**International Journal of Biology Research**
**ISSN: 2455-6548**
**Impact Factor: RJIF 5.22**
**www.biologyjournal.in**
**Volume 2; Issue 4; October 2017; Page No. 23-27**

# Sequence based evolutionary relationships among annelids

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

**Abstract**
Annelids are important and diverse group of animals with worldwide distribution. Evolutionary history of this group is still a not resolved. In the current study different coding gene sequences were retrieved and large scale phylogenetic analysis were carried out using multiple methods. The relative study indicates coding genes as phylogenetically informative. The trees showed more or less similar species clustered together but did not form distinct clades as per their lifestyles and morphological similarities. The result also showed that several species appear to be polyphyletic and several unrelated species appear to share the same clade. This may be attributed to adaptive radiation or mutations.

**Keywords:** annelid, species, coding genes, sequences, monophyletic, polyphyletic, phylogeny, clade

## Introduction
Annelids are soft-bodied, bilaterally symmetrical segmented animals distributed throughout the world, from deep ocean bottoms to high mountain glaciers. Despite their success and ecological importance, the evolutionary history of the group isstill a mystery. Numerous species of annelids are knownand it is likely that many still remain unknown and undescribed. Recent phylogenetic analyses have led to profound changes in the view that the Annelida, as traditionally formulated, is a natural, monophyletic group. The objectives of the present study were survey and retrieval of selected coding gene sequences from worldwide databaseand phylogenetic study of different speciesusing multiple methods.

## Materials and Methods
### Retrieval of sequences and taxon sampling
The gene sequences belonging to phylum Annelida were retrieved from NCBI-GenBank database (Benson *et al*., 2013) [1] using entrez key word search and PERL script. The sequences were filter searched and sequences were selected referring to specific genes. The sequences were sorted based on gene types using Bioedit software version 7.0.5.3 (Hall, 1999) [3]. Altogether nine gene categories of genes were retrieved and separated. These gene sequences were considered for further analysis.

### Multiple sequence alignment and Phylogenetic analysis
The retrieved gene sequences were saved and fasta formatted for multiple sequence alignment. The sequences were aligned using CLUSTAL W (Thomption *et al*., 1994). For pair wise sequence alignment the gap opening penalty and extension penalties were 15 and 6.66 respectively. The aligned file was exported for phylogenetic analysis. Five different methods (ML, NJ, ME, UPGMA and MP) were adopted to perform phylogenetic analysis using MEGA 7 software (Kumar *et al*., 2016) [5]. The branch length and consistency, retention and composites indices are shown (Table 1).

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

| S. No | Gene | Sum of Branch Length | | | | | Consistency Index | Retention Index | Composite Index |
|---|---|---|---|---|---|---|---|---|---|
| | | ML | NJ | ME | UPGMA | MP | | | |
| 1 | atp6 (ATP synthase 6) | 0.122 | -995 | 4.801 | 4.949 | 260 | 0.398 | 0.901 | 0.361 |
| 2 | cox1(cytochrome c oxidase subunit1) | -5380 | 2.770 | 2.770 | 2.758 | 1045 | 0.258 | 0.864 | 0.228 |
| 3 | gap (glyceraldehyde 3-phosphate dehydrogenase) | -3725 | 0.619 | 0.6197 | 0.616 | 527 | 0.625 | 0.589 | 0.434 |
| 4 | metallothionein 2 | -1135 | 0.793 | 0.793 | 0.694 | 187 | 0.694 | 0518 | 0.446 |
| 5 | ND2 (NADH dehydrogenase subunit2) | -2511 | 3.462 | 3.462 | 3.469 | 402 | 0.965 | 0.965 | 0.940 |
| 6 | ND4 (NADH dehydrogenase subunit4) | 2090.65 | 2.828 | 2.828 | 2.819 | 533 | 0.936 | 0.979 | 0.929 |
| 7 | RAG1(Recombination Activation Gene1) | -4303 | 0.494 | 0.494 | 0.495 | 580 | 0.592 | 0.468 | 0.434 |
| 8 | RAG2(Recombination Activation Gene2) | -2525 | 0.131 | 0.131 | 0.134 | 152 | 0.655 | 0.631 | 0.544 |
| 9 | aprA (alkaline metalloproteinaseA) | -1929 | 0.912 | 0.912 | 0.921 | 330 | 0.756 | 0.756 | 0.593 |

ML: Maximum Likelihood, NJ: Neighbour Joining, ME: Minimum Evolution, UPGMA: Unweighted Pair Group Method with Arithmetic Mean, MP: Maximum Parsimony, CI: Consistency Index, RI: Retention Index and CI: Composite Index.

All characters were equally weighted and unordered. Alignment gaps were treated as missing data. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap was 500 replicates. The evolutionary distances were computed.

## Results

### Maximum Likelihood Tree

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The trees with the highest log-likelihood are shown (Table 2).

**Table 2:** Gene sequences with highest log-likelihood values in the ML trees

| S. No. | Gene | Highest log-likelihood | S. No | Gene | Highest log-likelihood | S. No | Gene | Highest log-likelihood |
|---|---|---|---|---|---|---|---|---|
| 1 | atp6 | -1101.5774 | 4 | MT2 | -1135.3252 | 7 | RAG1 | -4303.0987 |
| 2 | cox1 | -5380.1513 | 5 | ND2 | -2511.4474 | 8 | RAG2 | -2528.0024 |
| 3 | gap | -3725.2005 | 6 | ND4 | -2094.1472 | 9 | aprA | -1929.9670 |

Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated (Fig. 1).

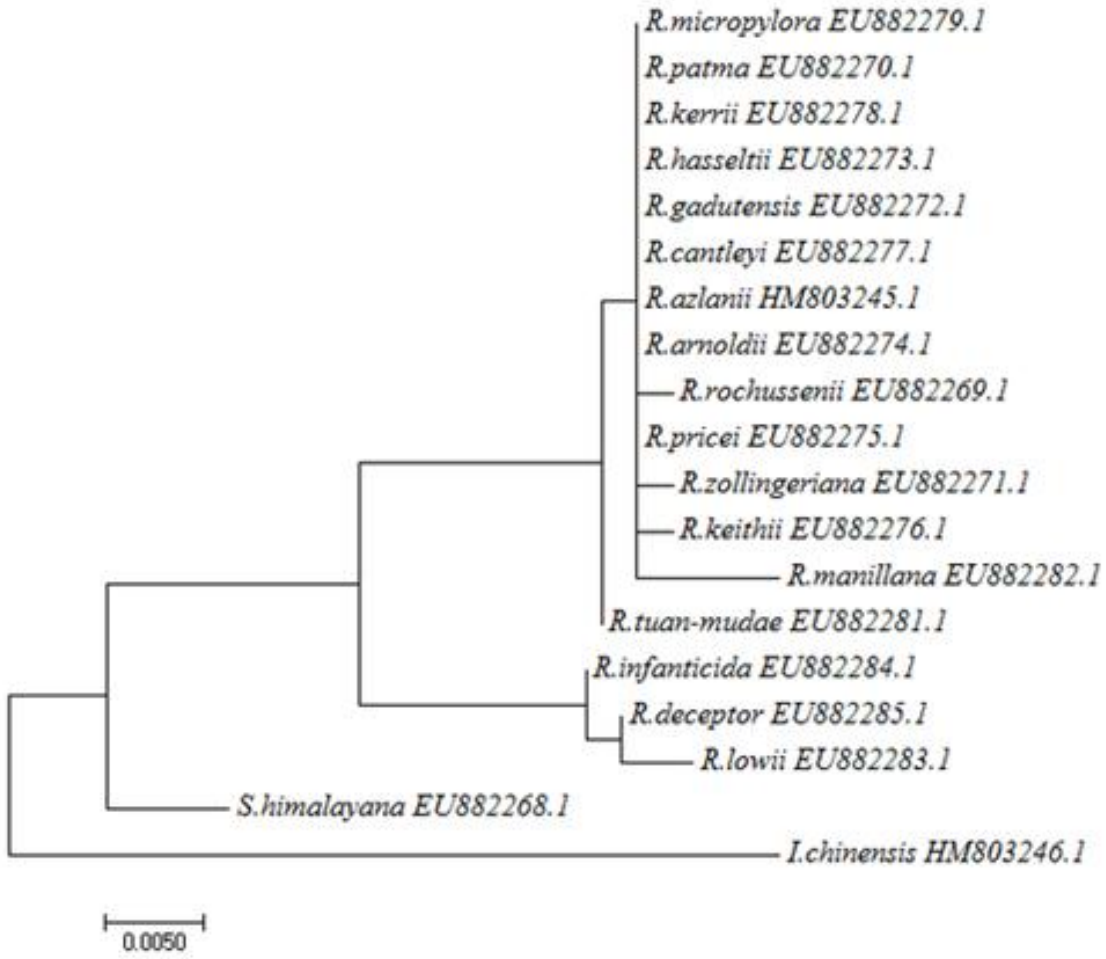


**Fig 1:** ML tree based on atp6 gene

### Neighbor Joining Tree

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987) [6]. The optimal trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al*., 2004) [7] and are in the units of the number of base substitutions per site.

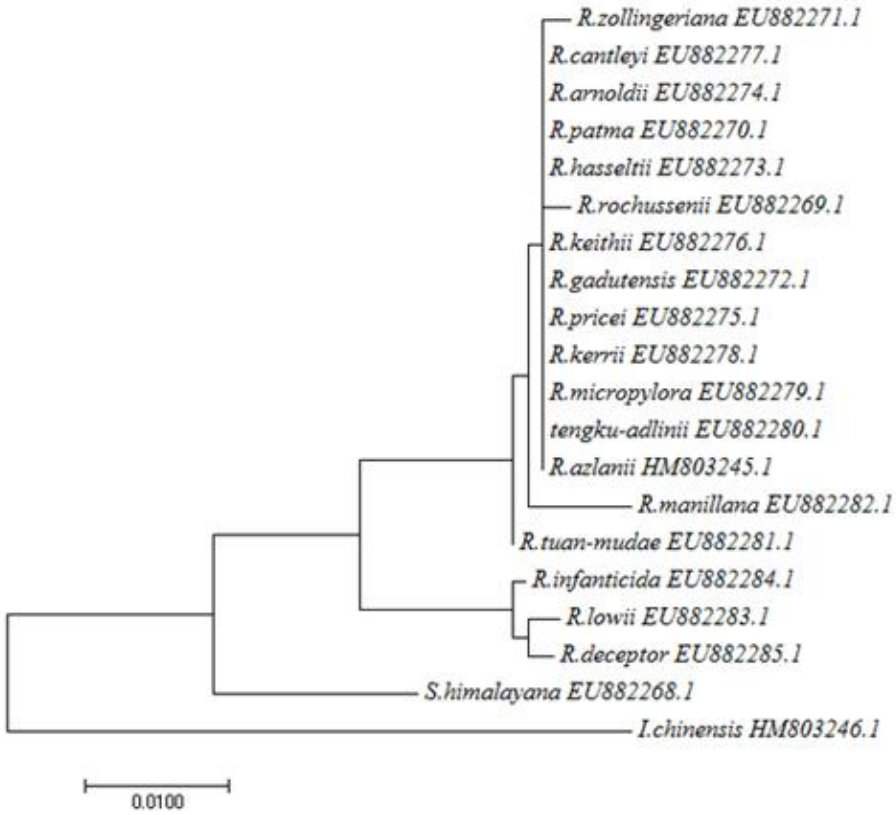


**Fig 2:** NJ tree based on atp6 gene

**Minimum Evolution Tree**

The evolutionary history was inferred using the Minimum Evolution method (Rzhetskyand Nei, 1992). The treesare drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura *et al.*, 2004) [7] and were in the units of the number of base substitutions per site. The ME trees were searched using the Close-Neighbor-Interchange (CNI) algorithm (Nei and Kumar, 2000) [5].

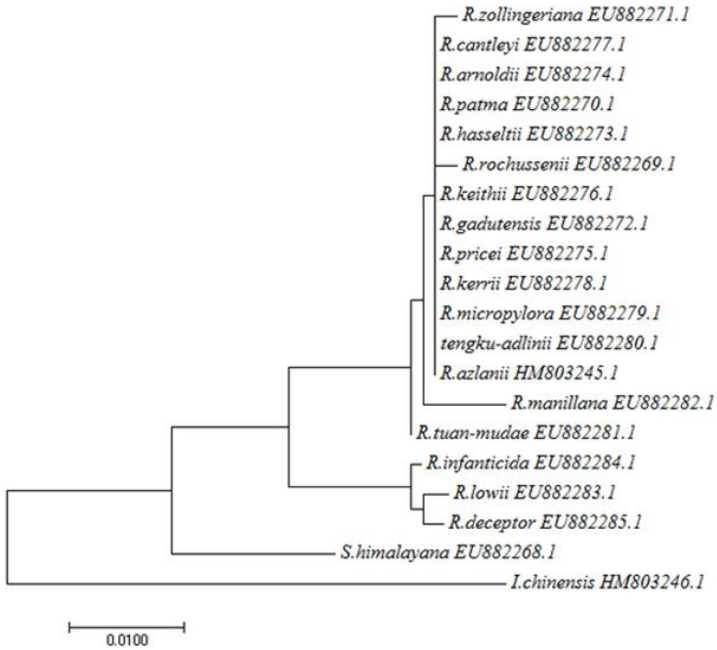


**Fig 3:** ME tree based on atp6 gene

**UPGMA Tree**

The evolutionary history was inferred using the UPGMA method (Sneath and Sokal, 1973) [11]. The optimal trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. The evolutionary distances were computed using the Maximum Composite Likelihood method and were in the units of the number of base substitutions per site.

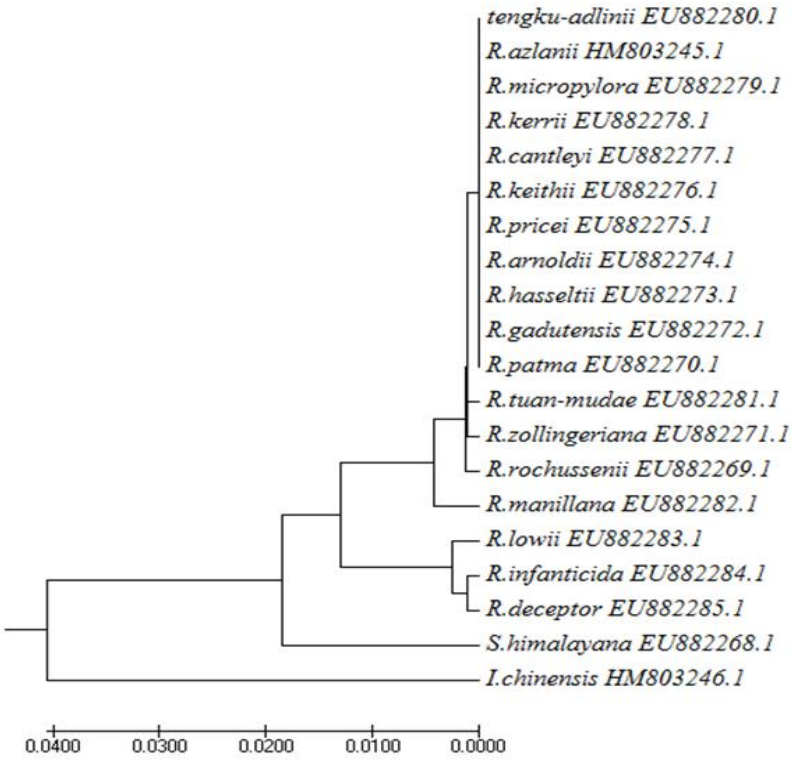


**Fig 4:** UPGMA tree based on atp6 gene

**MP Tree**

The evolutionary history was inferred using the Maximum Parsimony method. The MP trees were obtained using the Subtree-Pruning-Regrafting (SPR) algorithm with search level 0 in which the initial trees were obtained by the random addition of sequences (10 replicates). All positions containing gaps and missing data were eliminated.

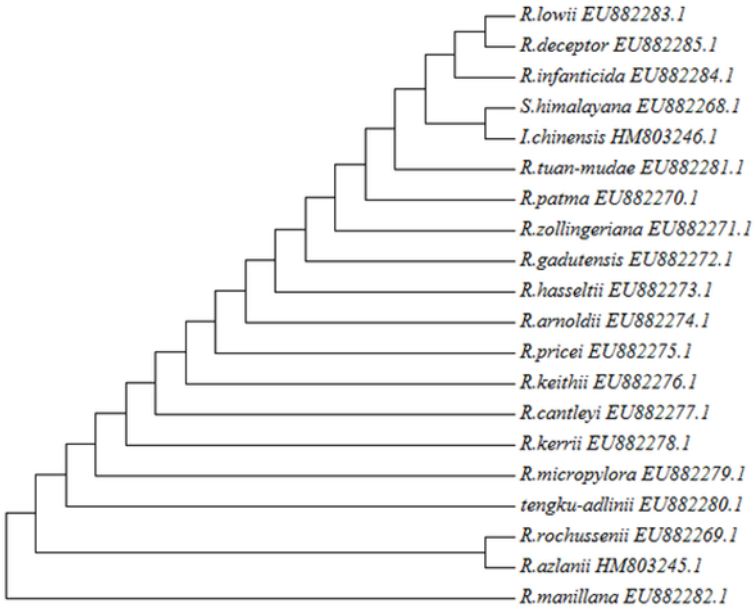


**Fig 5:** MP tree based on atp6 gene

## Conclusion

Identification of unknown species by means of morphology and non-coding genes only may result in unconvinced specimen identifications resulting in to false negatives or positives. In the current study the coding genes were preferred to inspect comprehensive phylogeny in Annelids. The relative study reveals coding genes seem to be phylogenetically informative at the species level. Phylogenetic trees were investigated by different methods to infer evolutionary relationships. The trees showed more or less similar species clustered together but did not form distinct clades as per their lifestyles and morphological similarities. The result also indicated that several species appear to be polyphyletic and several 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) [2]. Moreover, primary sequences often contain insertions and deletions (indels) making alignment difficult beyond intraspecific levels (Kruger and Gargas, 2008) [4]. But still it can be assumed here that phylogenetic analyses using coding gene sequences could be a very productive approach in understanding annelid evolution. Some trees showed similar species remain clustered together with few alterations and this may be assumed by possible adaptive radiation or mutations. With development of new algorithms, efforts should be targeted to increase taxonomic representation in coding genes to analyze the patterns of change through time in morphological, behavioral and developmental characters.

## References


1. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, *et al*. Gen Bank. Nucleic Acids Res. 41(Database issue). 2013, 36-42.
2. Cai L, Hyde KD, Taylor PWJ, Weir B, Waller J, Abang MM, *et al*. A polyphasic approach for studying Colletotrichum. Fungal Diversity. 2009; 39:183-204.
3. Hall TA. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nuc Acids Symp Ser. 1999; 41:95-98.
4. Kruger D, Gargas A. Secondary structure of ITS2 rRNA provides taxonomic characters for systematic studies d a case in Lycoperdaceae (Basidiomycota). Mycological Research. 2008; 112:316-330.
5. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 2016; 33(7):1870-1874.
6. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 1987; 4:406-425.
7. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA). 2004; 101:11030-11035.
8. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucl. Acids Res. ; 1994; 22:4673-4680
9. Rzhetsky A, Nei M. A simple method for estimating and testing minimum evolution trees. Molecular Biology and Evolution. 1992; 9:945-967.
10. Nei M, Kumar S. Molecular Evolution and Phylogenetics. Oxford University Press, New York, 2000.
11. Sneath PHA, Sokal RR. Numerical Taxonomy. Freeman, San Francisco, 1973.