International Journal of Zoology Studies
ISSN: 2455-7269
Impact Factor: RJIF 5.14
www.zoologyjournals.com
Volume 2; Issue 5; September 2017; Page No. 234-237

Bioinformatics and RNA secondary structure study of metazoans

Bibhuti Prasad Barik, Biswajyoti Narayana Sahoo
Post Graduate Department of Zoology, Khallikote University, Brahmapur, Odisha, India

Abstract
Metazoa are multicellular eukaryotes including all animals with differentiated tissues. ITS rDNA sequences were used to infer for phylogenetic affiliations among metazoan species. Phylogenetic trees were investigated by different methods and one of most parsimonious trees was selected to infer phylogeny. The tree showed more or less similar species clustered together but did not form distinct clades as per their lifestyles. The result indicated that several species appear to be polyphyletic and several unrelated species appear to share the same clade. Further studies are needed to resolve the deep phylogeny of metazoans and to have a better understanding of the early steps of nervous system evolution. The differentiation was further resolved by generating consensus RNA secondary structures of ITS rDNA gene sequences. Consensus RNA secondary structures analysis of different phyla revealed that species belonging to phylum arthropoda possess highest number of A-U, G-C and U-G pairing. The exceptions may be attributed to their adaptations in extreme environmental conditions. Consensus ITS rDNA gene RNA secondary structures showed that the structural features of all forms were quite distinct and different from each other. This observation supports the phylogenetic inference in which similarly named species clustered together based on their lifestyles. The results of the present study revealed that consideration of ITS rDNA gene phylogeny and its consensus RNA secondary structures could be used as possible a phylogenetic marker among diverse species of the animal kingdom for large scale data analysis.

Keywords: metazoans, bioinformatics, RNA

Introduction
Origin and phylogeny of metazoa has long been disputed and is a puzzle to the evolutionist. In the current study an attempt has been made to study the molecular phylogeny of at least 10 phyla along with consensus RNA secondary structures. The objectives of the present study were to survey ITS rDNA sequences from world wide database, phylogenetic study of different animal groups based ITS rDNA sequences and differentiation of metazoan phyla based on RNA secondary structure study.

Materials and Methods
Sequence Retrieval and Taxon Sampling
ITS rDNA gene sequences (3300) of diverse metazoan species available in Gen Bank (Benson *et al.*, 2015) [1] database were retrieved using entrez key word search and PERL script. The sequences were filter searched and sequences were selected referring to ten phyla. These gene sequences were considered for phylogenetic analysis and their corresponding RNA secondary structures were also predicted.

Multiple Sequence Alignment and Phylogenetic Analysis
Multiple sequence alignment (MSA) was performed using CLUSTALW program with default settings. Phylogenetic trees were generated by different methods by MEGA7 (Kumar *et al.*, 2016) [6]. All characters were equally weighted and unordered. Alignment gaps were treated as missing data (Table 1).

Table 1: Branch length and indices of CI, RI and CI.

Phylum	Sum of branch length					Consistency Index	Retention Index	Composite Index
	ML	NJ	ME	UPGMA	MP			
Bryozoa	-7223.51	0.41	0.41	0.41	890	0.92	0.88	0.81
placozoa	-990.13	0.01	0.01	0.01	9	0.88	0.50	0.44
Porifera	-15586.68	5.71	5.71	5.59	3609	0.50	9.93	0.47
Cnidaria	-26873.03	63.92	63.92	63.61	5643	0.17	0.57	0.10
Ctenophore	-4394.11	0.11	0.11	0.11	107	0.96	0.76	0.73
Arthropoda	-108865.44	17.59	17.59	17.59	26768	0.51	0.82	0.42
Mollusca	-4196.75	0.48	0.48	0.49	541	0.95	0.98	0.93
platyhelminthes	-38368.20	3.02	3.02	2.98	5229	0.57	0.89	0.51
Nematoda	-98935.81	5.22	5.22	5.15	20217	0.33	0.92	0.31
Chordata	-50922.14	0.53	0.53	0.53	8240	0.79	0.91	0.72
All Phyla	-340419.02	-	-	-	85202	0.31	0.83	0.26

RNA Secondary Structure Prediction

The ITS rDNA sequences were used to generate consensus RNA secondary structures using RNAalifold (Bernhart *et al.*, 2008). It performed multiple alignments and generated consensus secondary structures of each category using realistic energy model for RNAs. The MFE structure of an RNA sequence is the secondary structure that contributes a minimum of free energy. This structure is predicted using a loop-based energy model and the dynamic programming algorithm introduced by Zuker *et al.* (Zuker *et al.*, 1981) [10]. As an RNA secondary structure can be uniquely decomposed into loops and external bases the loop-based energy model treats the free energy *F (s)* of an RNA secondary structure *s* as the sum of the contributing free energies F_L of the loops *L* contained in *s*. According to the chosen energy parameter set and a given temperature (defaults to 37 °C) the secondary structure *s* that minimizes *F (s)* is computed.

Results

Phylogenetic Analysis using Maximum Parsimony Method

The evolutionary history was inferred using the Maximum Parsimony method (Figure 1). The consistency index is 0.31, the retention index is 0.83, and the composite index is 0.26 for all sites and parsimony-informative sites. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 1 in which the initial trees were obtained by the random addition of sequences 10 replicates. The analysis involved 589 nucleotide sequences. There were a total of 12993 positions in the final dataset.

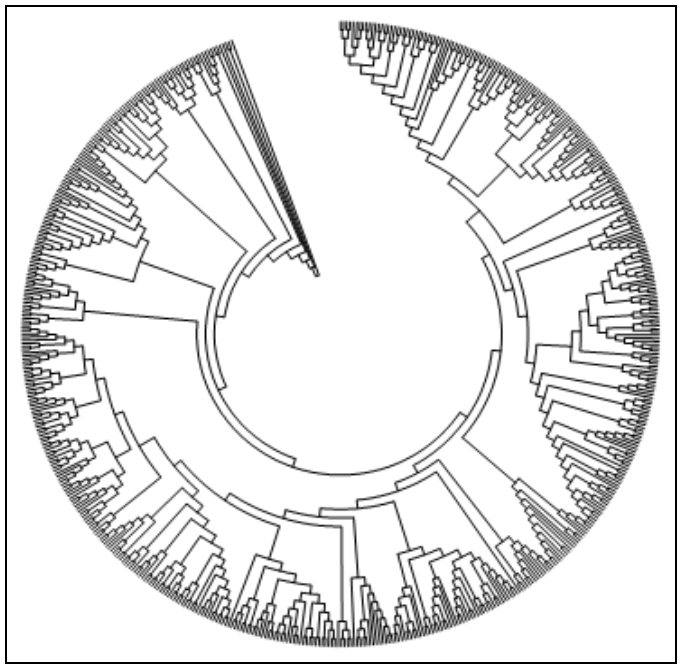

Fig 1: Circular Maximum Parsimony Tree.

The phylogenetic tree generated by maximum parsimony method resulted into 30 major clades, 31 minor clades and 11 independent lineages. Out of the major clades, clade no. 25, 29 and 30 covered species of Porifera only. Clade 25 (*L.clathria and P.heteroraphis*), clade 29 (*P.laughlini, Agelas sp. and A.willeyana*) and clade 30 (*Pione sp. and C.celata*). Clade no. 26 and 27 covered species of Cnidaria only. Clade 26 (*D.putteri, D.suensoni, Dendronephthya sp., D.gigantean, D.castanea, D.spinifera and D.aurea*) and clade 27 (*P.heliodiscus, P.mutuki and P.tuberculosa*). 5 clades (clade 1 (*C. abominator, C. peccator, C.iolambdis* and *C.mulrennani*), clade 2 (*A.arabiensis, A.annulipes, A.clowi and A.farauti*), clade 5(*C.tesquorum, P.parasiluri, S.undulatus, S.polycolpus, S.major*), clade 9 (*S.suzukii, S.rufibasis, S.sakishimaense* and *S.quinquestriatum*), clade 10 (*S.intermedium, S.ornatum and S. trifasciatum*) covered exclusively species of Arthropods. Clade 3 covered species of Mollusca only (*C.nippona, C.ariakensis and C.gigas*). Only clade 4 covered species of platyhelminthes (*G.arcuatus, G.gondae,G.branchicus, G.turnbulli, G.micropsi, G.rugiensis, G.gurleyi, G.kobayashii, G.longoacuminatus, G.derjavini, G.pungitii, G.salaries and G.gasterostei*). Clade 12 to 24 covered species of nematodes only. Clade 12 (*C.cotti,C. hypopthalmichthys and P.skrjabini*), clade 13 (*S.bicornutum, S.ceratophorum, S.riobrave, S.pakistanense, S.abbasi, S.scapterisci, S.carpocapsae, S.steinernema, S.karii, S.longicaudum, S.diaprepesi, S.arenarium, S.cubanum, S.glaseri, S.neocurtillae, S.monticolum, S.oregonense and S.kraussei*), clade 14 (*N.americanus and Necatorsp.*), clade 15 (*B.hellenicus, B.hylobianum,B.gerberi, B.paracorneolus, B.pinasteri, B.platzeri, B.clavicauda, B.africanus, B.doui, B.luxuriosae and B.mucronatus*), clade 16 (*H.glycines, H.schachtii, H.elachista, H.latipons, H.filipjevi and H.avenae*), clade 17 (*Pratylenchus sp. and P.hippeastri*), clade 18 (*P.coffeae and P.loosi*), clade 19 (*P.pinguicaudatus, M.javanica, D.destructor and T.semipenetrans*), clade 20(*P.convallariae, P.penetrans and P.brachyurus*), clade 21 (*P.neglectus, P.fallax and P.vulnus*), clade 22 (P.mediterraneus, P.thornei and P.crenatus), clade 23 (*N.aberrans, Nacobbus sp. and N.bolivianus*) and clade 24 (*R.similis, P.zeae and P.bhatti*). 2 clades (clade 7 (*T.putitora, T.mosal, P.chelynoides and T.progeneius*), clade 8 (*N.hexagonolepis, T.mussullah and T. khudree*) covered exclusively species of Chordata.

Clade no. 6 is covered jointly by the species of Bryozoa (*P.geimermassardi, P.repens, P.vaihiriae, H.punctata, P.casmiana, C.mucedo and F.sultana*), Arthropoda (*S.pallens, P.couloniana, Archiblatta sp., M.maculate, A.tessellata, C.tentans and B.discoidalis*) and Cnidaria (*T.bryosalmonae*).

Clade no.11 is covered jointly by the species of Arthropoda (*O.cynotis*), Placozoa (*T.adhaerens*) and Platyhelminthes (*D.huronense, D.indistinctum, D.baeri, C.funduli, H.taichui, Centrocestus sp., H.pumilo and C.sinensis*).

Clade no. 28 is covered jointly by the species of Cnidaria (*Cupressopathes sp., Myriopathes sp., C.abies, C.pumila, M.myriophylla, R.reticulate, Stichopathessp, S.cylindrical, Distichopora sp. and M.exaesa*), Ctenophora (Bolinopsis sp.)And Porifera (*Dysidea sp., S.schultzei and Vaceletiasp*).

Out of the minor clades 4 no. of species belonging to phylum porifera got their place in 4 different no. of minor clades in the current maximum parsimony tree (*A.vastus, S.capricorn, E.fluviatilis, Timea sp*.). In minor clade no. 3, species belonging to phylum Cnidaria is present exclusively (*Aurelia sp*.). 25 no of species belonging to phylum Arthropoda got their places in 12 no. of minor clades in the current maximum parsimony tree (*C.pilosus, C.erraticus, C.atratus, C.cedecei, D.neobrevipes and D.bravipes. L.invasa, Q.erythrinae and*

A.miltoni, A.bahia, Prosimulium sp., T.tibblesi, C.dacotensis and S.mututa, S.annulus, S.euryadminiculum, S.aureum, S.quebecense, S.tuberosum, S.longistylatum and S.pictipes, S.aureohirtum, S.decorum and S.venustum, S. intermedium). Only minor clade 7 is being covered by the species of Mollusca phylum (*O.denselamellosa*). Minor clade 8 is exclusively comprised of phylum Platyhelminthes (*S.acheilognathi*) In 10 different minor clades there are 21 no. of species belonging to the phylum Nematoda in the current maximum parsimony tree (*C.splendens, P.nana, B.germanica, G.pulchrum, D.brevis and D.folliculorum, S.affine and S.intermedium, P.goodeyi, G.pallida, G.rostochiensis, G.artemisiae and C.estonica, Heterodera sp. and H.goettingiana, P.floridensis, P.jaehni, P.agilis and P.scribneri, P.gutierrezi, P.bolivianus*). In two minor clades 10 and 11 it is exclusively covered by the species of Chordata in the current maximum parsimony tree (*T.tor, C.carpio, O.aureus, B.Taurus and B.bubalis*). In the current maximum parsimony tree altogether 11 no. of independent lineages were spotted belonging to two phyla namely Arthropoda and Nematoda.

RNA Secondary Structure Prediction

Further the variations of different lifestyles was investigated in RNA secondary structures level since RNA molecules perform biological functions that fold into specific secondary and tertiary structures that are more conserved than sequences. The compensatory mutations in organisms can be studied in their RNA secondary structures and recently evolutionary models that address such structural variations have been proposed. The result of the present study revealed that consensus RNA secondary structures of all phyla are different from each other.The RNA secondary structures were predicted using a loop-based energy model and the dynamic programming algorithm. The minimum free energy observed for 10 different phyla were: Bryozoa:1096.57, Placozoa:262.80, Porifera:172.34, Cnidaria:357.54, Ctenophora:368.00, Arthropoda:457.18, Mollusca:58.73, Platyhelminthes:717.16, Nematoda:320.64 and Chordata:598.80 kcal/mol. respectively.

Discussion and Conclusion

Morphological disparity among the key lineages of metazoans has encouraged numerous conflicting phylogenetic hypotheses. Unfortunately, because of varying interpretations of features as derived or plesiomorphic, a lack of clear synapomorphies, and often unclear character homology, the ability of morphology to resolve such deep phylogenetic events is limited. Morphological and traditional molecular phylogenetic approaches have failed to robustly reconstruct mollusc phylogeny. Notably, several recent phylogenomic studies have significantly advanced our understanding of metazoan evolution by using sequences derived from genome and transcriptome Morphological disparity among the key lineages of metazoans has encouraged numerous conflicting phylogenetic hypotheses. Unfortunately, because of varying interpretations of features as derived or plesiomorphic, a lack of clear synapomorphies, and often unclear character homology, the ability of morphology to resolve such deep phylogenetic events is limited. Morphological and traditional molecular phylogenetic approaches have failed to robustly reconstruct metazoan phylogeny. Notably, several recent phylogenomic studies have significantly advanced our understanding of metazoan evolution by using sequences derived from genome and transcriptome data. The phylogenetic relationships within the kingdom Animalia (Metazoa) have long been questioned. Focusing on the lowest eukaryotic multicellular organisms, the metazoan phylum Porifera (sponges), it remained unsolved if they evolved multicellularity independently from a separate protist lineage (polyphyly of animals) of derived from the same protist group as the other animal phyla (monophyly). After having analyzed genes typical for multicellularity (adhesion molecules/receptors and a nuclear receptor), Muller (1995) [8] presented evidence that Porifera should be placed in the kingdom Animalia. We therefore suggest a monophyletic origin for all animals. Ctenophores have traditionally been treated as eumetazoans, but some recent whole genome studies have revived the idea that they are, rather, the sister group to all other metazoans. This deep branching position implies either that nervous systems have evolved twice, in Ctenophora and in Eumetazoa, or that an ancestral metazoan nervous system has been lost in sponges and placozoans. However, that phylogenetic-tree construction artifact may have placed ctenophores too deep in the metazoan tree. Although homology of complex biological traits can be assessed without knowing the exact phylogeny, based on statistical principles (well established for molecular sequences, more difficult for complex cellular and organismic data such as gene expression patterns), clarifying the deep branching order of metazoans will undoubtedly help to resolve evolutionary position.

In the current study ITS rDNA gene sequences were retrieved from world wide database. Phylogenetic relationships were inferred using different methods. The clustering was not uniform. Evolutionary diversifications were noticed. The ITS rDNA based phylogeny revealed that species belonging to particular phylum did not cluster as per their lifestyles and several species appear to be polyphyletic while several other unrelated species appear to share the same clade. This may be due to wrongly named species submitted to GenBank as in some cases (Cai *et al.*, 2009) [3]. Moreover, primary sequences often contain insertions and deletions (indels) making alignment difficult beyond infraspecific levels (Kruger and Gargas, 2008). The evolutionary trees showed similar species remain cluster together with few alterations. This may be attributed to adaptive radiation or mutations. For especially for deep nodes, the outcomes of molecular phylogenetics should always be compared with, and eventually validated by, all the expertise in the field, merging to a widely accepted phylogenetic hypothesis, encompassing the whole evidence from protein to morphology. Further improvements of the present work along with proteome structural will increase the available dataset either by exploiting more molecular markers or by further enlarging the sample, with special reference to some underrepresented groups.

One of the interesting finding of the current study was that chordates, molluscs and nematodes are found in separate groups as distinctively in the maximum parsimony tree. Our findings are also supported with the studies carried out by

Meglécz *et al.*, (2012) [7]. The most recent classes or orders of Chordates still retain the pattern of their common ancestor. However, within older groups, such as classes of Arthropods, the phylogenetic pattern has been scrambled by the long independent evolution of the lineages. *We landed in a simple phylogenetic placement of molluscs with disagreement to the findings of Kocot et al., (2011)* [4] *hypothesizing possible multiple origins.*

Further the variations of different lifestyles was investigated in RNA secondary structures level since RNA molecules perform biological functions that fold into specific secondary and tertiary structures that are more conserved than sequences. The compensatory mutations in organisms can be studied in their RNA secondary structures and recently evolutionary models that address such structural variations have been proposed (Srivastava *et al.*, 2011) [9]. *The result of the present study revealed that consensus RNA secondary structures of phyla are different from each other.* Arthropods possess highest number of A-U, G-C and U-G paring. The exceptions may be attributed to their adaptations in extreme environmental conditions. The phylogenetic affinity of certain phylum with those of other may be attributed to the fact that transformation of habit might have taken place due to compensatory base mutations resulting into variations in certain stem-loops and pseudoknots in the RNA structure across evolution but conserving many of the structural elements. While in other phyla there might have been greater degree of mutation followed by insertions and deletions of bases, base pairs and whole stems resulting into variation in the length of corresponding double helical regions and other structural elements. Such variations in RNA secondary structures due to mutations involving indels and substitutions resulting into variation in the lengths of double helical regions of RNA have been reported by Srivastava *et al.* (2011) [9].

References

1. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW *et al.* Gen Bank, Nucleic Acids Res. 43 Database issue: D30-35. Doi: 10.1093/nar/gku1216, 2015.
2. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RN Aalifold: improved consensus structure prediction for RNA alignments. BMC Bioinformatics, 2008, 9:474.
3. Cai L, Hyde KD, Taylor PWJ, Weir B, Waller J, Abang MM, Johnston PR *et al.* A polyphasic approach for studying *Colletotrichum*. Fungal Diversity, 2009; 39:183-204.
4. Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A *et al.* Phylogenomics reveals deep molluscan relationships, *Nature*, 2011; 477:452-456.
5. Kruger D, Gargas A. Secondary structure of ITS2 rRNA provides taxonomic characters for systematic studies-a case in Lycoperdaceae Basidiomycota. Mycological Research. 2008; 112(3):316-330.
6. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. And Evol. 2016; 33(7):1870-1874.
7. Meglécz E, Nève G, Biffin E, Gardner MG. Breakdown of Phylogenetic Signal: A Survey of Microsatellite Densities in 454 Shotgun Sequences from 154 Non Model Eukaryote Species. PLoS ONE. doi:10.1371/journal. pone.0040861. 2012; 7(7):e40861.
8. Muller WE. Molecular phylogeny of Metazoa animals: monophyletic origin. Naturwissenschaften. 1995; 82(7):321-329.
9. Srivastava A, Cai L, Mrazek J, Malmberg RL. Mutational Patterns in RNA Secondary Structure Evolution Examined in Three RNA Families. PLoS ONE, 20484. doi:10.1371/journal. pone. 0020484 PMID:21698102, 2011; 6(6).
10. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information Nucleic Acid Res. 9, 1981.