Protein Cell 2011, 2(2): 161–170
DOI 10.1007/s13238-011-1017-2

Protein & Cell

RESEARCH ARTICLE

Structural diversity of eukaryotic 18S rRNA and its impact on alignment and phylogenetic reconstruction

Qiang Xie[1] ✉, Jinzhong Lin[2], Yan Qin[2], Jianfu Zhou[3], Wenjun Bu[1] ✉

[1] Department of Zoology and Developmental Biology, College of Life Sciences, Nankai University, Tianjin 300071, China
[2] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3] College of Information Technical Science, Nankai University, 94 Weijin Road, Tianjin 300071, China
✉ Correspondence: qiangxie@nankai.edu.cn (Q. Xie), wenjunbu@nankai.edu.cn (W. Bu)

Received January 21, 2011 Accepted January 30, 2011

ABSTRACT

Ribosomal RNAs are important because they catalyze the synthesis of peptides and proteins. Comparative studies of the secondary structure of 18S rRNA have revealed the basic locations of its many length-conserved and length-variable regions. In recent years, many more sequences of 18S rDNA with unusual lengths have been documented in GenBank. These data make it possible to recognize the diversity of the secondary and tertiary structures of 18S rRNAs and to identify the length-conserved parts of 18S rDNAs. The longest 18S rDNA sequences of almost every known eukaryotic phylum were included in this study. We illustrated the bioinformatics-based structure to show that, the regions that are more length-variable, regions that are less length-variable, the splicing sites for introns, and the sites of A-minor interactions are mostly distributed in different parts of the 18S rRNA. Additionally, this study revealed that some length-variable regions or insertion positions could be quite close to the functional part of the 18S rRNA of Foraminifera organisms. The tertiary structure as well as the secondary structure of 18S rRNA can be more diverse than what was previously supposed. Besides revealing how this interesting gene evolves, it can help to remove ambiguity from the alignment of eukaryotic 18S rDNAs and to improve the performance of 18S rDNA in phylogenetic reconstruction. Six nucleotides shared by Archaea and Eukaryota but rarely by Bacteria are also reported here for the first time, which might further support the supposed origin of eukaryote from archaeans.

KEYWORDS secondary structure diversity, tertiary structure diversity, 18S rRNA, Foraminifera, Euglenida

INTRODUCTION

The ribosomal RNAs (rRNAs) are among the most widely known functional macromolecules. Because rRNAs play the key functional role in the ribosome (Noller, 1991; Green and Noller, 1997; Nissen et al., 2000; Ramakrishnan, 2002), resolving the structures of rRNAs is critical for understanding the details of the function of rRNAs and ribosomes. The molecular structures of a few 16S and 23S rRNAs have been resolved in the last decade (Ban et al., 2000; Schluenzen et al., 2000; Wimberly et al., 2000; Harms et al., 2001; Spahn et al., 2001; Yusupov et al., 2001). The 18S rRNAs of eukaryotes are more variable in length than the homologous 16S rRNAs averaging 1.5 kb, but ranging from about 1.5 kb to over 4.5 kb. Previous comparative studies of 18S rRNAs have suggested that the major length-variable regions are distributed on the surface of the molecules (Spahn et al., 2001; Wuyts et al., 2001; Chandramouli et al., 2008), while the splicing sites for introns are clustered in the inner region (Jackson et al., 2002; Chandramouli et al., 2008). Nearly all of the variability is contributed by 3 major variable regions, namely, V2, V4, and V7, which were named by Neefs et al. (1991). All of these studies suggest that the 18S rRNA be highly conservative.

An increasing number of 18S rDNA sequences with unusual lengths have been documented in GenBank recently (Crease and Colbourne, 1998; Cunningham et al., 2000; Giribet and Wheeler, 2001; Busse and Preisfeld, 2003;

Pawlowski et al., 2003). The 18S rDNA is probably the most frequently sequenced gene in eukaryotes and the complete or nearly complete sequences of over 6500 rDNAs had been available seven years ago (Wuyts et al., 2004). There are totally 375,786 issues containing 18S with various lengths in GenBank now. When especially long 18S rDNA sequences were examined further in this study, more length-variable positions and higher length variability were discovered within the 18S rDNA. In the rDNAs of the forams (Foraminifera) in particular, several unique length-variable regions or inserted sequences exist, which are not present in the 18S rDNAs of other eukaryotes. Some of these unique insertions are located unexpectedly close to the functional part of the 18S rRNA. These results suggest that the tertiary structure as well as the secondary structure of 18S rRNA is more diverse than what was supposed earlier.

The highly variable length of 18S rDNAs raises a great impediment to the aligning of the length-variable regions across taxa. Some methods have been developed to align RNA coding genes based on the secondary structure, and/or describe the substitution models of the stem and loop regions these years (Schöniger and von Haeseler, 1994; Jow et al., 2002; Hudelot et al., 2003; Siebert and Backofen, 2005; Telford et al., 2005; Seibel et al., 2006; Wolf et al., 2008; Schultz and Wolf, 2009; Stocsits et al., 2009; Keller et al., 2010). The use of these secondary-structure-based methods of alignment is still restricted by the length, the completeness, the level of length variation and the number of sequences. So these methods are mainly used in the analyses of tRNA, internal transcribed spacers (ITS) and some parts of nuclear rDNA with moderate level of length variation. Due to the complexity of length variation, the parts with hyper-extensive length are better manually removed from the original sequences before inferring phylogeny (Xie et al., 2008, 2009).

Due to the differences in taxa sampled and ambiguous alignment, the findings of previous phylogenetic studies based on 18S rDNAs vary extensively from each other (Cavalier-Smith and Chao, 1996; Kumar and Rzhetsky, 1996; Van de Peer and De Wachter, 1997; Burki et al., 2002; Kostka et al., 2004; Nikolaev et al., 2004; Polet et al., 2004; Shalchian-Tabrizi et al., 2006; Shalchian-Tabrizi et al., 2007). Additionally because of the effects of sparse taxon sampling and heterogeneity in the sequence data, the results of the studies based on multiple proteins or protein-coding genes are also quite different from each other (Baldauf et al., 2000; Philippe et al., 2004; Harper et al., 2005; Rodríguez-Ezpeleta et al., 2005; Burki and Pawlowski, 2006; Burki et al., 2007; Hackett et al., 2007; Patron et al., 2007; Kim and Graham, 2008). Only Opisthokonta, also known as the Fungi-Metazoa group, appears in all of the studies of eukaryote interrelationships (Parfrey et al., 2006). Most of these studies also support a clade composed of Stramenopila and Alveolata (Keeling et al., 2005). Because the taxon sampling in new phylogenomic studies insists to be based on the existing 18S-derived phylogenetic scheme with the most comprehensively sampled taxa, the studies of 18S rDNAs are still necessary and a standardized approach for using 18S in phylogenetic studies is desirable.

In this study, 138 eukaryote taxa were sampled so that almost all of the known phyla were included for our phylogenetic reconstruction of eukaryotes. The secondary-structure model of 18S rRNA makes it possible to extract the maximum length of common parts from the original sequences and to reconstruct phylogeny based on those parts. This approach can minimize the interference of ambiguous alignment and keep the number of informative sites as high as possible. It will let future phylogenetic studies based on 18S rDNAs be more comparable.

RESULTS AND DISCUSSION

The diversity of the secondary and tertiary structures of 18S rRNAs

As seen in illustrations of the secondary and tertiary structures (Fig. 1 and 2), the positions of the regions that are more length-variable, regions that are less length-variable, the splicing sites for introns, and the sites of A-minor interactions are distributed in different parts of 18S rRNA. Twenty-six positions of introns are summarized in the illustrations of the secondary structure of 18S rRNA (a–z in Fig. 1) and the tertiary structure of 16S rRNA (pink symbols in Fig. 2). The introns are distributed in the inner region, mostly around the translation functional region (Fig. 2). The position and variability of each length-variable region are also summarized (Fig. 1 and 2). There are 24 common length-variable regions (A–X in Fig. 1) and 15 specific regions (dark blue and light blue symbols in Fig. 1 and 2) that can be found in forams and a few other eukaryotes. Fourteen of the 24 length-variable regions are clustered in the V2, V4, and V7 regions, and these large regions contribute to more than 70% of the total length-variability (approximately 4.7 kb out of 6.5 kb). Of the remaining length-variable regions, 10 are medium-sized with approximately 100–300 bases (the 4 regions in V8 are viewed as 1 region), while the other 12 regions are small with approximately 40 bases or less (the 2 regions in V6 are viewed as 1 region). The details of the species and length information of their introns and length-variable regions are provided in the supplementary files (Fig. S1 and S2). The sequences/taxa with the highest recorded length for each length-variable region are listed in Table 1.

Some novel results suggest that 18S rRNA can be more diverse than what has been believed earlier. In the secondary structure, some specific length-variable regions in the 18S rDNAs of forams are present quite close to the positions of some introns (light blue symbols in Fig. 1 and 2). In the tertiary structure, some insertions that are less length-variable occur quite close to the functional part (light blue dots in Fig. 2).

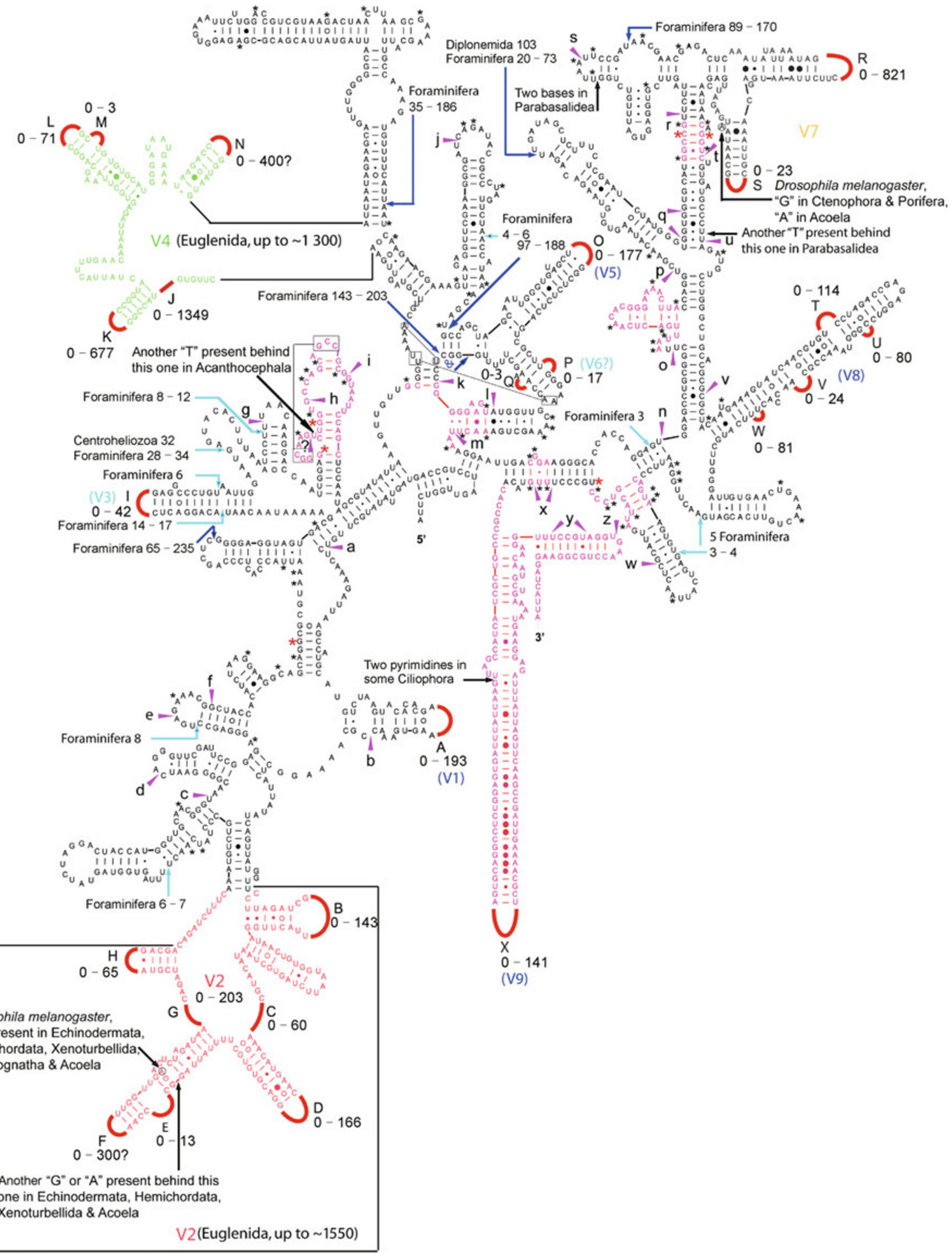


Figure 1. Positions of the introns (a–z) and length-variable regions (A–X) of eukaryotic 18S rRNA in the secondary structure. The sequence is that of the 18S rDNA of *Drosophila melanogaster* (GenBank Accession No. M21017). The positions q, x, and y indicate 3 ranges of possible intron positions. The red curved lines labeled as A to X represent the 24 length-variable regions. Black, light blue, and dark blue arrows indicate positions where there would be specific insertions of 1–2, approximately 30 or less, and approximately 100–300 bases, respectively. Base pairing is indicated as follows: standard canonical pairs by lines (G-C and A-U), wobble G:U pairs by dots (G·U), A:G or A:C pairs by open circles (A○G and A○C), and other non-canonical pairs by filled circles (e.g., A●A). The 81 black asterisks (*) indicate the bases shared by all of Eukaryota, Archaea and Bacteria. The 6 red asterisks (*) indicate the bases solely shared by Eukaryota and Archaea. The bases labeled in purple indicate the translation-functional regions.

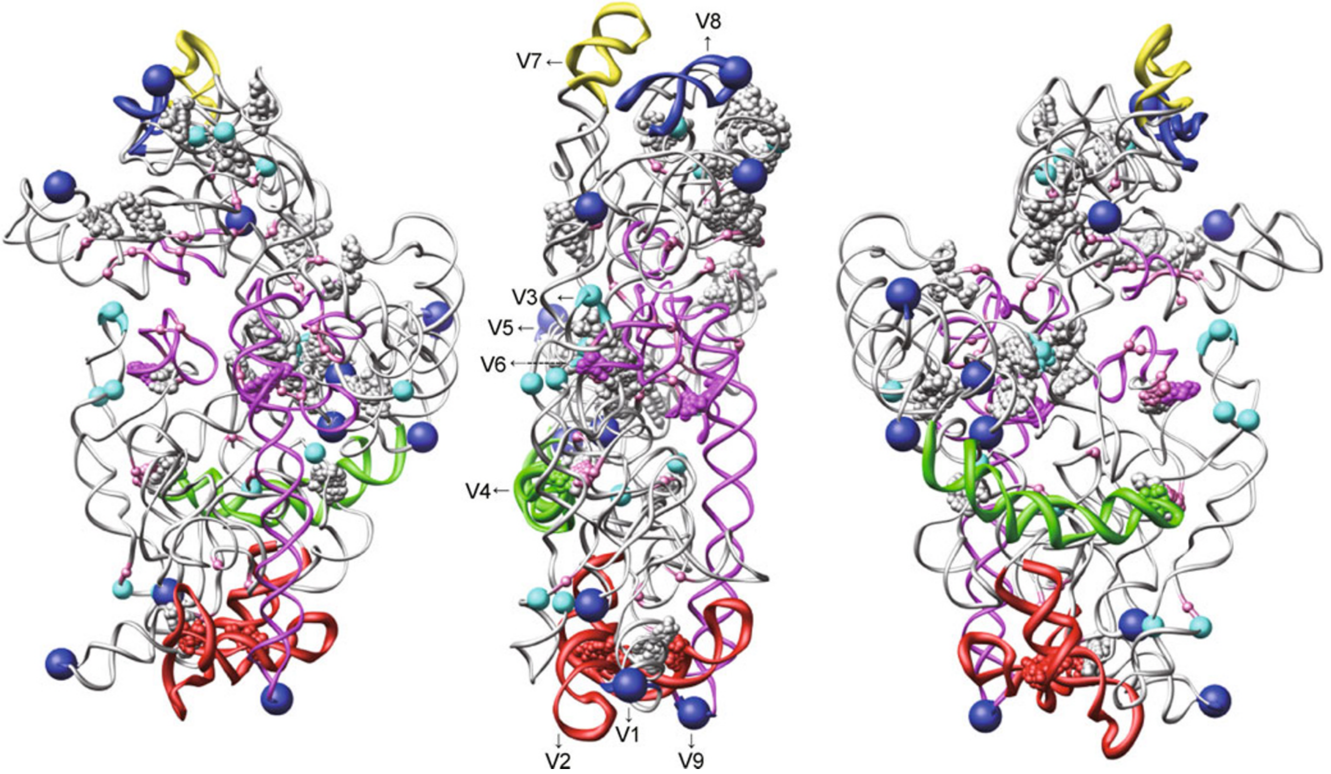


Figure 2. Positions of the introns and length-variable regions of eukaryotic 18S rRNA marked onto the tertiary structure of the 16S rRNA of *Thermus thermophilus* in intersubunit surface view (A), side view (B) and cytoplasm surface view (C). The functional translation domains (purple) and the variable regions V1 (dark blue), V2 (red), V3 (light blue), V4 (green), V5 (dark blue), V6 (light blue), V7 (yellow), V8 (dark blue), and V9 (dark blue) are illustrated by colored lines. The remaining grey lines are the other length-conserved regions. The positions of introns are illustrated by pink pellets and lines. The longer specific insertions (> 35 nts) are illustrated by dark-blue including blue larger pellets, while the smaller ones (< 35 nts) are illustrated by light-blue including smaller pellets. The positions of A-minor interactions (24) are also included and illustrated by the smallest pellets that are the same colors as that of the regions wherein they are located.

The shortest length of eukaryotic 18S rDNA can be ~1.5 kb (Diplomonadida: *Spironucleus salmonicida*, GenBank Acc. No. DQ812526; Microsporidia: *Spraguea lophii*, GenBank Acc. No. AF033197), which is approximately the length of the prokaryotic 16S rDNA. The longest length of 18S rDNA without introns can be more than 4.5 kb (Euglenida: *Distigma sennii*, GenBank Acc. No. AF386644 and AY062001), which is longer than most 28S rDNAs. The longest variable region is region J in V4 (Fig. 1), which is 1349 bases in *Cubaris murina* (Crustacea: Isopoda). Although the most extensive regions are distributed on the external cytoplasmic side of 18S rRNA which may be little functional and the press of selection is quite low, the massive local length-variability of 1349 bases is quite unusual. Additionally, the positions of length-variable regions or insertions can be divided into 3 levels according to the extents of their variability and their distances from the deep-lying translation-functional region. The regions V2, V4, and V7 are the most variable in length and the farthest from the functional region. The smaller V1, V5, V8 and V9 regions are less variable in size and closer to the functional region (Fig. 2). The V3 and V6 regions are the least variable in length and can be quite close to the deep intron positions or even to the functional region.

Few 18S rDNAs with introns have a mature rRNA segment longer than 2 kb. The longest 18S rDNA, introns included, is nearly 6.4 kb (Amoebozoa: *Diderma niveum*, GenBank Acc. No. AM231291), and its mature rRNA is approximately 1.9 kb. The longest mature rRNA segment of an 18S rDNA containing introns is 2.3 kb long (Amoebozoa: *Acanthamoeba* sp., GenBank Acc. No. AY176047, ~3.6 kb). Note that this is considerably shorter than the longest intron-free 18S rDNA, the 4.5 kb gene of the euglenid mentioned above. It is probably that some unknown mechanism may inhibit the coexistence of multiple introns and lengthened variable regions in the same rDNA. Conversely, the longest mature rRNA segment of 18S containing introns (2.3 kb) is still considerably longer than most common intron-free 18S rDNAs, which are 1.8–1.9 kb in length. This suggests that the structure of the 18S rRNA can still lengthen its variable regions to some extent, even in the presence of introns.

Table 1 The 18S rDNA sequences with the highest local extension

LVR[a]	GenBank Acc. No.	Sequence length	Species	Taxon-specific extension
A	DQ408641	3579	*Uvigerina phlegeri*	Order Foraminifera
B	AF106036	3725	*Distigma proteus*	Genus *Distigma*
C	AY305011	2818	*Actinosphaerium eichhornii*	⩽ Class Actinophryidae[b]
D	AJ010592	2040	*Guillardia theta*	⩽ Genus *Guillardia*[b]
E	–	–	–	–
F	AF386644	4503	*Distigma sennii*	Genus *Distigma*
G	X77784	3316	*Xenos vesparum*	Order Strepsiptera
H	AY305011	2818	*Actinosphaerium eichhornii*	⩽ Class Actinophryidae[b]
I	AJ318228	3553	*Eggerelloides scabrum*	Order Foraminifera
J	AJ287064	3537	*Cubaris murina*	Crinocheta
K	AY596366	2579	*Hainanjapyx jianfengensis*	Class Diplura
L	AY268037	2746	*Multicilia marina*	Unknown
M	–	–	–	–
N	L23799 AF386644	2741 4503	*Phreatamoeba balamuthi* *Distigma sennii*	Family Mastigamoebidae Genus *Distigma*
O	AF106036	3725	*Distigma proteus*	Genus *Distigma*
P	DQ122380	2492	*Sappinia diploidea*	Genus *Sappinia*
Q	–	–	–	–
R	AJ243681	2873	*Gigantolina magna*	⩾ Genus *Gigantolina*[c]
S	AY769863	1960	*Entamoeba invadens*	⩽ Entamoebidae[b]
T	AJ318224	4066	*Astrorhiza triangularis*	Order Foraminifera
U	AF386644	4503	*Distigma sennii*	Genus *Distigma*
V	AJ318224 AJ318223	4066 4140	*Astrorhiza triangularis* *Astrammina rara*	Order Foraminifera
W	DQ408644	3507	*Discorbis rosea*	Order Foraminifera
X	AY305011	2818	*Actinosphaerium eichhornii*	⩽ Class Actinophryidae[b]

[a] LVR, length-variable region; [b] ⩽ stands for at most; [c] ⩾ stands for at least.

These results demonstrate that the length-variable regions can be present fairly close to the translation functional region of the 18S rRNA. For example, the sites around V3 and V6 regions in Fig. 1. On one aspect, these variations still follow the rule that, the closer to the functional part, the shorter of the length-variation. On the other aspect, the secondary and tertiary structures of the 18S rRNA are more diverse than was believed previously. Considering the correlation between structure and function, there should be some differences, more or less, between the function (maybe translational efficiency) of ribosomes of different species.

These results also provide modular recognition of 18S rRNA. When combined with the secondary structure, the sites of length variation or introns can serve as markers to improve the determination of large-scale positional homology inside 18S rDNA and thereby can be helpful for identifying the real length-conserved parts.

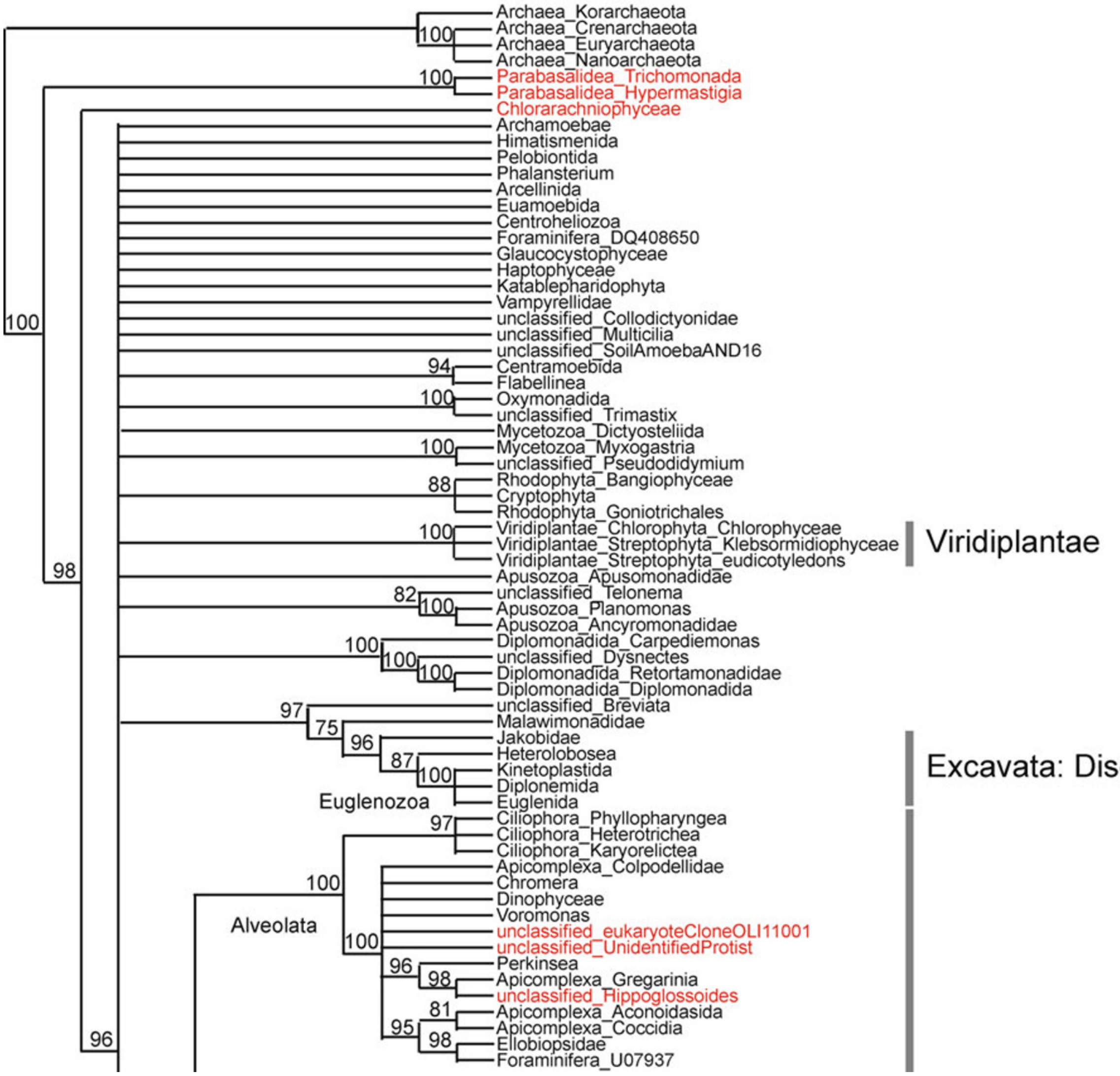
Experimentally deleting the length-variable regions constitutes is the standard approach for evaluating functions of these regions (Sweeney et al., 1994). Because a large length-variable region may actually comprise several separate length-stable regions as well as smaller length-variable regions, our present results can help in determining precisely which parts to delete in such deletion tests.

A potentially important discovery is that, with the help of the secondary-structure model, six nucleotides shared by Archaea and Eukaryota but not congruently by Bacteria are found in the length-conserved parts of 18S rDNAs (the six scattered red asterisks in Fig. 1, and Fig. S3 and S4). Starting at the 5′ end after V2, the first three nucleotides are all Gs, while in bacteria, they are all Ys (C or T). The fourth shared nucleotide, a C, is T in all of the ~100 bacteria sampled except for NC_004307 (Actinobacteria), NC_000918 (Aquificae) and NC_006461 (Deinococcus-Thermus). The fifth shared nucleotide, the G paired with the C, is A in all Bacteria, except in the same three sequences mentioned. The final nucleotide shared by Archaea and Eukaryota, a T (U in Fig. 1), is C in all bacteria except for NC005027 (Planctomycetes). These six shared nucleotides can be viewed as the relics of the origin of eukaryotes from archaeans and support the endosymbiotic theory, which says eukaryotes originate from the endosymbiosis of bacteria into archaeans and result in bacteria and archaean nucleoid evolving into organelles and nucleus of eukaryotic cells respectively (Margulis, 1970). No bases shared solely by Eukaryota and Bacteria could be found in the dataset.

Although this evidence only supports eukaryotic rRNA originating from archaean or "pre-archaean" rRNA, to the exclusion of the bacterial rRNA, but not eukaryotic nucleus from archaean nucleoid, it is still reasonable to suppose so. One reason is that the horizontal gene transfers from organelles into nucleus can make eukaryotic genome seems like chimeric. Another reason is that probably few genes like rDNAs can reserve relics for events which happened about 2 billion years ago. All of these can greatly reduce the numbers of possible sources of evidences.

The phylogeny of Eukaryota based on the length stable parts of 18S rDNAs

The phylogenetic study based on the length-conserved parts of 18S rDNAs yielded the following results (Fig. 3). First, the monophyly of the clade (Rhizaria + Stramenopila) is witnessed, and the corresponding posterior probability is 100%. Compared with the corresponding value (71%) in a previous study (Shalchian-Tabrizi et al., 2006), this value is more convincing. And this clade was also supported by the evidence of expressed sequence taqs (ESTs) (Burki et al., 2007). In the other phylogenetic studies based on 18S rDNA, Stramenopila had been shown to evolve prior to the divergence of animals and fungi and green plants and red algae (Cavalier-Smith and Chao, 1996; Kumar and Rzhetsky, 1996), or have close relationship to Alveolata (Van de Peer and De Wachter, 1997; Nikolaev et al., 2004), or be position



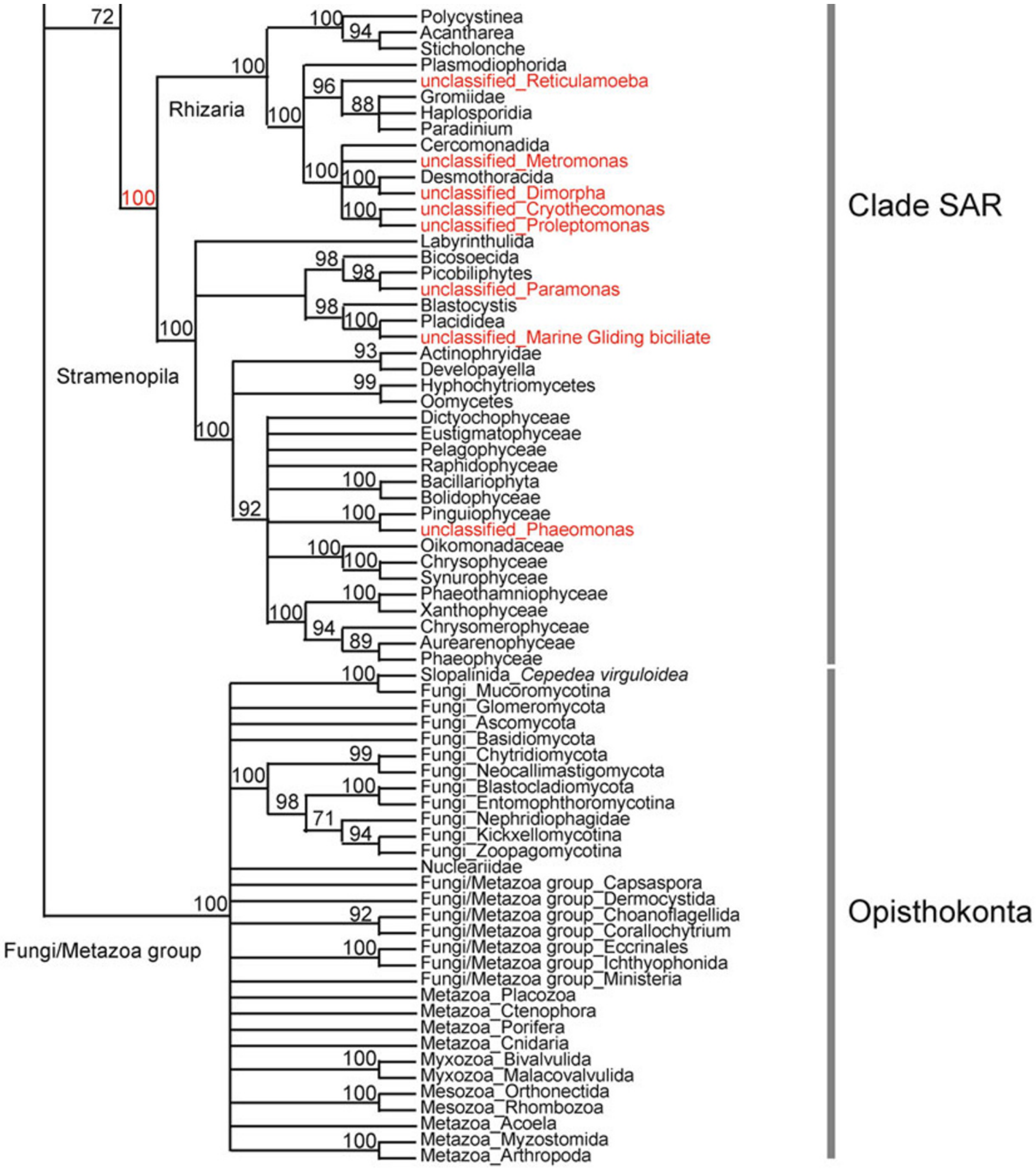


Figure 3. Cladogram of Eukaryota phylogeny based on the length-conserved parts of 18S rDNAs. The number above each branch is the posterior probability value in the Bayesian analysis. Only the values above 70% are shown. The words and number in red indicate the major new findings of this phylogenetic reconstruction.

uncertain (Burki et al., 2002; Kostka et al., 2004; Polet et al., 2004). As the sampled taxa in these studies all had plenty representatives of Alveolata, Rhizaria and Stramenopila, the taxon sampling is largely equivalent between different studies. So the inconsistency between different studies is very probably due to the ambiguity of alignment, which is caused by length extensive variation of 18S rDNA.

Second, the cladogram gave highly-supported positions to many unclassified eukaryotes. This will ease the quick identification of newly found protists based on 18S rDNAs in the future, and it makes 18S rDNA a good candidate for the DNA barcode of protists. Third, the cladogram is rooted with Archaea and the results supposed Parabasalidea and Chlorarachniophyceae to be among the basal lineages of eukaryotes. Nearly all cladograms in the previous studies have been unrooted or rooted with some eukaryotes, and that may greatly reduce the chance to know more about the details of the very early branching of eukaryotes.

CONCLUSIONS

The analyses on the 18S rDNA sequences of Foraminifera suggest that the positions of some variable length sites can be quite close to the translation functional parts of 18S rRNA.

The positions and lengths of the introns and variable regions of the 18S rDNAs of all eukaryotes show great diversity and suggest the tertiary structures as well as the secondary structures of 18S rRNAs can be more diverse than have been thought. The supposed origin of eukaryotic nucleus from archaean nucleoid is supported by six nucleotides shared by Archaea and Eukaryota.

The performance of 18S rDNA in phylogenetic studies can be improved by using only the length-conserved parts in the model of secondary structure. This may help to quickly determine what a protist is based on the sequence of its 18S rDNA, and evade the interference of extensive length variation.

MATERIALS AND METHODS

The longest 18S rDNA sequences from organisms of almost every known eukaryotic phylum were included. To compensate for differences in species diversity, the taxa were sampled at the class level in most Metazoa but at the order level in the diverse Hexapoda. The positions of introns were obtained from the summarization of GenBank records, or determined by the unusual disorder of length conserved regions in the results of the alignment program, CLUSTAL X (Thompson et al., 1997). The secondary structures of 18S rRNA were reconstructed by RNAStructure 4.6 (Mathews et al., 2004). The current comparative studies, on which the common secondary-structure model was based, involved the methods used in previous studies (Cannone et al., 2002), in which co-variation is the basic principle: the fewer the secondary structural elements, especially the paired regions, are destroyed by each sequence, the better the model is. The positions of the length-variable regions were summarized on the basis of the consensus result of the reconstructed secondary structures. Because of their exceptionally great lengths, the sequences of forams were additionally aligned independently, and aligned with the length-stable regions of the other sequences. There are three keys which may help to distinguish insertions from introns. First, insertions are obviously group specific while introns seldom. Second, insertions are usually much shorter when compared to introns. Additionally, the lengths of homologous insertions in different species within a taxon are continuously variable, while the lengths of homologous introns are disruptively variable. Third, when extensively long sequences of different organisms are independently aligned with common-length sequences without introns, the supposed positions of specific insertions and introns have never been completely the same.

In this study, the positions of the introns and the length-variable regions were marked in the secondary-structure model of the eukaryotic 18S rRNA with most detailed supplementary information up to now. At the tertiary level, as no fine structure of eukaryotic organisms is available, these positions were marked in the resolved structure of the 16S rRNA of *Thermus thermophilus*. The highly length-conserved parts between prokaryotes and eukaryotes can make it be extrapolated to eukaryote 18S rRNA. The positions of A-minor interactions (Noller, 2005) are also included.

For phylogenetic reconstructions, the best nucleotide substitution model was found to be GTR+I+Γ under AIC by Treefinder (Jobb et al., 2004). Based on this model, the program MrBayes 3.1.2 was run in parallel (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Parallel Mrbayes @ BioHPC). The parameters were set as follows: samplefreq = 1000, diagnfreq = 1000, nchains = 4, nst = 6, rates = invgamma. The number of generations run was 15,000,000. The number of the burn-in generations was 11,472,000. The data matrix file for the phylogenetic study, in nexus format, is provided as Fig. S4.

ACKNOWLEDGEMENTS

We thank Di Wu (Mr.), Gang Wang (Dr.) and Xiaoguang Liu (Dr.) of Nankai University for the helps in parallel running of the program MrBayes. This work was supported by the National Education Project in Basic Science for Special Subjects (Insect Systematics) [grant No. J0630963], the National Science Foundation Project [Grant No. 30970350] and the National Science Foundation Project for Distinguished Young Scholars [Grant No. 30725005].

ABBREVIATIONS

ESTs, expressed sequence taqs; ITS, internal transcribed spacers; rRNAs, ribosomal RNAs; tRNA, transfer RNA

REFERENCES

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290, 972–977.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289, 905–920.

Burki, F., Berney, C., and Pawlowski, J. (2002). Phylogenetic position of *Gromia oviformis* Dujardin inferred from nuclear-encoded small subunit ribosomal DNA. Protist 153, 251–260.

Burki, F., and Pawlowski, J. (2006). Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. Mol Biol Evol 23, 1922–1930.

Burki, F., Shalchian-Tabrizi, K., Minge, M., Skjaeveland, Å., Nikolaev, S.I., Jakobsen, K.S., and Pawlowski, J. (2007). Phylogenomics reshuffles the eukaryotic supergroups. PLoS One 2, e790.

Busse, I., and Preisfeld, A. (2003). Systematics of primary osmotrophic euglenids: a molecular approach to the phylogeny of *Distigma* and *Astasia* (Euglenozoa). Int J Syst Evol Microbiol 53, 617–624.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Müller, K.M., *et al.* (2002). The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. [Correction: BMC Bioinformatics 3, 15.] BMC Bioinformatics 3, 2.

Cavalier-Smith, T., and Chao, E.E. (1996). Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. J Mol Evol 43, 551–562.

Chandramouli, P., Topf, M., Ménétret, J.F., Eswar, N., Cannone, J.J., Gutell, R.R., Sali, A., and Akey, C.W. (2008). Structure of the mammalian 80S ribosome at 8.7 A resolution. Structure 16, 535–548.

Crease, T.J., and Colbourne, J.K. (1998). The unusually long small-subunit ribosomal RNA of the crustacean, *Daphnia pulex*: sequence and predicted secondary structure. J Mol Evol 46, 307–313.

Cunningham, C.O., Aliesky, H., and Collins, C.M. (2000). Sequence and secondary structure variation in the *Gyrodactylus* (Platyhelminthes: Monogenea) ribosomal RNA gene array. J Parasitol 86, 567–576.

Giribet, G., and Wheeler, W.C. (2001). Some unusual small-subunit ribosomal RNA sequences of Metazoans. Am Mus Novit 3337, 1–16.

Green, R., and Noller, H.F. (1997). Ribosomes and translation. Annu Rev Biochem 66, 679–716.

Hackett, J.D., Yoon, H.S., Li, S., Reyes-Prieto, A., Rümmele, S.E., and Bhattacharya, D. (2007). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. Mol Biol Evol 24, 1702–1713.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. Cell 107, 679–688.

Harper, J.T., Waanders, E., and Keeling, P.J. (2005). On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol 55, 487–496.

Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M., and Higgs, P.G. (2003). RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. Mol Phylogenet Evol 28, 241–252.

Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Jackson, S.A., Cannone, J.J., Lee, J.C., Gutell, R.R., and Woodson, S.A. (2002). Distribution of rRNA introns in the three-dimensional structure of the ribosome. J Mol Biol 323, 35–52.

Jobb, G., von Haeseler, A., and Strimmer, K. (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4, 18.

Jow, H., Hudelot, C., Rattray, M., and Higgs, P.G. (2002). Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol Biol Evol 19, 1591–1601.

Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., Pearlman, R.E., Roger, A.J., and Gray, M.W. (2005). The tree of eukaryotes. Trends Ecol Evol 20, 670–676.

Keller, A., Förster, F., Müller, T., Dandekar, T., Schultz, J., and Wolf, M. (2010). Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. Biol Direct 5, 4.

Kim, E., and Graham, L.E. (2008). EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. PLoS One 3, e2621.

Kostka, M., Hampl, V., Cepicka, I., and Flegr, J. (2004). Phylogenetic position of *Protoopalina intestinalis* based on SSU rRNA gene sequence. Mol Phylogenet Evol 33, 220–224.

Kumar, S., and Rzhetsky, A. (1996). Evolutionary relationships of eukaryotic kingdoms. J Mol Evol 42, 183–193.

Margulis, L. (1970). Origin of Eukaryotic Cells. New Haven, Connecticut: Yale University Press..

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A 101, 7287–7292.

Neefs, J.M., Van de Peer, Y., De Rijk, P., Goris, A., and De Wachter, R. (1991). Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 19, 1987–2015.

Nikolaev, S.I., Berney, C., Fahrni, J.F., Bolivar, I., Polet, S., Mylnikov, A.P., Aleshin, V.V., Petrov, N.B., and Pawlowski, J. (2004). The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci U S A 101, 8066–8071.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science 289, 920–930.

Noller, H.F. (1991). Ribosomal RNA and translation. Annu Rev Biochem 60, 191–227.

Noller, H.F. (2005). RNA structure: reading the ribosome. Science 309, 1508–1514.

Parallel Mrbayes @ BioHPC. (2011). http://cbsuapps.tc.cornell.edu/mrbayes.aspx

Parfrey, L.W., Barbero, E., Lasser, E., Dunthorn, M., Bhattacharya, D., Patterson, D.J., and Katz, L.A. (2006). Evaluating support for the current classification of eukaryotic diversity. PLoS Genet 2, e220.

Patron, N.J., Inagaki, Y., and Keeling, P.J. (2007). Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol 17, 887–891.

Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J., Gooday, A.J., Cedhagen, T., Habura, A., and Bowser, S.S. (2003). The evolution of early Foraminifera. Proc Natl Acad Sci U S A 100, 11494–11498.

Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: impact of missing data on large alignments. Mol Biol Evol 21, 1740–1752.

Polet, S., Berney, C., Fahrni, J., and Pawlowski, J. (2004). Small-subunit ribosomal RNA gene sequences of Phaeodarea challenge the monophyly of Haeckel's Radiolaria. Protist 155, 53–63.

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. Cell 108, 557–572.

Rodríguez-Ezpeleta, N., Brinkmann, H., Burey, S.C., Roure, B., Burger, G., Löffelhardt, W., Bohnert, H.J., Philippe, H., and Lang, B. F. (2005). Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol 15, 1325–1330.

Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., *et al.* (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102, 615–623.

Schöniger, M., and von Haeseler, A. (1994). A stochastic model for the evolution of autocorrelated DNA sequences. Mol Phylogenet Evol 3, 240–247.

Schultz, J., and Wolf, M. (2009). ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics. Mol Phylogenet Evol 52, 520–523.

Seibel, P.N., Müller, T., Dandekar, T., Schultz, J., and Wolf, M. (2006). 4SALE—a tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics 7, 498.

Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vaulot, D., Minge, M.A., Le Gall, F., Romari, K., Throndsen, J., Botnen, A., Massana, R., *et al.* (2006). Telonemia, a new protist phylum with affinity to chromist lineages. Proc Biol Sci 273, 1833–1842.

Shalchian-Tabrizi, K., Kauserud, H., Massana, R., Klaveness, D., and

Jakobsen, K.S. (2007). Analysis of environmental 18S ribosomal RNA sequences reveals unknown diversity of the cosmopolitan phylum Telonemia. Protist 158, 173–180.

Siebert, S., and Backofen, R. (2005). MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. Bioinformatics 21, 3352–3359.

Spahn, C.M.T., Beckmann, R., Eswar, N., Penczek, P.A., Sali, A., Blobel, G., and Frank, J. (2001). Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. Cell 107, 373–386.

Stocsits, R.R., Letsch, H., Hertel, J., Misof, B., and Stadler, P.F. (2009). Accurate and efficient reconstruction of deep phylogenies from structured RNAs. Nucleic Acids Res 37, 6184–6193.

Sweeney, R., Chen, L., and Yao, M.C. (1994). An rRNA variable region has an evolutionarily conserved essential role despite sequence divergence. Mol Cell Biol 14, 4203–4215.

Telford, M.J., Wise, M.J., and Gowri-Shankar, V. (2005). Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. Mol Biol Evol 22, 1129–1136.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24, 4876–4882.

Van de Peer, Y., and De Wachter, R. (1997). Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. J Mol Evol 45, 619–630.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M. Jr, Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature 407, 327–339.

Wolf, M., Ruderisch, B., Dandekar, T., Schultz, J., and Müller, T. (2008). ProfDistS: (profile-) distance based phylogeny on sequence—structure alignments. Bioinformatics 24, 2401–2402.

Wuyts, J., Perrière, G., and Van De Peer, Y. (2004). The European ribosomal RNA database. Nucleic Acids Res 32, D101–D103.

Wuyts, J., Van de Peer, Y., and De Wachter, R. (2001). Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Res 29, 5017–5028.

Xie, Q., Tian, X., Qin, Y., and Bu, W. (2009). Phylogenetic comparison of local length plasticity of the small subunit of nuclear rDNAs among all Hexapoda orders and the impact of hyper-length-variation on alignment. Mol Phylogenet Evol 50, 310–316.

Xie, Q., Tian, Y., Zheng, L., and Bu, W. (2008). 18S rRNA hyper-elongation and the phylogeny of Euhemiptera (Insecta: Hemiptera). Mol Phylogenet Evol 47, 463–471.

Yusupov, M.M., Yusupova, G.Z.H., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 A resolution. Science 292, 883–896.