Molecular Phylogenetics and Evolution 50 (2009) 310–316

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

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

Qiang Xie [a], Xiaoxuan Tian [a], Yan Qin [b], Wenjun Bu [a,*]

[a] Institute of Entomology, College of Life Sciences, Nankai University, 94 Weijin Road, Nankai District, Tianjin 300071, China
[b] National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 12 May 2008
Revised 25 October 2008
Accepted 30 October 2008
Available online 7 November 2008

Keywords:
Secondary structure
rDNA
Hexapoda
Phylogenetic

ABSTRACT

The SSU nrDNA (18S), is one of the most frequently sequenced molecular markers in phylogenetic studies. However, the length-hyper-variation at multiple positions of this gene can affect the accuracy of alignment greatly and this length variation makes alignment across arthropod orders a serious problem. The analyses of Hexapoda phylogeny is such a case. A more clear recognition of the distribution of the length-variable-regions is needed. In this study, the secondary structure of some length-variable-regions in the SSU nrRNA of Arthropoda was adjusted by the principle of co-variation. It is found that the extent of plasticity of some length-variable-region can extraordinarily be higher than 600 bases in hexapods. And the numbers of hyper length-variable-regions are largest in Strepsiptera and Sternorrhyncha (Hemiptera). Our study shows that some length-variable-regions can serve as synapomorphies for some groups. The phylogenetic comparison also suggested that the expansion of a lateral bulge could be the origin of a helix.

1. Introduction

It has been well known that the lengths of certain regions in the small subunit of nuclear ribosomal DNA (SSU nrDNA, which is also known as 18S rDNA), are not conservative among different groups and sometimes the difference can be huge. Length differences are seldom considered at a fine scale in phylogenetic studies. The reasons may include that the local length variations are not always noticed and/or included, or the relatively length-conservative parts in the complete sequences are favored in phylogenetic reconstructions.

Hexapods amount to more than half of the recognized species biodiversity of cellular life (Groombridge, 1992), yet relationships among hexapod orders remain controversial. Although it is clear that alignment rather than reconstruction algorithms *per se* have a large influence on phylogenetic results (Goldman, 1998; Morrison and Ellis, 1997), there are few comparisons of the nrDNAs that include all Hexapoda orders. The SSU nrDNA has been used in almost all molecular phylogenetic studies of Hexapoda (Caterino et al., 2000). These studies often generate different results for some groups, such as the monophyly of Ellipura (Whiting et al., 1997; Wheeler et al., 2001; Giribet et al., 2004; Kjer, 2004; Luan et al., 2005; Kjer et al., 2006a), the sister group of Neoptera (Whiting et al., 1997; Wheeler et al., 2001; Kjer, 2004; Terry and Whiting, 2005; Kjer et al., 2006a), the position of Zoraptera (Wheeler et al., 2001; Terry and Whiting, 2005; Yoshizawa and Johnson, 2005), the phylogeny of Paraneoptera (Wheeler et al., 2001; Kjer, 2004), and etc. The large subunit of nuclear rDNA (LSU nrDNA) of a-few-hundred-base long and/or different protein coding genes were included variously in the data sets (Whiting et al., 1997; Wheeler et al., 2001; Giribet et al., 2004; Terry and Whiting, 2005; Kjer et al., 2006a). Occasionally, homeobox genes (Rokas et al., 1999) or protein sequences (Nardi et al., 2003a) were also used. However, convincing evidence has not been provided for these genes (Delsuc et al., 2003; Nardi et al., 2003b).

In previous studies, it seemed that when the taxon sampling was good in a study, only partial but not complete sequences of SSU nrDNAs were used, and when complete sequences were used in a study, some orders or suborders with hyper-extensive SSU nrDNAs were not sampled (Wheeler et al., 2001; Kjer, 2004; Ogden et al., 2005; Terry and Whiting, 2005). It should be noted that, certain order-level taxa were sometimes excluded from data sets in the phylogenetic studies of Hexapoda (e.g., Strepsiptera and Sternorrhyncha in Hemiptera), while the SSU nrDNAs of these taxa generally were extraordinarily long ones. Among the hyper-extensive SSU nrDNAs, Strepsiptera had the longest one (Gillespie et al., 2005). Whether the complete sequences of SSU nrDNAs were used, and how gaps were treated in alignment also varied among different studies. Generally, the indels were either removed (Kjer, 2004;

* Corresponding author. Fax: +86 22 23503711.
E-mail address: wenjunbu@nankai.edu.cn (W. Bu).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.10.025

Luan et al., 2005) or analyzed by direct optimization procedures (Wheeler, 1996; Wheeler et al., 2001; Wheeler and Gladstein, 2003; Terry and Whiting, 2005).

The model of SSU nrRNA secondary structure in previous studies was stable for most length-variable-regions (LVRs). However, the estimated local structures of a few LVRs are not stable yet, which has become an impediment in recognizing the plastic extent of each LVR in all hexapods orders and in determining the treatment of LVRs in phylogenetic studies based on SSU nrDNA. Although the phenomenon of the local length variation among SSU nrDNAs of different groups has been known for some years, the comparison of each length-variable-region for all hexapod orders and the consideration on the effect of the hyper-extensive regions, i.e., the mass insertions at some positions of SSU nrDNA, on alignment are few. The hyper-extensive regions often make regional-position-homology inaccurate in alignment. Knowing the details of the length variation in each hexapod order is important to learn its effect on the alignment, as shown in Fig. 1. If the details of both LVRs (LVR1 and LVR2) in each sampled taxon (a and b, Fig. 1) are known, we are able to judge whether the regional homology have been properly determined in alignment by comparing the actual length of alignment with the length summation of all of the length-stable regions and the longest ones of each length-variable-region.

Complete SSU nrDNA sequences of each Hexapoda order are available in GenBank. In some groups, the length expansions are extraordinary. The whole-length of the longest SSU nrDNA (Strepsiptera) was almost twice as long as that of the shortest ones (Chalwatzis et al., 1995). And the extension of plasticity can vary significantly among suborders or even families, for example in Strepsiptera and Hemiptera, as well as among orders. To know the position and plastic extension of each LVR, all Hexapoda orders should be included and the groups with high length plasticity should be sampled intensively due to the length diversity of their rDNAs. This will also facilitate the determination on the position of each LVR in alignment and the selection among various local secondary structures of each LVR in the construction of consensus secondary structure model of SSU nrRNA. Although secondary structure models of LVRs have been available in Diptera (model organism *Drosophila melanogaster*) (Cannone et al., 2006), Hemiptera (Ouvrard et al., 2000), Hymenoptera (Gillespie et al., 2006), Strepsiptera (Gillespie et al., 2005), and etc., these models vary in V2, V4 and V7 regions (the region nomenclature follows Neefs et al., 1993). Gillespie et al. (2005) compared some LVRs among different groups of Hexapoda. In addition, four orders, Mantophasmatodea, Megaloptera, Raphidoptera, and Zoraptera, and suborder Sternorrhyncha (Paraneoptera: Hemiptera) are sampled in this study. And the plasticity of each LVR is shown in illustrations instead of in numerals.

Both incomplete taxon sampling and vagueness of positional homology in alignment in large scales have hindered studies on the alignment of Hexapoda SSU nrDNA. This study focuses on the phylogenetic comparison of each LVR and the role of the hyper-length-variation in the alignment of Hexapoda SSU nrDNA. Sequences from each Hexapoda order are included (see Section 2). The aim of this work is not to reconstruct a more reliable Hexapoda phylogeny based on SSU nrDNA, but to clarify an impediment in reconstructing Hexapoda phylogeny based on SSU nrDNA.

2. Methods

2.1. Secondary structure model rebuilding

The thermodynamic secondary structures of regions V2, V4 and V7 (*sensu* Neefs et al., 1993), whose alignments have been inconsistent in previous studies, were calculated using the program RNA structure 4.5 (Mathews et al., 2004). Only differences among various models were re-calculated. The re-calculated result in this study, along with the alternative corresponding parts in existing models, were then selected under the principle of co-variation: the fewer the secondary structural elements, especially the paired regions, are destroyed by each sequence, the better the model is (Gutell et al., 1985, 2000, 2002). We simplified the principle in this study as: the fewer the paired regions are destroyed by each sequence, the better the model is. Or, the longer the stems are kept by each sequence, the better the model is.

The length plasticity information was plotted in Fig. 3B, with reference to the phylogenetic pattern shown in Fig. 3A. The width of each short vertical line in Fig. 3B equals to $3 \times \mathrm{Log}_{10}N$, with N standing for the length of each LVR in each taxon.

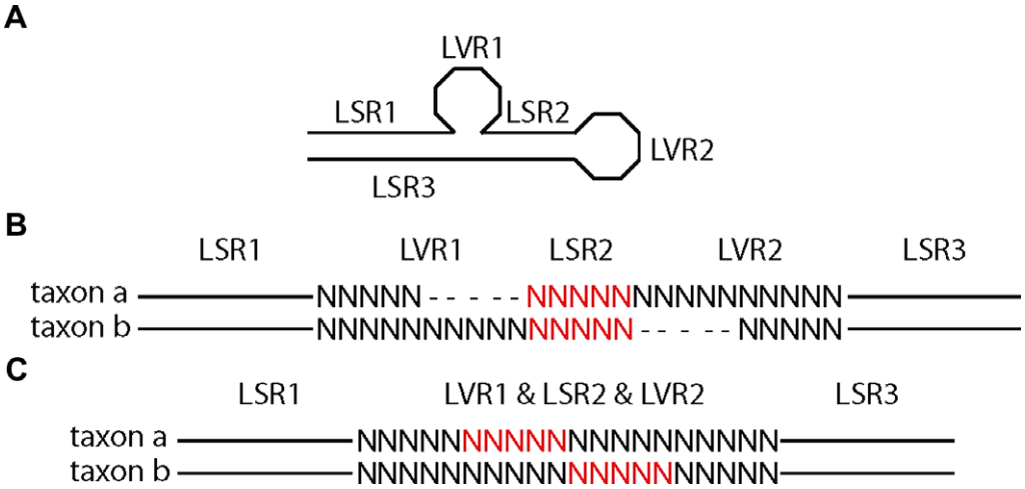

Fig. 1. (A) A schematic presentation of the secondary structure of SSU nrDNA. The lateral and terminal bulges are two length-variable-regions (LVR1 and 2, respectively) and the flanking stems are length-stable regions (LSR1–3). (B) In taxa a and b, the length of LVR1 are 5 bp and 10 bp, respectively, and the length of LVR2 are 10 bp and 5 bp, respectively. The length of LSR2 is 5 bp. If the regional homologies are properly determined in alignment, the length summation of LVR1, LSR2 and LVR2 will be 25-column-long in alignment result. (C) If the regional homologies are improperly determined in alignment, the length summation of LVR1, LSR2 and LVR2 will be shorter than 25-column-long in alignment result.

2.2. Phylogenetic reconstruction

All Hexapoda orders were sampled. Most orders were sampled to suborder-level while some were sampled to family-level as they do not have suborder taxonomic hierarchies. All orders with great SSU nrDNA length variations were also sampled to family-level to accommodate thorough length representatives in the order (see Supplementary information). It may appear that the taxa sampling is biased in some orders in terms of species diversity. However, the length diversity of each LVR in Hexapoda orders is not correlative to species diversity. The extent of length plasticity within each order has to be considered in taxon sampling. Thus, even though the species diversity in Strepsiptera is significantly lower than that of Coleoptera, Strepsiptera deserves four representatives in our taxon sampling as does Coleoptera because of its extreme length variation of the SSU nrDNAs. In phylogenetic reconstruction, all LVRs were removed from the original sequences. The cladogram search was done with MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The parameters were set as: ngen = 5,000,000, samplefreq = 100, nchains = 4, nst = 6, rates = invgamma, burnin = 3,000,000. The respective monophyly of Hexapoda, Pterygota, Eumetabola, Paraneoptera, Phasmatodea, Orthoptera, Hemiptera, Coleoptera, and Hymenoptera are constrained for the reason that they are largely accepted by now (Carmean et al., 1992; Pashley et al., 1993; Whiting et al., 1997; Wheeler et al., 2001; Giribet et al., 2004; Kjer, 2004; Terry and Whiting, 2005; Kjer et al., 2006a). As it is better to leave the phylogenetic schemes referred in this study unresolved than to be biased, the branches that were unstable in previous studies are collapsed, e.g., the relationships between Odonata, Ephemeroptera, and Neoptera (Whiting et al., 1997; Wheeler et al., 2001; Kjer, 2004; Terry and Whiting, 2005; Kjer et al., 2006a).

3. Results

The consensus model of the secondary structure of SSU nrDNA is shown in Fig. 2A. It is mostly as same as the model based on *D. melanogaster* (Cannone et al., 2006), with only two local differences: the structure of V2 agrees with Gillespie et al. (2005); and the 3′ region of E23 is favored by the principle of co-variation. The LVRs O and P of the 3′ region of E23 are located at bulges in all sequences in the model shown in Fig. 2A, whereas they are located at stem regions in some sequences in the previous models. When some hyper-variable-regions were close to each other in the primary structure of nucleic acids, it would be impossible to reasonably recover their secondary structures solely by thermodynamic methods and alignment results. Therefore, the adjusted model should be the one with nearly identical thermodynamic stability among multiple alternative structures and is favored by the principle of co-variation and the compensatory and/or semi-compensatory base changes.

In the adjusted model, the local length plasticity of Hexapoda SSU nrDNA were distributed in 24 independent LVRs except for specific indels (Fig. 2A). Some of them were very close to each other and the length-stable regions between them were quite short. Most LVRs were lateral or terminal bulges. The LVRs with the highest length plasticity were all in lateral or terminal bulges. Most LVRs were restrained in three regions, which were previously named as V2, V4 and V7 (Neefs et al., 1993). The sum of variations for the ten variable bulges in V4 (I∼P) accounted for nearly half of the total variations for all LVRs, whereas the six bulges in V2 (B∼G) accounted for about 20% and the four bulges in V7 (S∼V) accounted for about a quarter. Therefore, the three major length-variable-regions in hexapods SSU nrDNA/RNA account for nearly all length variations.

Entognatha *sensu* (Hennig, 1953, 1969) include Collembola, Diplura and Protura. Among these three basal Hexapoda groups, there are no large expansions in Collembola, whereas there are distinct common expansions in Diplura and Protura at positions D, K, L and W (Fig. 3B). Among Hexapoda, Diplura and Protura even share a unique insertion position in the stem where the LVR W positioned upstream (Fig. 2A). In Collembola, Diplura, and Protura, the lengths of various LVRs, e.g., A, J, Q and S, are order-specific (shown in the Supplementary information).

Among Insecta *sensu* Kristensen (1991, 1995) or Ectognatha *sensu* Hennig, 1953, 1969), the length of LVR L of "Apterygota" is distinctly shorter than that of "Paleoptera". And the latter is distinctly shorter than that of Neoptera. As for "Apterygota" and "Paleoptera", the length of each LVR is relatively consistent.

Among "Polyneoptera" there are several order-specific expansions at various bulges, including S in Mantophasmatodea and P in Dictyoptera. The number of length-identical regions is greatest between Zoraptera and Dermaptera (19 out of the total 24 regions, see Supplementary information). Among "Polyneoptera", the lengths of regions J, K, M and T are unique in Haplocerata (= Zoraptera + Dermaptera, *sensu* Terry and Whiting, 2005). Plecoptera has specific expansions at U and W and specific length at J.

Among Paraneoptera, Sternorrhyncha in Hemiptera has the most expansive regions, in terms of degree of plasticity and number of hyper-variable-regions. Psocodea *sensu* Lyal (1985) and Thysanoptera rank the second and the third, respectively, in Paraneoptera. On the other hand, LVR lengths in Euhemiptera, the sister group of Sternorrhyncha, and the remaining Hemiptera, are the most stable and shortest. Bulge lengths are rarely group-specific in Holometabola. Strepsiptera has the most expansive regions in terms of degree of plasticity and number of hyper-variable-regions. Additionally, the previously reported AT-rich phenomenon (Kjer, 2004; Gillespie et al., 2005) occurs in some hyper-expansive bulges.

Overall, in different groups, expansions in the entire length of nrDNA occur in different regions. Strepsiptera has the largest number of hyper-expansive bulges in all Hexapoda, followed by Sternorrhyncha in Hemiptera. Although the length of SSU nrRNA in Diplura, Mantophasmatodea, and Neuropterida is long, most expansions in these groups are restricted in one or two bulges. In terms of degree of plasticity and number of hyper-variable-regions, the expansive regions in Phthiraptera and Plecoptera are less than those in Strepsiptera and Sternorrhyncha and largely more than those in Diplura, Mantophasmatodea, and Neuropterida.

4. Discussion

The lengths of SSU nrDNAs are also highly expansive in some archaea, algae, and fungi. However, these length expansions are different from those in Arthropoda, Mollusc (Lindgren and Daly, 2007) and other invertebrates. The length variation in archaea, algae, and fungi is mainly caused by the existence of group I introns whereas that of Arthropoda and other invertebrates is caused by the expansion in LVRs. Based on what we know from GenBank data, group I introns mostly exist in single-cellular organisms, while LVRs are mainly observed in multi-cellular invertebrates.

Although LVRs are mostly located at bulges on the consensus model built on secondary structure, not all bulges are length-variable. In fact, most bulges are length-conservative. The length-conservative regions form a skeleton of the model, which is universal in Arthropoda. In the adjusted secondary structure model of Hexapoda SSU nrRNA, the stem length between some hyper-variable bulges can be as short as a few base-pairs, e.g., the ones flanking K and T. These short intervals are always concealed in the gap-rich regions in alignments and can hardly be identified.

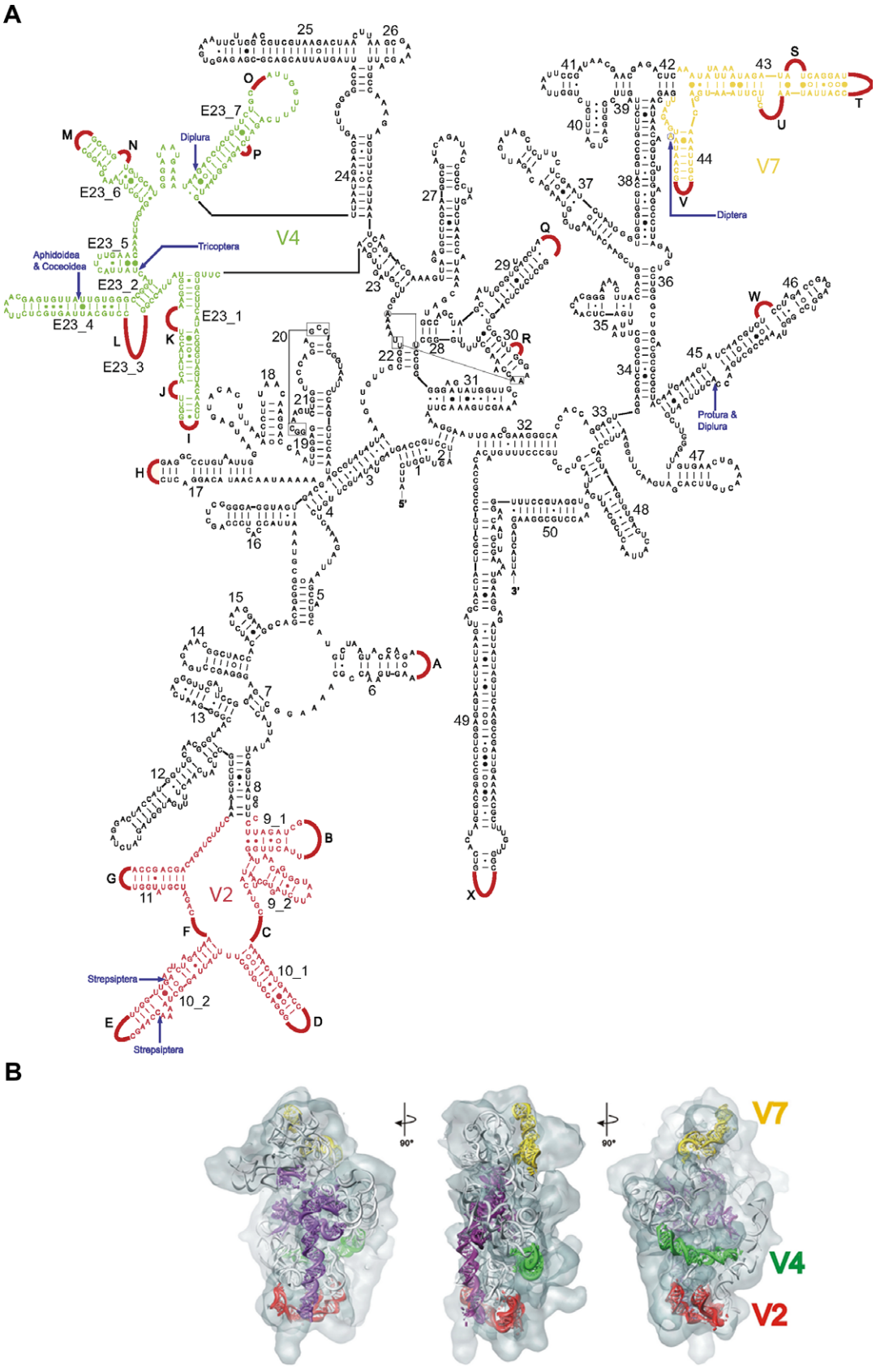

Fig. 2. (A) Adjusted secondary structure model of the SSU nrRNA in Hexapoda. The sequence was from *Drosophila melanogaster* (GenBank Accession No. M21017). Red curved lines labeled as A to X represent 24 length-variable-regions. Blue arrows indicate positions where there would be specific insertions in some taxa. Base pairing is indicated as follows: standard canonical pairs by lines (G–C, A–U), wobble G:U pairs by dots (G•U), A:G or A:C pairs by open circles (A○G, A○C), and other non-canonical pairs by filled circles (e.g., A•A). (B) the 16S rRNA from *T. thermophilus* (PDB:1S1H) fitted into the cryo-electron map of yeast small ribosomal subunit from 3D-EM database (Accession No. 1067). Left: intersubunit surface view; middle: side view; right: cytoplasm surface view. The translation functional domains (purple), V2 (red), V4 (green), and V7 (yellow). For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.

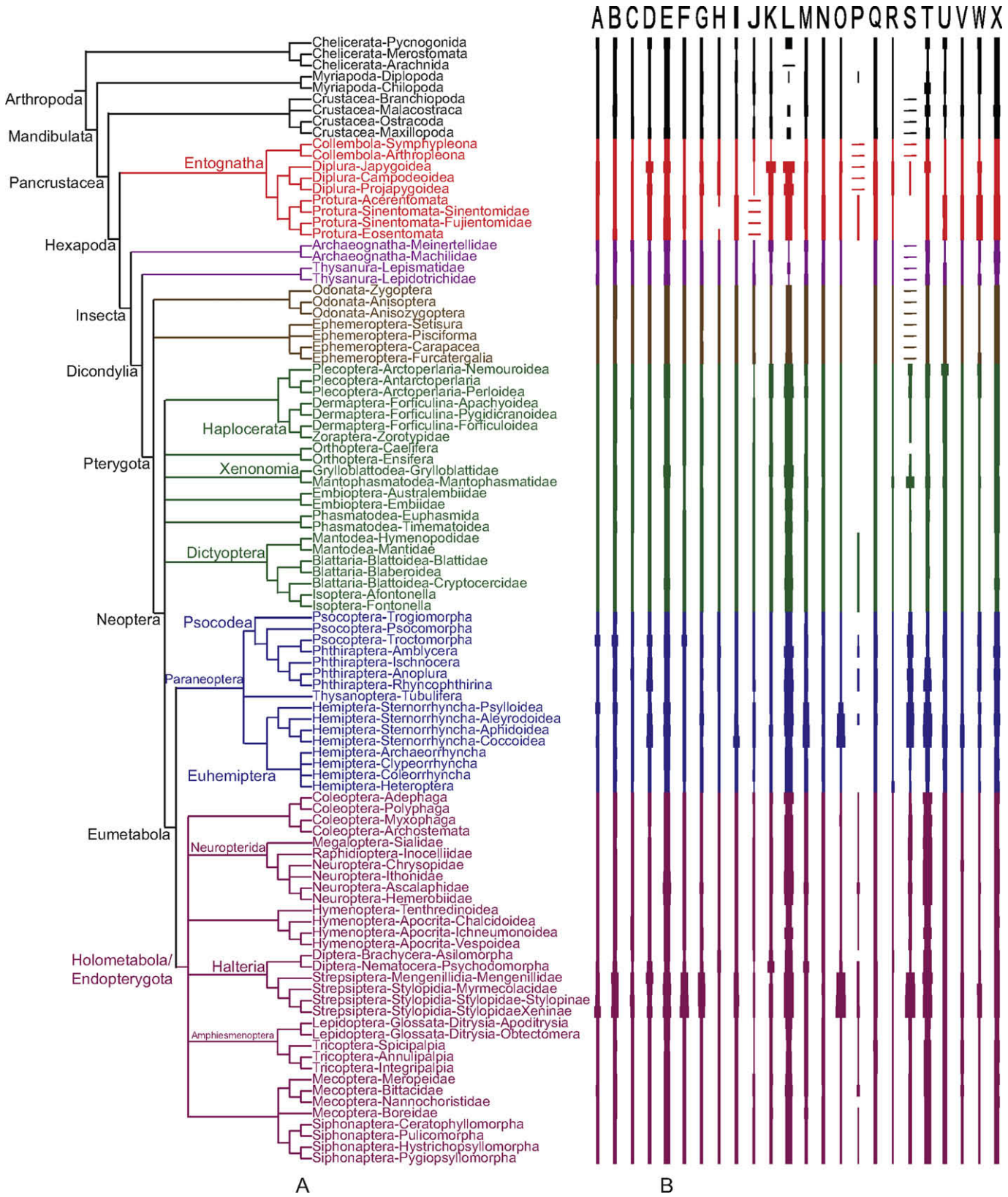

Fig. 3. Comparison of bulge length at different LVR with reference to Hexapoda phylogeny. (A) The hypothesized phylogeny of Hexapoda based on Bayesian analysis on the length-stable part of SSU nrDNAs and existing studies. (B) Plotting the length of variable regions of each operational taxonomic unit (OTU) in the width of a short line, in proportion to the logarithm of its length. Zero and single base length are indicated by "–" and blank, respectively.

When some LVRs are close to each other along the sequence and the level of length plasticity is high in at least one of them, these LVRs always collapse and are merged as one region in alignment. This phenomenon caused the loss of their respective region-positional homology on the whole. The occurrence of the region-positional homology loss could be detected by the comparison of the

number of columns in alignment and the sum of the length of the stable regions along with the longest one of each LVR. If the former is shorter than the latter, then it is very likely that region-mergence has taken place. Taking the sequences sampled in this study for example, if all of the regional homology are well determined, the length of an ideal alignment should reach about 4200 columns. It seems that this condition can hardly be fulfilled by the algorithms in current computer-based alignment programs. For rDNA, alignment should be based on the biological background given by the rRNA secondary structure model. And in this circumstance, manual alignments are needed (Kjer et al., 2006b). At present, it seems that some hyper-variable-regions, e.g., LVRs L and T (Fig. 2A), have to be removed in aligning to avoid the loss of positional homology in large-range, whereas other LVRs may be retained to maximize phylogenetic information in nrDNAs. As some expansive regions do not impact the region-positional homology, probably the regions that can jeopardize the large-range positional homology in alignment can be defined as the hyper-extensive ones. This condition is not as same as those reported in small ranges (Xie et al., 2005), where the simultaneous positional homology of the paired bases in the stems are manually recovered if they are not generated in alignment.

The LVR distribution pattern varies from one position to another according to the Hexapoda phylogeny. Therefore, from the view of Hexapoda phylogeny, it would be unreasonable to define whole-length expansions as synapomorphies of the descendents in some order or suborder, as has been done previously (Campbell et al., 1994). No bulge expansion has single origin according to Hexapoda phylogeny. However, some bulge expansions can be defined as synapomorphies. In the instance of Entognatha phylogeny, the monophyly of Ellipura (= Collembola + Protura) has been morphologically recognized by tremendous amount of studies (Hennig, 1953, 1969; Kristensen, 1975, 1981, 1995, 1998; Boudreaux, 1979; Kukalová-Peck, 1987, 1991; Stys and Bilinski, 1990; Stys et al., 1993; Kraus and Kraus, 1994; Stys and Zrzavy, 1994; Koch, 1997; Kraus, 1998, 2001; Ax, 1999; Carpenter and Wheeler, 1999; Edgecombe et al., 2000; Wheeler et al., 2001) yet the Diplura + Protura clade has never been considered by morphologists (Giribet et al., 2004). Although the support for Ellipura based on internal anatomical and external morphological characters may not be unambiguous (Bitsch and Bitsch, 1998, 2000) and the monophyly of Diplura + Protura is stable in molecular analysis and the combined analysis of molecular and morphological data (Giribet and Ribera, 2000; Giribet and Wheeler, 2001; D'Haese, 2002; Luan et al., 2005), few synapomorphies for Diplura + Protura have been raised. The concepts "molecular morphometrics" and "morpho-molecular" were mentioned by Billoud et al. (2000) and Ouvrard et al. (2000), respectively. In this study, the unique insertion position shared by Diplura and Protura among Hexapoda and the expansions shared by them among Entognatha can serve as synapomorphies. For the support of Dermaptera + Zoraptera, there are also synapomorphies indicated by secondary structure. The hyper-expansions at position O can serve as synapomorphies for Strepsiptera and Sternorrhyncha. It should be noted that, as all LVRs are removed from the data matrix in inferring phylogenetic relationships (see Section 2), the length expansions have not contributed to the topology of phylogenetic tree. Furthermore, the monophyly of Diplura + Protura and Dermaptera + Zoraptera can still be recovered in Bayesian analysis with 100% node support values, even if these nodes are not constrained *a priori* (see Section 2).

In terms of plasticity level of the Hexapoda LVRs, although V4 and V7 have been known for a decade as positions where length variation could occur, few studies have shown that the length plasticity of a lateral bulge as well as a terminal one could exceed 300 nts, as shown in E23-3, E23-7 and helix 43 (position L, O and S in Fig. 2A). In Diplura and Protura, E23-3 and E23-4 always exist as one helix in the secondary structure reconstruction and the length plasticity extent can reach higher than 600 nts within Diplura. The expansions at positions L and S suggest that the expansion of the lateral bulge could be a mode for the origin of a helix. Additionally, the phylogenetic comparison of the LVRs also shows that the phenomenon of length plasticity of SSU nrDNA takes place more or less in all Hexapoda orders. This study compared the secondary structures of SSU nrDNAs among Hexapoda orders thoroughly and phylogenetically. As a consequence, more questions have been raised. For example, is the local length plasticity in rRNA identical to those in rDNA? Are length variations at different bulges related? What effects, if any, are caused by the LVRs in the tertiary structure and its corresponding functions? The fit-in position of V2, V4, and V7 in Fig. 2B shows the quasi-localization of these three regions in the small ribosome subunit serially. Compared to the translation function core (Fig. 2B left, purple colored), which is prominently distributed in the intersubunit surface that interacts with large ribosomal subunit and is responsible for translational decoding functionality, the V2, V4, and V7 regions are located predominantly in the surfaces of the subunit that are exposed to cytoplasm (Fig. 2B, right). A good comparison given by the side view (Fig. 2B, middle) reveals that the LVRs are located faraway from the universally conserved translation function center on the small ribosomal subunit. Instead, they reside in the cytoplasm surface that is relevant to species/order-specificity as concluded by structural analysis of the ribosomes between prokaryotes and eukaryotes (Wilson and Nierhaus, 2003). Studies on the LSU nrRNA also suggested that LVRs usually do not have direct effects on the translation function of the ribosome (Gerbi, 1985; Sweeney et al., 1994), thus their structural constraints are not very rigorous. However, it seems that when an unusually high plastic extent is found, these questions may deserve reconsideration. Undoubtedly, if the length expansions in rRNA are largely the same as that in rDNA, the hyper-extensive ones will be energy-inefficient. Because the knowledge on model organisms is far from completion, it is crucial to introduce the view and methods of phylogenetic comparison into molecular structural biology.

Acknowledgments

We thank Prof. Jin-Tang Dong (Emory University, USA), Prof. Karl Kjer (Rutgers University, USA), Dr. Xin Zhou (University of Guelph, Canada), Dr. Joseph Gillespie (Texas A&M University, USA), Dr. Jinsheng Yang (Massachusetts General Hospital, Harvard Medical School, USA) and three anonymous reviewers for reviewing the manuscript and providing very helpful advices. We thank Mr. Jinzhong Lin and Dr. Xiaopeng Zhu (Institute of Biophysics, Chinese Academy of Sciences, China.) and Ms. Li-Ping Wang (Nankai University, China) for the preparation of illustrations and the helpful discussions. This project was supported by National Natural Science Foundation of China (No. 30600063, 30725005), the National Education Project in Basic Science for Special Subjects (Insect Systematics, No. J0630963, Research Fund for the Doctoral Program of Higher Education (No. 20050055027) and National Basic Research Program of China (No. 2006CB910903).

Appendix A. Supplementary data

Detailed expansion information for each taxon at each length-variable bulge is provided in the supplementary information. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.10.025.

References

Ax, P., 1999. Das System der Metazoa II. Ein Lehrbuch der phylogenetischen Systematik. Gustav Fischer Verlag, Stuttgart.

Billoud, B., Guerrucci, M.A., Masselot, M., Deutsch, J.S., 2000. Cirripede phylogeny using a novel approach: molecular morphometrics. Mol. Biol. Evol. 17, 1435–1445.

Bitsch, C., Bitsch, J., 1998. Internal anatomy and phylogenetic relationships among apterygote insect clades (Hexapoda). Ann. Soc. Entomol. Fr. NS 34, 339–363.

Bitsch, C., Bitsch, J., 2000. The phylogenetic interrelationships of the higher taxa of apterygote hexapods. Zool. Screen. 29, 131–156.

Boudreaux, H.B., 1979. Arthropod Phylogeny with Special Reference to Insects. Wiley, New York.

Campbell, B.C., Steffen-Campbell, J.D., Gill, R.J., 1994. Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha: Aleyrodidae) inferred from 18S rDNA sequences. Insect Mol. Biol. 3, 73–88.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2006. Available from: <http://www.rna.icmb.utexas.edu/>.

Carmean, D., Kimsey, L.S., Berbee, M.L., 1992. 18S rDNA sequences and holometabolous insects. Mol. Phylogenet. Evol. 1, 270–278.

Carpenter, J.M., Wheeler, W.C., 1999. Cladística numérica, análisis simultáneo y filogenia de hexápodos. In: Melic, A., de Haro, J.J., Méndez, M., Ribera, I. (Eds.), Filogenia y Evolución de Arthropoda. Bol. Soc. Entomol. Aragon. 26, 333-346.

Caterino, M.S., Cho, S., Sperling, F.A.H., 2000. The current state of insect molecular systematics: a thriving tower of Babel. Annu. Rev. Entomol. 45, 1–54.

Chalwatzis, N., Baur, A., Stetzer, E., Kinzelbach, R., Zimmermann, F.K., 1995. Strongly expanded 18S rRNA genes correlated with a peculiar morphology in the insect order of Strepsiptera. Zoology 98, 115–126.

Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?". Science 301, 1482d.

D'Haese, C.A., 2002. Phylogeny of apterygote hexapods. Abstracts, 20th Annual Meeting of the Willi Hennig Society. Cladistics 18, 220.

Edgecombe, G.D., Wilson, G.D.F., Colgan, D.J., Gray, M.R., Cassis, G., 2000. Arthropod cladistics: combined analysis of Histone H3 and U2 snRNA sequences and morphology. Cladistics 16, 155–203.

Gerbi, S.A., 1985. Evolution of ribosomal DNA. In: MacIntyre, R.J. (Ed.), Molecular Evolutionary Genetics. Plenum, New York, pp. 419–517.

Gillespie, J.J., McKenna, C.H., Yoder, M.J., Gutell, R.R., Johnston, J.S., Kathirithamby, J., Cognato, A.I., 2005. Assessing the odd secondary structural properties of nuclear small subunit ribosomal RNA sequences (18S) of the twisted-wing parasites (Insecta: Strepsiptera). Insect Mol. Biol. 14, 625–634.

Gillespie, J.J., Johnston, J.S., Cannone, J.J., Gutell, R.R., 2006. Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): structure, organization, and retrotransposable elements. Insect Mol. Biol. 15, 657-686.

Giribet, G., Ribera, C., 2000. A review of arthropod phylogeny: new data based on ribosomal DNA sequences and direct character optimization. Cladistics 16, 204–231.

Giribet, G., Wheeler, W.C., 2001. Some unusual small-subunit ribosomal RNA sequences of Metazoans. Am. Mus. Novit. 3336, 1–14.

Giribet, G., Edgecombe, G.D., Carpenter, J.M., D'Haese, C.A., Wheeler, W.C., 2004. Is Ellipura monophyletic? A combined analysis of basal hexapod relationships with emphasis on the origin of insects. Org. Divers. Evol. 4, 319–340.

Goldman, N., 1998. Effects of sequence alignment procedures on estimates of phylogeny. Bioessays 20, 287–290.

Groombridge, B., 1992. Global Biodiversity. Chapman and Hall, London.

Gutell, R.R., Cannone, J.J., Konings, D., Gautheret, D., 2000. Predicting U-turns in ribosomal RNA with comparative sequence analysis. J. Mol. Biol. 300, 791–803.

Gutell, R.R., Lee, J.C., Cannone, J.J., 2002. The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struc. Biol. 12, 301–310.

Gutell, R.R., Weiser, B., Woese, C.R., Noller, H.F., 1985. Comparative anatomy of 16S-like ribosomal RNA. Prog. Nucl. Acid Res. 32, 155–216.

Hennig, W., 1953. Kritische Bemerkungen zum phylogenetischen System der Insekten. Beitr. Ent. 3 Sonderheft: 1-85. [English edition with supplementary notes in 1981: W. Hennig: Insect Phylogeny. J. Wiley & Sons, New York].

Hennig, W., 1969. Die Stammesgeschichte der Insekten. Waldemar Krammer, Frankfurt am Main.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.

Kjer, K.M., Carle, F.L., Litman, J., Ware, J., 2006a. A molecular phylogeny of Hexapoda. Arthropod Syst. Phylogeny 64, 35–44.

Kjer, K.M., Gillespie, J.J., Ober, K.A., 2006b. Structural Homology in Ribosomal RNA, and a Deliberation on POY. Arthropod Syst. Phylogeny 64, 159–164.

Koch, M., 1997. Monophyly and phylogenetic position of the Diplura (Hexapoda). Pedobiologia 41, 9–12.

Kraus, O., 1998. Phylogenetic relationships between higher taxa of tracheate arthropods. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Chapman & Hall, London, pp. 295–303.

Kraus, O., 2001. "Myriapoda" and the ancestry of the Hexapoda. Ann. Soc. Entomol. Fr. NS 37, 105–127.

Kraus, O., Kraus, M., 1994. Phylogenetic system of the Tracheata (Mandibulata): on "Myriapoda": insecta interrelationships, phylogenetic age and primary ecological niches. Verh. Naturwiss. Ver. Hamburg 34, 5–31.

Kristensen, N.P., 1975. The phylogeny of hexapod "orders". A critical review of recent accounts. Zool. Syst. Evolutionsforsch. 13, 1–44.

Kristensen, N.P., 1981. Phylogeny of insect orders. Annu. Rev. Entomol. 26, 135–157.

Kristensen, N.P. 1991. Phylogeny of extent hexapods. In: Naumann, I. D. (Ed.), The Insects of Australia, vol. 1. 2nd ed. CSIRO, Melbourne University Press, Carlton, pp. 125–140.

Kristensen, N.P., 1995. Forty years' insect phylogenetic systematics. Zool. Beitr. NF 36, 83–124.

Kristensen, N.P., 1998. The groundplan and basal diversification of the hexapods. In: Fortey, R.A., Thomas, R.H. (Eds.), Arthropod Relationships. Chapman & Hall, London, pp. 281–293.

Kukalová-Peck, J., 1987. New Carboniferous Diplura, Monura, and Thysanura, the hexapod groundplan, and the role of thoracic side lobes in the origin of wings (Insecta). Can. J. Zool. 65, 2327–2345.

Kukalová-Peck, J., 1991. Fossil history and the evolution of hexapod structures. In: Naumann, I.D. (Ed.), The Insects of Australia. CSIRO, Australia, Melbourne, pp. 141–179.

Lindgren, A.R., Daly, M., 2007. The impact of length-variable data and alignment criterion on the phylogeny of Decapodiformes (Mollusca: Cephalopoda). Cladistics 23, 1–13.

Luan, Y., Mallatt, J.M., Xie, R., Yang, Y., Yin, W., 2005. The phylogenetic positions of three basal-hexapod groups (Protura, Diplura, and Collembola) based on ribosomal RNA gene sequences. Mol. Biol. Evol. 22, 1579–1592.

Lyal, C.H.C., 1985. Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera). Syst. Entomol. 10, 145–165.

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

Morrison, D.A., Ellis, J.T., 1997. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14, 428–441.

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003a. Hexapoda origins: monophyletic or paraphyletic? Science 299, 1887–1889.

Nardi, F., Spinsanti, G., Boore, J.L., Carapelli, A., Dallai, R., Frati, F., 2003b. Response to comment on "Hexapod origins: monophyletic or paraphyletic?". Science 301, 1482e.

Neefs, J.-M., Van de Peer, Y., De Rijk, P., Shapelle, S., De Wachter, R., 1993. Compilation of small ribosomal subunit RNA structures. Nucl. Acids Res. 21, 3025–3049.

Ogden, T.H., Whiting, M.F., Wheeler, W.C., 2005. Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer. Cladistics 21, 259–302.

Ouvrard, D., Campbell, B.C., Bourgoin, T., Chan, K.L., 2000. 18S rDNA secondary structure and phylogenetic position of Peloridiidae (Insecta, Hemiptera). Mol. Phylogenet. Evol. 16, 403–417.

Pashley, D.P., McPheron, B.A., Zimmer, E.A., 1993. Systematics of holometabolous insect orders based on 18S ribosomal RNA. Mol. Phylogenet. Evol. 2, 132–142.

Rokas, A., Kathirithamby, J., Holland, P.W.H., 1999. Intron insertion as a phylogenetic character: the *engrailed* homeobox of Strepsiptera does not indicate affinity with Diptera. Insect Mol. Biol. 8, 527–530.

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

Stys, P., Bilinski, S., 1990. Ovariole types and the phylogeny of hexapods. Biol. Rev. Cambridge Philos. Soc. 65, 401–429.

Stys, P., Zrzavy, J., 1994. Phylogeny and classification of extant Arthropoda: review of hypotheses and nomenclature. Eur. J. Entomol. 91, 257–275.

Stys, P., Zrzavy, J., Weyda, F., 1993. Phylogeny of the Hexapoda and ovarian metamerism. Biol. Rev. 68, 365–379.

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

Terry, M.D., Whiting, M.F., 2005. Mantophasmatodea and phylogeny of the lower neopterous insects. Cladistics 21, 240–257.

Wheeler, W. C., and D. Gladstein. 2003. POY: Program and documentation available from: <http://research.amnh.org/scicomp/projects/poy.php>.

Wheeler, W.C., 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12, 1–9.

Wheeler, W.C., Whiting, M., Wheeler, Q.D., Carpenter, J.M., 2001. The phylogeny of the extent Hexapoda orders. Cladistics 17, 113–169.

Whiting, M.F., Carpenter, J.C., Wheeler, Q.D., Wheeler, W.C., 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46, 1–68.

Wilson, D.N., Nierhaus, K.H., 2003. The ribosome through the looking glass. Angew. Chem. Int. Ed. Engl. 42, 3464–3486.

Xie, Q., Bu, W.-J., Zheng, L.-Y., 2005. The Bayesian phylogenetic analysis of the 18S rRNA sequences from the main lineages of Trichophora (Insecta: Heteroptera: Pentatomomorpha). Mol. Phylogenet. Evol. 34, 448–451.

Yoshizawa, K., Johnson, K.P., 2005. Aligned 18S for Zoraptera (Insecta): Phylogenetic position and molecular evolution. Mol. Phylogenet. Evol. 37, 572–580.