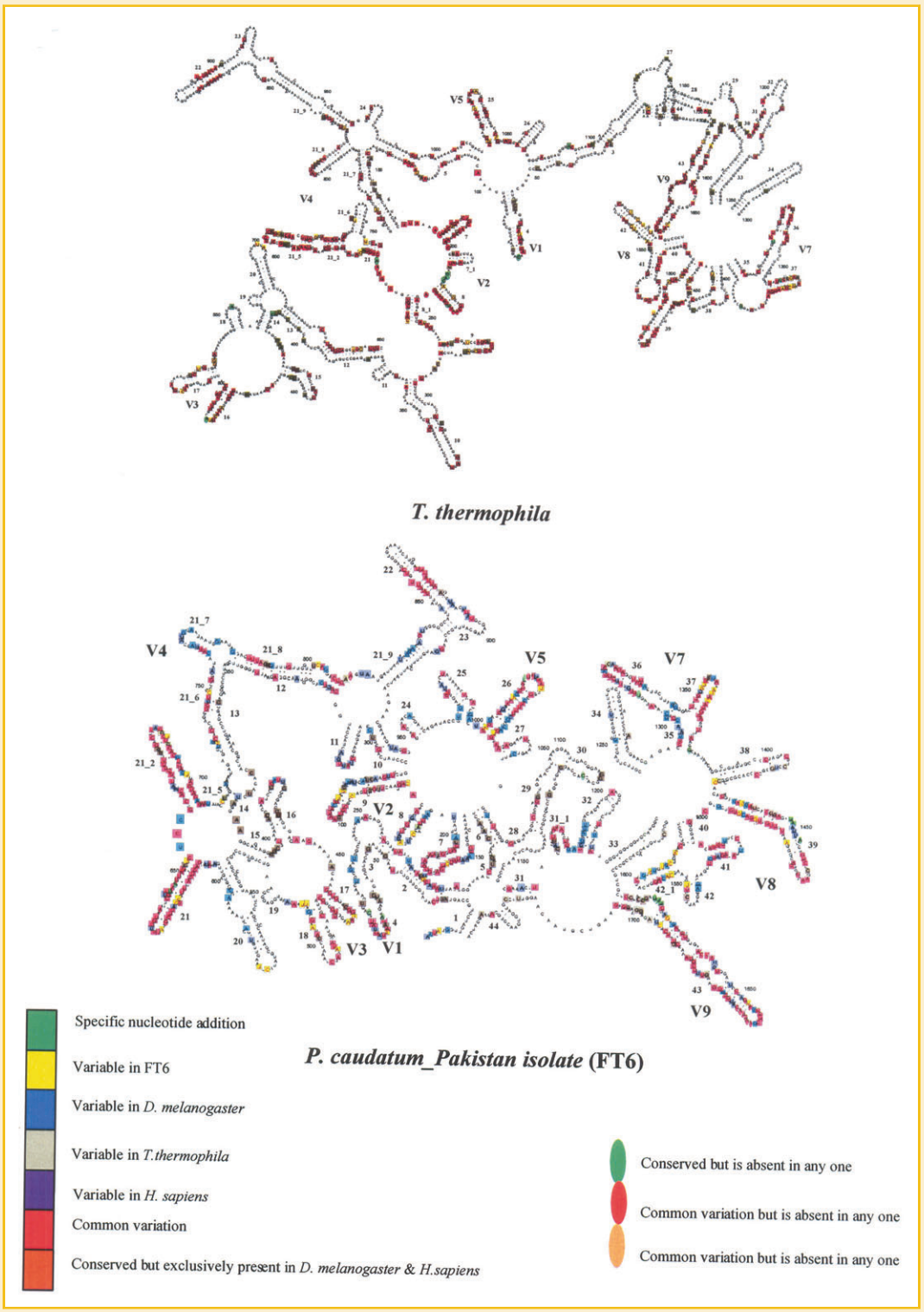

Fig. 5. Comparison of secondary structure of 18S rRNA of *Paramecium caudatum pakistanicus* (FT6) and *Tetrahymena thermophila*. This structure has also been used to compare with those of *Drosophila melanogaster* (Supplementary Fig. 3) and *Homo sapiens* (Supplementary Fig. 4).

TABLE IV. Differences in Lengths and Base Compositions of the Core and the Variable Regions of 18S rRNA of *Paramecium caudatum pakistanicus (FT6), Tetrahymena thermophila, Drosophila melanogaster* and *Homo sapiens*

Regions	*Paramecium caudatum pakistanicus*		*Tetrahymena thermophila*		*Drosophila melanogaster*		*Homo sapiens*	
Core segments	Length(bp) 978	GC (%) 45	Length (bp) 979	GC (%) 44	Length (bp) 980	GC (%) 46	Length (bp) 985	GC (%) 51
Variable regions								
V1	19	21	19	42	21	23	22	59
V2	106	39	124	37	135	39	173	71
V3	65	43	68	39	70	40	68	42
V4	222	43	221	38	314	34	243	61
V5	34	35	29	48	35	54	35	68
V7	85	34	86	38	192	35	92	59
V8	144	47	141	41	140	43	142	56
V9	96	47	86	46	108	43	110	62

CONCLUSIONS

Three different *Paramecium* species viz., *P. multimicronucleatum, P. jenningsi* and *P. caudatum pakistanicus* were identified from the local environment, based on 18S rRNA analysis. The isolate identified as *P. caudatum pakistanicus* here, however showed weak similarity with the 18S rRNA sequence of *P. caudatum* from the data bank. In fact the clade containing this species was at quite a distance from other species in the phylogenetic tree, suggesting their novelty from this part of the world. This isolate has been designated as *Paramecium caudatum pakistanicus*. This article also considers secondary structure of 18S rRNA as an additional and perhaps more reliable topological marker for species discrimination and for determining possible phylogenetic relationship between the ciliate species. On the basis of secondary structural comparison among *P. caudatum*, *T. thermophila*, *D. melanogaster*, and *H. sapiens* it can be deduced that variable regions are more helpful in differentiating the species at interspecific level rather than at intraspecific level. V3 region was observed as the least variable throughout the kingdom. V2 and V7 are the longest expansion segments of *D. melanogaster*. Nearly all of the A.U richness of this species is contributed by the expansion segments of V2, V4, and V7. It is also interesting to note that there is continuous mutational bias towards G.C base pair composition in *H. sapiens* making it more favorable and energetically stable by releasing it for the need of RNA–RNA or RNA–protein interaction for stabilization.

ACKNOWLEDGMENTS

The financial support from University of the Punjab, Lahore, Pakistan and Research Centre, College of Science, King Saud University, Riyadh, Kingdom of Saudi Arabia is gratefully acknowledged.

REFERENCES

Breiner HW, Foissner W, Stoeck T. 2008. The search finds an end: colpodidiids belong to the Class Nassophorea (Ciliophora). J Eukary Microbiol 55:100–102.

Chen W, Zhang Y. 2012. Comparative analysis of RNA molecules. MATCH Commun Math Comput Chem 67:253–268.

Curds CR. 1982. The ecology and role of protozoa in aerobic sewage treatment processes. Annu Rev Microbiol 36:27–46.

Curds CR, Gates MA, Roberts D-MCL. 1983. British and other freshwater ciliated protozoa, Part II: Ciliophora, Oligohymenophora and Polyhymenophora. Keys and notes for the identification of the free living genera. In: Karmack DM Barnes RSK eds. Synopsis of the British fauna No. 23, London: Cambridge University Press. pp 1–474.

Fokin SI. 2011. *Paramecium* genus: biodiversity, some morphological features and the key to the main morphospecies discrimination. Protistology 6:227–235.

Fokin SI, Przyboś E, Chivilev SM, Beier CL, Horn M, Skotarczak D, Wodecka B, Fujishima M. 2004. Morphological and molecular investigations of *Paramecium schewiakoffi* sp. nov. (Ciliophora, Oligohymenophorea) and current status of paramecia distribution and taxonomy. Eur J Protistol 40:225–243.

Gagnon S, Bourbeau D, Levesque RC. 1996. Secondary structure and features of the 18S, 5.8S and 26S ribosomal RNAs from the Apicomplexan parasite *Toxoplasma gondii*. Gene 173:129–135.

Gao S, Huang J, Li J, Song W. 2012. Molecular phylogeny of the cyrtophorid ciliates (Protozoa, Ciliophora, Phyllopharyngea). PLoS ONE 7:e33198.

Gao F, Katz LA, Song W. 2013. Multigene-based analyses on evolutionary phylogeny of two controversial ciliate orders: Pleuronematida and Loxocephalida (Protista, Ciliophora, Oligohymenophorea). Mol Phylogen Evol 68:55–63.

Hofacker IL, Fekete M, Stadler PF. 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol 319:1059–1066.

Hoshina R, Hayashi S, Imamura N. 2006. Secondary structure of two regions in expansion segments ES3 and ES6 with the potential of forming a tertiary interaction in eukaryotic 40S ribosomal subunits. Acta Protozool 45:377–386.

Hwang UW, Ree HI, Kim W. 2000. Evolution of hypervariable regions, V4 and V7, of insect 18S rRNA and their phylogenetic implications. Zool Sci 17:111–121.

Lee JC, Gutell RR. 2012. A comparison of the crystal structures of eukaryotic and bacterial SSU ribosomal RNAs reveals common structural features in the hypervariable regions. PLoS One 7:38203.

Letsch HO, Kuck P, Stocsits RR, Misof B. 2010. The impact of rRNA secondary structure consideration in alignment and tree reconstruction: Simulated data and a case study on the phylogeny of hexapods. Mol Biol Evol 27:2507–2521.

Li L, Shao C, Song W, Lynn DH, Chen Z, Shin MK. 2009. Does *Kiitricha* (Protista, Ciliophora, Spirotrichea) belong to *Euplotida* or represent a primordial spirotrichous taxon? With suggestion to establish a new subclass Protohypotrichia. Int J Syst Evol Microbiol 59:439–446.

Lukashenko NP. 2009. Molecular evolution of ciliates (ciliophora) and some related groups of protozoans. Russian J Genet 45:1013–1028.

Lynn DH, Struder-Kypke M. 2002. Phylogenetic position of Licnophora, Lechriopyla, and Schizocaryum, three unusual ciliates (*Phylum Ciliophora*) endosymbiotic in echinoderms (*Phylum Echinodermata*). J Eukary Microbiol 49:460–468.

Marande W, López-García P, Moreira D. 2009. Eukaryotic diversity and phylogeny using small and large-subunit ribosomal RNA genes from environmental samples. Environ Microbiol 11:3179–3188.

Milyutina IA, Aleshin VV, Mikrjukov KA, Kedrova OS, Petrov NB. 2001. The unusually long small subunit ribosomal RNA gene found in a mitochondriate amoeboflagellate, *Pelomyxa palustus:* its rRNA predicted secondary structure and phylogenetic implications. Gene 272:131–139.

Nelles L, Fang BL, Volckaert G, Vandenberghe A, De Wachter R. 1984. Nucleotide sequence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. Nucl Acids Res 12:8749–8768.

Ouvrard D, Campbell BC, Bourgoin T, Chan KL. 2000. 18S rRNA secondary structure and phylogenetic position of Peloridiidae (Insecta, Hemiptera). Mol Phylogen Evol 16:403–417.

Przybos E, Fokin S. 2000. Data on the occurrence of species of the *Paramecium aurelia* complex world-wide. Protistology 1:179–184.

Przybos E, Greczek-Stachura M, Potekhin A, Rautian M. 2007. Strains of *Paramecium* decaurelia (Ciliophora, Protozoa) from Russia with molecular characteristics of other known strains of the species. Folia Biol (Krakow) 55:3–4.

Przybos E, Greczek-Stachura M, Prajer M, Potekhin A, Cotsinia A. 2008. Two species of the *Paramecium aurelia* complex (Ciliophora, Protista) from the Black Sea region (Russia) with their RAPD-PCR fingerprints characteristics. Protistology 5:207–212.

Przybos E, Hori M, Fokin SI. 2003. Strains of *Paramecium quadecurelia* from Namibia, Africa; genetic molecular studies. Acta Protozool 42:357–360.

Przybos E, Nevo E, Pavlicek T. 2002. Distribution of species of the *Paramecium aurelia* complex in Israel. Acta protozool 41:293–295.

Przybos E, Prajer M, Greczek-Stachura M. 2005. Molecular analysis (RAPD-PCR) of inter-strain hybrids of the *Paramecium aurelia* species complex (Ciliophora, Protozoa). Folia Biol 53:3–4.

Przybos E, Tarcz S, Fokin S. 2009. Molecular polymorphism of *Paramecium tetraurelia* (Ciliophora, Protozoa) in strains originating from different continents. Folia Biol (Kraków) 57:1–2.

Przybos E, Tarcz S, Greczek-Stachura M, Surmacz M. 2011. Polymorphism of Paramecium pentaurelia (Ciliophora, Oligohymenophorea) strains revealed by rDNA and mtDNA sequences. Eur Protistol 47:138–143.

Przybos E, Tarcz S, Potekhin A, Rautianc M, Prajer M. 2012. A two-locus molecular characterization of *Paramecium calkinsi*. Protist 163:263–273.

Przybos E, Tarcz S, Schmidt H, Czubatinski L. 2008. First Stand of *Paramecium octaurelia* in Europe and molecular characteristics of other known strains of this species. Folia Biol (Krakow) 57:1–2.

Regensbogenova M, Kisidayova S, Michalowsk T, Javorsky P, Moon-van der Staay SY, Moon-van der Stay GW, Hackstein JHP, McEwan NR, Jouany JP, Newbold JC, Pristas P. 2004. Rapid identification of rumen protozoa by restriction analysis of amplified 18S rRNA gene. Acta Protozool 43:219–224.

Sambrook J, Russell DW. 1989. Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Shakoori AR, Rehman A, Haq R. 2004. Multiple metal resistance in the ciliate protozoan, *Vorticella microstoma*, isolated from industrial effluents and its potential in bioremediation of toxic wastes. Bull environ Contam Toxicol 72:1046–1051.

Shang H, Song W, Warren A. 2003. Phylogenetic positions of two ciliates, *Paranophrys magna* and *Mesanophrys carcini* (Ciliophora: Oligohymenophorea), within the subclass Scuticociliatia inferred from complete small subunit rRNA gene sequences. Acta Protozool 42:171–181.

Srivastava A, Cai L, Mrazek J, Malmberg RL. 2011. Mutational patterns in RNA secondary structure evolution examined in three RNA families. PLoS One6: e 20484.

Strüder-kypke MC, Wrigh ADG, Fokin SI, Lynn DH. 2000. Phylogenetic relationships of the genus *Paramecium* inferred from small subunit rRNA gene sequences. Mol Phylogenet Evolut 14:122–130.

Strüder-Kypke MC, Wright ADG, Jerome CA, Lynn DH. 2001. Parallel evolution of histophagy in ciliates of the genus *Tetrahymena*. BMC Evolut Biol 1:5.

Ueno R, Huss VAR, Urano N, Watabe S. 2007. Direct evidence for redundant segmental replacement between multiple 18S rRNA genes in a single *Prototheca* strain. Microbiology 153:3879–3893.

Varriale A, Torelli G, Bernardi G. 2008. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. RNA 14:1492–1500.

Voigt O, Erpenbeck D, Wörheide G. 2008. Molecular evolution of rDNA in early diverging Metazoa: First comparative analysis and phylogenetic application of complete SSUrRNA secondary structures in Porifera. BMC Evolut Biol 8:69.

Weigand AM, Dinapoli A, Klussmann-Kolb A. 2012. 18S rRNA variability map for gastropoda. J Mollus Stud 78:151–156.

Woodruff LL. 1921. The structure, life history, and intrageneric relationships of *Paramecium calkinsi*. sp. nov. Biol Bull 41:171–180.

Xia X, Xie Z, Kjer KM. 2003. 18S ribosomal RNA and tetrapod phylogeny. Syst Biol 52:283–295.

Xia Q, Lin J, Quin Y, Zhou J, Bu W. 2011. Structural diversity of eukaryotic 18S rRNA and its impact on alignment and phylogenetic reconstruction. Protein Cell 2:161–170.

Yi Z, Song W, Gong J, Warren A, Al-Rasheid KAS, Al-Arifi S, Al-Khedhairy AA. 2009. Phylogeny of six oligohymenophoreans (Protozoa, Ciliophora) inferred from small subunit rRNA gene sequences. Zool Scrip 38:323–331.

Yu S, Wang Y, Rédei D, Xie Q, Bu W. 2013. Secondary structure models of 18S and 28S rRNAs of the true bugs based on complete rDNA sequences of *Eurydema maracandica* Oshanin, 1871 (Heteroptera, Pentatomidae). Zoo Keys 319:363–377.

Zimmermann J, Jahn R, Gemeinholzer B. 2011. Barcoding diatoms: Evaluation of the V4 subregion on the 18SrNA, including new primers and protocols. Organi Divers Evolut 11:173–192.

Zhang Q, Simpson A, Song W. 2012. Insights into the phylogeny of systematically controversial haptorian ciliates (Ciliophora, Litostomatea) based on multigene analyses. Proc Biol Sci 279:2625–2635.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.