Secondary structure and phylogenetic utility of the ribosomal large subunit (28S) in monogeneans of the genus *Thaparocleidus* and *Bifurcohaptor* (Monogenea: Dactylogyridae)

Anshu Chaudhary · Hridaya Shanker Singh

Received: 3 May 2011 / Accepted: 9 June 2012 / Published online: 3 July 2012
© Indian Society for Parasitology 2012

Abstract Present communication deals with secondary structure of 28S rDNA of two already known species of monogeneans viz., *Bifurcohaptor indicus* and *Thaparocleidus parvulus* parasitizing gill filaments of a freshwater fish, *Mystus vittatus* for phylogenetic inference. Secondary structure data are best used as accessory taxonomic characters as their phylogenetic resolving power and confidence in validity. Secondary structure of the 28S rDNA transcript could provide information for identifying homologous nucleotide characters, useful for cladistic inference of relationships. Such structure data could be used as taxonomic character. The study supports that species-level sequence variability renders 28S sequence as a unique window for examining the behavior of fast evolving, non-coding DNA sequences. Apart from this it also confirms that molecular similarity present in various species could be host-induced.

Keywords *Bifurcohaptor indicus* · *Thaparocleidus parvulus* · 28S · Secondary structure predictions · Sfold · Maximum parsimony · Phylogenetic analysis

Introduction

The novel approach of molecular morphometrics that relies both on traditional morphological comparison and on molecular sequence comparison by measuring the structural parameters of the 28S secondary structure homologies

(geometrical features, bond energies, base composition etc.) is recently being used to study the phylogenetic relationships of various species (Billoud et al. 2000). It is well known that rRNA is highly conserved throughout evolution; therefore, the secondary-structure elements of the RNA molecule, like the helices, loops, bulges, and separating single-stranded portions, can be considered phylogenetic characters (Zwieb et al. 1981; Schultz et al. 2005; Grajales et al. 2007). RNA secondary structures are particularly useful in systematics because they include characteristics, not found in the primary sequence, that give ‘morphological’ information. It is an established fact that rRNA structure is highly conserved throughout evolution, as most of the folding is functionally important despite primary sequence divergence (Caetano-Anolles 2002). At the same time, the 28S offers sequence motifs that are useful for the development of monogenean DNA barcoding. This is a technique that uses short DNA sequences from a standardized region of the genome as a diagnostic “biomarker” for species and it is becoming must in species identification (Ebach and Holdrege 2005; Moritz and Cicero 2005; Savolainen et al. 2005; Summerbell et al. 2005).

Phylogenetic relationships among groups and species have been studied intensively using the 28S rRNA sequences. However, none of these studies has considered structural components of the 28S rRNA molecule. Since the structural 28S database contains secondary structures, re-analysis of the phylogeny of monogenean of the genus *Thaparocleidus* and *Bifurcohaptor* using secondary structure information seems useful. The objectives of the current study is to validate monogenean parasites, *Bifurcohaptor indicus* (Jain 1958) and *Thaparocleidus parvulus* (Gussev 1976) Lim (1996) from gill filaments of *Mystus vittatus* and to construct phylogeny using rRNA secondary structures supplementing the primary sequence analysis.

A. Chaudhary · H. S. Singh (✉)
Molecular Taxonomy Laboratory, Department of Zoology,
Ch.C.S. University, Meerut 250004, UP, India
e-mail: hridaya_singh@rediffmail.com

Methodology

Parasite material and DNA isolation

Monogeneans were collected in live form from gills of freshly killed fish, *M. vittatus* (Bloch, 1794) at Hastinapur (29°01'N and 77°45'E), (Meerut), U.P., India. Holotype and paratype slides have been deposited in the museum of Department of Zoology (Voucher No HS/Monogenea/2009/02 & HS/Monogenea/2009/05), Ch.C.S. University, Meerut, U.P., India. For the purpose of DNA extraction, parasites were processed singly. DNA was then extracted from individual worms using Qiagen DNeasy Tissue Kit as per the manufacturer's instructions.

DNA amplification, sequencing and analysis

rDNA region spanning the 28S region was amplified from the genomic DNA obtained from the *B. indicus* (Jain 1958) and *T. parvulus* (Gussev 1976) Lim (1996) monogeneans with the help of PCR using the universal primers, forward (5'-AC-CCGCTGAATTTAAGCAT-3') in combination with reverse primer (5'-CTCTTCAGAGTACTTTTCAAC-3'). PCR amplification was performed following conditions: after an initial denaturation at 94 °C for 3 min, 35 cycles of 94 °C for 30 s (denaturation), 56 °C for 45 s (annealing), 72 °C for 10 min. Resultant PCR products were separated by electrophoresis through 1.5 % agarose gels in TBE buffer, stained with ethidium bromide, transilluminated under ultraviolet light. PCR products were purified using Chromous PCR clean up Kit (#PCR 10) and sequenced in both directions using PCR primers on an automated sequencer by DNA sequencing services of Chromous Biotech Ltd. DNA sequences were put to further analysis by using various Bioinformatics tools including similarity search BLAST (URL <http://www.ncbi.nlm.nih.gov/blast>) for query DNA sequence.

Molecular phylogenetic analysis

Initially, the sequences were aligned using ClustalW multiple alignment (Thompson et al. 1994) with the default gap and extension penalties used by this program. Sequence alignment was analyzed in the MEGA 5 (Tamura et al. 2011) for construction of the phylogenetic trees using Neighbour joining, Maximum Parsimony and UPGMA methods. While reconstructing the NJ tree, correcting for transition bias, the Kimura-2-parameter model was used to estimate the distances. The trees from NJ, MP, UPGMA were drawn to scale, branch lengths calculated using the average pathway method and shown in the units of the number of changes over the whole sequence. Subsequently most-parsimonious tree was obtained using the close-neighbor-interchange algorithm. The robustness of the inferred phylogeny was assessed using a bootstrap

procedure with 1,000 replications. Both the sequences were submitted in the gene bank under the accession numbers GU830881 and GU014844.

Motif identification

28S sequence motifs were identified from aligned sequences of *B. indicus* and *T. parvulus* data set using MEME software (Bailey and Elkan 1994). In order to further evaluate the potential use of the 28S sequence motifs for species identification of *B. indicus* and *T. parvulus*, we directly tested the 28S motifs, using the BLAST algorithm against the non-redundant GenBank database of the National Center for Biotechnology Information (nr at NCBI). The BLAST analysis investigated motifs that showed either conserved or slightly degenerated sites obtained from the two monogenean species. The outputs were showing significantly high scores in terms of percent identity and low expect (*E*) values for both species.

Prediction of secondary structures

Secondary structures of 28S sequences of *B. indicus* and *T. parvulus* species were reconstructed by aligning their sequences. For structural modeling, reconstruction of secondary-structure models, thermodynamic criteria and covariation analyses have been applied. RNA secondary structure was determined using Sfold, based on statistical sample of Boltzmann ensemble for secondary structures (Ding and Lawrence 2003). Subsequently, inferred structure was examined for stems, loops and bulges. RNA was folded at a fixed temperature of 37 °C, and the structure chosen from different output files was the desired one with the highest negative free energy if various similar structures were obtained. Since, GC content is known to influence structural energy therefore; GC percentage was determined using GC calculator (http://www.genomicsplace.com/gc_calc.html). All non-DNA characters were striped before computing. Energy levels of presumptive secondary structure were then calculated with Mfold (Jaeger et al. 1989; Zuker et al. 1999). 28S sequences with secondary-structure format were aligned using the MARNA web server (Siebert and Backofen 2005) based on both the primary and secondary structures. As the default setting, base deletion was scored 2.0, base mismatch 1.0, arc removing 2.0, arc breaking 1.5, and mismatch 1.8 with ensemble of shaped structures.

Results

Construction of phylogenetic trees

28S nucleotide sequences of *B. indicus* (Jain 1958) and *T. parvulus* (Gussev 1976) Lim (1996) were compared with

Fig. 1 Phylogenetic trees of 28S sequences of *B. indicus* and *T. parvulus*. **a** Neighbor joining, **b** maximum parsimony, **c** UPGMA

sequences of other monogenean species, available at Genbank. The fragments of amplified DNA were found to be 344 and 329 bp, respectively. 28S sequences in the

BLAST hit results show that the query 28S sequences were most similar to the sequence of *C. proximus* (Gussev 1976), that belonged to the same family. Phylogenetic analysis

Fig. 2 Showing discover motifs. Each site is identified by the name of the sequence where it occurs and the position in the sequence where the site begins. The p value gives the probability of a random

string having the same match score or higher. This is referred to as the position p value. **a** Motif 1 **b** Motif 2, **c** Motif 3

Fig. 3 Combined block diagram of 3 motifs. Non-overlapping sites with a p value better than 0.0001. The height of the motif “block” is proportional to $-\log(p \text{ value})$, truncated at the height for a motif with a p value of $1e^{-10}$. The occurrences of the motif in sequences are shown as coloured blocks on a line. The p value of an occurrence is

the probability of a single random subsequence the length of the motif. When the DNA strand is specified ‘+’, it means the motif appears from left to right on the sequence, and ‘−’ means the motif appears from right to left on the complementary strand. (Color figure online)

using NJ, MP and UPGMA methods showed that the topology is similar among the trees obtained (Fig. 1a, b, c). Evolutionary divergence between sequences was also estimated and the number of base substitutions per site from analysis between sequences is shown in Fig. 3.

Identification of 28S motifs

Nucleotide sequences of *B. indicus* and *T. parvulus* were tested for motifs. In all 10 motifs were discovered by the software from the aligned sequences (Fig. 2a, b, c). Total 3 best 28S motifs were tested by BLAST analysis against the generalized GenBank database at NCBI (Fig. 3). Out of the 3 motifs, motif 1 (Table 1) exhibited exact or perfect matches with that of the monogenean species from different geographical isolates with 100 % identity and significant *E* value scores of $1e^{-19}$.

Secondary structure analysis

G+C contents for the regions of rDNA (28S) of *B. indicus* and *T. parvulus* is found to be 46.5 and 49.2 %, respectively. Probability profiling approach by Ding and Lawrence (2001) reveals target sites that are commonly accessible for a large number of statistically representative structures in the target RNA. This novel approach bypasses the long-standing difficulty in accessibility evaluation due to limited representation of probable structures due to high statistical confidence in predictions. Minimum free energy (MFE) is estimated by summing individual energy contributions from base pair stacking, hairpins, bulges, internal loops and multi-branch loops. RNA secondary structure consists of stems and loops. From the data we were able to draw the rRNA secondary structure where each residue is identified by a base pair, the backbone and the hydrogen

Table 1 BLAST output of best 28S sequence motif against NCBI data base

Motif 1 pattern (5'-3' end)	Length (bp)	Number of best hits	Identity (%)	<i>E</i> value
CGGATTGCTTGAGAATGCAGTCC AAAGTGGGTGGTAACTCCATCCAAGG	50	100	100	$1e^{-19}$

Fig. 4 Predicted 28S RNA secondary structure and their structure formation enthalpies. **a** *B. indicus*, **b** *T. parvulus*

Fig. 5 Distribution of different loops in 28S region in two species *B. indicus* and *T. parvulus*

bonds are represented as dots between the base pair (Fig. 4a, b). Mainly five types of loops are present in RNA secondary structure viz., interior, hairpin, exterior, multi and bulge (Fig. 5). A single-stranded region, predicted by both the MFE structure and the ss-count has low probabilities on the probability profile ($W = 4$ bases, as described by Ding and Lawrence (2001)). The ss-count statistics gives the propensity of a base to be unpaired, as measured by the frequency with which it is unpaired in a group of the optimal and suboptimal foldings within a specified increment of the MFE. At nucleotide position i , the probability that nucleotide i , $i + 1$, $i + 2$, $i + 3$ (i.e. fragment width $W = 4$) are all single stranded is plotted against I which gave a 3D energy landscape plot of the sampled ensemble and representative structures.

3D energy landscape plot for *B. indicus* is shown (Fig. 6a). The optimal number of clusters determined by software is two. Structures belonging to the clusters are marked as solid dots of two different colors i.e. light blue and purple. The MFE structure and the ensemble centroid are both in the largest cluster (light blue color), with a probability of 0.506. The coordinates for a structure are (MDS axis 1, MDS axis 2, free energy), where the horizontal axis are from multi-dimensional scaling (MDS) and the vertical axis is the free energy of a secondary structure. The coordinates are for the MFE structure (−52.77, −4.97, −109.80), for ensemble centroid (3.96, 0.99, −67.80), for the centroid of cluster 1 (29.24, −2.05, −87.20) and for centroid of cluster 2 (−36.99, −6.25, −94.40).

3D energy landscape plot of *T. parvulus* is shown (Fig. 6b). In this structure, the optimal number of clusters determined by software is 2. Structures belonging to the clusters are marked as solid dots of two different colors. The MFE structure and the ensemble centroid are both in the largest cluster (light blue color), with a probability of 0.875. The coordinates for a structure are (MDS axis 1,

Fig. 6 3D energy landscape plot. **a** *B. indicus*, **b** *T. parvulus*

MDS axis 2, free energy), where the horizontal axis are from MDS and the vertical axis is the free energy of a secondary structure. The coordinates are for the MFE structure (−8.16, −2.74, −116.60), for ensemble centroid (−4.23, −1.91, −110.80), for the centroid of cluster 1 (−4.23, −1.91, −110.80) and for centroid of cluster 2 (58.41, −3.13, −104.00).

Clusters are sorted in descending order of cluster size. Cluster containing the MFE structure is marked using a red asterisk. Size of a cluster is the sampling estimate for the probability of the cluster i.e. the sample frequency of the cluster. MFE structure is excluded from the calculation of cluster sizes, and the sizes of all clusters sum to one.

Multi-dimensional scaling plot of the samples ensemble and representative structures MDS is a technique for representing high-dimensional objects in typically 2D (Kruskal and Wish 1977). Base pair distances are used as an input to MDS. Members of all other clusters are plotted as small circles.

In the present work, we applied a more objective approach for reconstruction of optimal alignment using

secondary-structure informations of *B. indicus* and *T. parvulus* 28S sequence. Manual construction of alignment is a complex and time consuming procedure, whereas application of the MARNA program generated a reliable alignment and represented a powerful tool (Fig. 7). It showed alignment of *B. indicus* and *T. parvulus* which evaluated both the sequence and structural similarity. The alignment optionally satisfies given constraints and allows unaligned fragments at the end of both sequences without penalty. The alignment is shown together in the form of the predicted structure (Fig. 8). The saturation decreases with the number of incompatible base pairs, thus, it indicates the structural conservation of the base pair.

Discussion

Molecular taxonomists are generally overwhelmed by complexity of smothering sequence information owing to their number and sibling status of species. Dearth of reliable molecular markers has led to an unquenched search to find new ones and utilizing the existing knowledge at different levels to unveil new patterns and phylogenetic inferences. Species identification and evolutionary inference generally use molecular and phylogenetic approaches on the 28S as target genomic loci as also reported by Chisholm et al. (2001); Olson et al. (2003); Lee et al. (2007). Differentiating between two species is not possible on the basis of

Fig. 7 Alignment with color annotation showing conservation of base pairs and annotated column identity between *B. indicus* and *T. parvulus* (Base pairs are colored using the same color code, such that

hue shows sequence conservation and saturation shows structural conservation). (Color figure online)

Fig. 8 2D plot of consensus secondary structure predicted from Fig. 7 (same color scheme has been used). The consensus structure is printed as a string of *dots* and *brackets* on *top* of the alignment. The string is well-bracketed, such that base pairs in the structure are indicated by corresponding *opening* and *closing brackets*. Furthermore, compatible base pairs are colored, where the hue shows the number of different types C–G, G–C, A–U, U–A, G–U or U–G of compatible base pairs in the corresponding columns. In this way the hue shows sequence conservation of the base pair. (Color figure online)

morphological, pathological or immunological findings only as it often lead to taxonomic confusions. In phylogenetic studies involving secondary structure analysis as a tool, RNA folding is used for refining the alignment. Structural energy seems to be the most important factor influencing the structural stability. The measurable structural parameters of the molecule are directly used as specific characters to construct a phylogenetic tree. These structures are inferred from the sequence of the nucleotides, often using energy minimization (Zuker 1994). The secondary structure with the lowest possible free energy value, the MFE structure is predicted and found to be the most stable secondary structure for the strand. 28S RNA structures from *B. indicus* and *T. parvulus*, the highest negative free energy is -108.50 and -116.00 kcal, respectively. This convergence at secondary

structural level may be due to the evolutionary pressure on 28S to maintain the RNA secondary structure involved in post-transcriptional processing of rRNA as also reported by other workers like Lalev and Nazar (1999), Shinohara et al. (1999). Critical changes in the rRNA folding pattern are brought about by sequence evolution in the 28S region. This have an important influence on the kinetics of precursor rRNA formation and ultimately on the efficient functioning of the rDNA cluster. Therefore, homologous recognizable characters on the secondary structures can be easily traced out, contrary to finding right counterpart for each nucleotide in every other sequence.

Moreover, predicted secondary structures of RNA, constructed from 28S sequences from *B. indicus* and *T. parvulus*, provided additional information for correct identification of the species. The secondary structure analysis of the same data also confirmed the results mentioned for primary sequence analysis as similar findings have also been reported by workers like Bachellerie and Michot (1989); Billoud et al. (2000). Representative structures, including the MFE structure, ensemble centroid and centroids of all colored clusters, are drawn as large dots in the graph. The units on the axes of the MDS plot only serve the purpose of indicating the relative positions of the objects (Chan et al. 2005). Therefore, structural model-based analyses of DNA sequence data have become increasingly important for phylogenetic inference. In the present study, we propose that *B. indicus* and *T. parvulus*, collected from same host have some minor molecular biological similarity that might be due to the following reasons: these two species are belonging to same family, as these two morphological types are exclusively and respectively found on same hosts.

Motif discovery can be viewed as a ‘needle in a haystack’ problem. The motif discovery algorithm is looking for a set of similar short sequences (the needle) in a set of much longer sequences (the haystack). The problem is easier when the motif instances are long and very similar to each other. It gets much harder when the motif instances are short and/or degenerate, or the input sequences are very long. Discovering motifs in a set of DNA sequences is a difficult task owing to the tendency of binding sites to be short and degenerate and owing to the fact that promoter regions are often difficult to identify precisely. The problem tends to be worse in eukaryotes than in prokaryotes and yeast because eukaryotic transcription factor binding sites tend to be shorter and more variable (Tompkins et al. 2005). A motif was considered highly specific to a monogenean species if it matched most or all the 28S sequences of this class but no other 28S of any other class species. Motif 1 is found to be highly conserved and showed 100 % nucleotide similarity with hundred species of monogeneans.

RNA secondary structure is crucial structure but determination of the correct structure and folding pattern of 28S is cumbersome. It is practically unfeasible to calculate the effect of parameters influencing the structural energy of the RNA structure by conventional experimental approaches. With exponential increase in sequences complexities in deriving interpretation and inferences from the accumulated data pose an infinite challenge. Different RNA folding algorithms also take into account the structural energy as the major determinant in furnishing RNA secondary structure models and conformation. Application of the secondary structure model of rRNA to phylogenetic analyses leads to trees with resolved relationships among clades and eliminates some artifactual support for misinterpreted relationships. The highly resolved topology in the phylogenetic tree suggests that a deep phylogenetic signal has been retained in the 28S sequences of existing species. Use of incorporating secondary-structure information, allows improved estimates of phylogeny and relationships among species. Besides phylogenetic and functional implications, the present results also enhance the value of 28S as a paradigm for DNA sequence analysis. The present analysis demonstrates that 28S also exhibits conserved sequence patterns, diagnostic in the form of motif. Moreover, species-level sequence variability renders 28S as a unique window for examining the behavior of fast evolving, non-coding DNA sequence. This molecular similarity could be host-induced.

Conclusion

Molecular study of the 28S sequences is a promising tool for monogenean species identification. RNA secondary secondary analysis could also be a valuable tool for distinguishing species because 28S secondary structure contains more information than the usual primary sequence alignment. During the study an attempt has also been made to find 28S motifs (≤ 50 bp in size) and its application as a tool for species identification. This approach can be further fine-tuned in resolving ambiguities using differences at the RNA structural level for identification of species complexes.

Acknowledgments We are thankful to Head, Department of Zoology, Ch.C.S. University, Meerut, for providing laboratory facilities. Funding for this study was provided by Department of Science and Technology, (grant number SR/SO/A543/2005) to HSS.

References

- Bachellerie JP, Michot B (1989) Evolution of large subunit rRNA structure. The 39 terminal domain contains elements of secondary structure specific to major phylogenetic groups. *Biochimie* 71:701–709
- Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36
- Billoud B, Guerrucci MA, Masselot M, Deutsch JS (2000) Cirripede phylogeny using a novel approach: molecular morphometrics. *Mol Biol Evol* 17:1435–1445
- Caetano-Anolles G (2002) Tracing the evolution of RNA structure in ribosomes. *Nucleic Acids Res* 30:2575–2587
- Chan CY, Lawrence CE, Ding Y (2005) Structure clustering features on the Sfold web server. *Bioinformatics* 21:3926–3928
- Chisholm LA, Morgana JAT, Adlard RD, Whittington ID (2001) Phylogenetic analysis of the Monocotylidae (Monogenea) inferred from 28S rDNA sequences. *Int J Parasitol* 31:1537–1547
- Ding Y, Lawrence CE (2001) Statistical prediction of single stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucleic Acids Res* 29:1034–1046
- Ding Y, Lawrence CE (2003) A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res* 31:7280–7301
- Ebach MC, Holdrege C (2005) DNA barcoding is no substitute for taxonomy. *Nature* 434:697
- Grajales A, Aguilar C, Sanchez J (2007) Phylogenetic reconstruction using secondary structures of internal transcribed spacer 2 (ITS2, rDNA): finding the molecular and morphological gap in Caribbean gorgonian corals. *BMC Evol Biol* 7:90
- Gushev AV (1976) The systematic composition of the Indian fauna, zoogeography and evolution of freshwater monogeneans. *Novaya Seriya* 35:5–32
- Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. *Proc Natl Acad Sci USA* 86:7706–7710
- Jain SL (1958) Monogenea of Indian freshwater fishes VII. *Bifurcator*, a new genus of freshwater Tetraonchinae from the gills filaments of two fishes from Lucknow. *J Parasitol* 44:388–394
- Kruskal JB, Wish M (1977) Multidimensional scaling. Sage, Beverly Hills
- Lalev AI, Nazar RN (1999) Structural equivalence in the transcribed spacers of pre-rRNA transcripts in *Schizosaccharomyces pombe*. *Nucleic Acids Res* 27:3071–3078
- Lee SU, Chun HC, Huh S (2007) Molecular phylogeny of parasitic platyhelminthes based on sequences of partial 28S rDNA and mitochondrial cytochrome c oxidase subunit I. *Korean J Parasitol* 45:181–189
- Lim LHS (1996) *Thaparocleidus* Jain, 1952, the senior synonym of *Silurodiscoides* Gussev, 1976 (Monogenea: Ancylo-discoidinae). *Syst Parasitol* 35:207–215
- Moritz C, Cicero C (2005) DNA barcoding: promise and pitfalls. *PLoS Biol* 2:1529–1531
- Olson PD, Cribb TH, Tkach VV, Bray RA, Littlewood DTJ (2003) Phylogeny and classification of the digenea (Platyhelminthes: Trematoda). *Int J Parasitol* 33:733–755
- Savolainen VR, Cowan S, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopaedia of life: an introduction of DNA barcoding. *Philos Trans R Soc Lond B* 360:1805–1811
- Schultz J, Maisel S, Gerlach D, Mueller T, Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11:361–364
- Shinohara ML, LoBuglio KF, Rogers SO (1999) Comparison of ribosomal DNA ITS regions among geographic isolates of *Cenococcus geophilum*. *Curr Genet* 35:527–535
- Siebert S, Backofen R (2005) MARN: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics* 21:3352–3359

- Summerbell RC, Levesque CA, Seifer KA, Bovers M, Fell JW, Diaz MR, Boekhout T, De Hoog GS, Stalpers J, Crous PW (2005) Microcoding: the second step in DNA barcoding. *Philos Trans R Soc Lond B* 360:1897–1903
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Thompson JD, Higgins DG, Gibson TJ (1994) ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat Biotechnol* 23:137–147
- Zuker M (1994) Prediction of RNA secondary structure by energy minimization. In: Griffin AM, Griffin HG (eds) *Computer analysis of sequence data, Part II*, vol 25, Chap 23. CRC Press Inc., Totowa, pp 267–294
- Zuker M, Mathews DH, Turner DH (1999) Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J, Clark BFC (eds) *RNA biochemistry and biotechnology*. Kluwer, Dordrecht, pp 11–43
- Zwieb C, Glotz C, Brimacombe R (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. *Nucleic Acids Res* 9:3621–3640