Phylogenetic study of nine species of freshwater monogeneans using secondary structure and motif prediction from India

Anshu Chaudhary* & Hridaya Shanker Singh

Molecular Taxonomy Laboratory, Department of Zoology, University Road, C. C. S. University, Meerut (UP), 250004, India; Anshu Chaudhary - Email: anshu8282@rediffmail.com; *Corresponding author

Abstract:
The present study was performed to identify and validate monogenean species from different piscine hosts using molecular tools. Nine species of freshwater monogeneans were collected from gills and skin of freshwater fishes at Hastinapur, Meerut, India. After microscopic examination, molecular analysis was performed utilizing 28S gene marker. Phylogenetic analysis indicated the validation and systematic position of these nine different monogeneans belongs to the Dactylogyridae and Gyrodactylidae families. The findings also confirm that the 28S rDNA sequence is highly conserved and may prove to be useful in taxonomic studies of parasitic platyhelminthes. Besides this, the study is also supplemented by molecular morphometrics that is based on 28S secondary structure homologies of nine monogenean species. The data indicate that 28S motifs *i.e.*, ≤ 50bp in size can also be considered a promising tool for monogenean species identification and their validation.

Key words: Monogenea; India, 28S rDNA, Secondary structure prediction, Motif, Phylogeny

Background:
One class of platyhelminth, Monogenea is found parasitic mostly on external surfaces and gills of freshwater and marine fishes. In this class, 53 families were recognized in the most recent phylogenetic analysis which is based on morphological characters, but omitted at least ten other 'families' because of uncertainties about origins and/or validity [1]. For many years, despite different aspects of study, the validation and phylogenetic position of the species of this class and their relationship between the sister groups is unresolved. Generally, monogenean identification is based on morphological criteria and morphometric analysis which allows a qualitative and quantitative approach in the analysis of several body parts of monogeneans [2]. The status of many Indian species of monogeneans is considered as *species inquirendae* [3]. Thus, there is need to evaluate the status of many of the Indian species on the basis of morphological as well as molecular features. During the course of this study, we selected large subunit rRNA as taxonomic tool because it is highly conserved across all domains of life [4, 5] and the expansion segments can vary greatly, even across recently diverged lineages [6, 7]. The 28S rDNA is useful for evaluating different levels of taxonomic divergence as they are ideal phylogenetic markers. Moreover, in the case of platyhelminth systematics, rDNA have been successfully used by workers in general and 28S rDNA in particular in estimation of the relationships among the Platyhelminthes [8]. Besides this, it has been known for at least two decade that reference to secondary structure can improve the assignment of positional homology in length heterogeneous data sets [9, 10] and structure-based alignments have also been shown to increase phylogenetic accuracy over automated approaches [11, 12]. As a supplement to the phylogenetic analyses, RNA secondary structure prediction and sequence motifs are as important as the sequence for the function as well

as in the functional RNA. Interactions in the base pairing in an RNA molecule are more important to the overall structure in RNA than any other interaction. Secondary structure study finds out the highly conserved elements in the structure of monogeneans that is found to be common in all species along with consensus structure prediction. The 28S rDNA segments consist of one or a series of putative helical and nonpairing regions that are useful for assessing different levels of taxonomic divergence as they are ideal phylogenetic markers [7]. Apart from this, a predictive approach for the identification of motifs that are conserved between different species was undertaken. The 28S region offers short sequence motifs that are also useful for monogenean species identification. These short DNA sequences are taken from a standardized region of the genome of all studied monogenean species and used as a diagnostic marker for species identification. Thus, with the goal to study the phylogenetic status of different species of monogeneans, the present investigation was made. During study, focus of our investigation was on monogenean species of families Dactylogyridae [13] which is the most diverse family and Gyrodactylidae that is known to be affected by abiotic factors in the macroenvironment [14]. The purpose of this study is to examine the taxonomic status, phylogenetic relationships and secondary structure prediction of 28 rDNA sequences from nine different species of monogeneans that represents nine genera and two families.
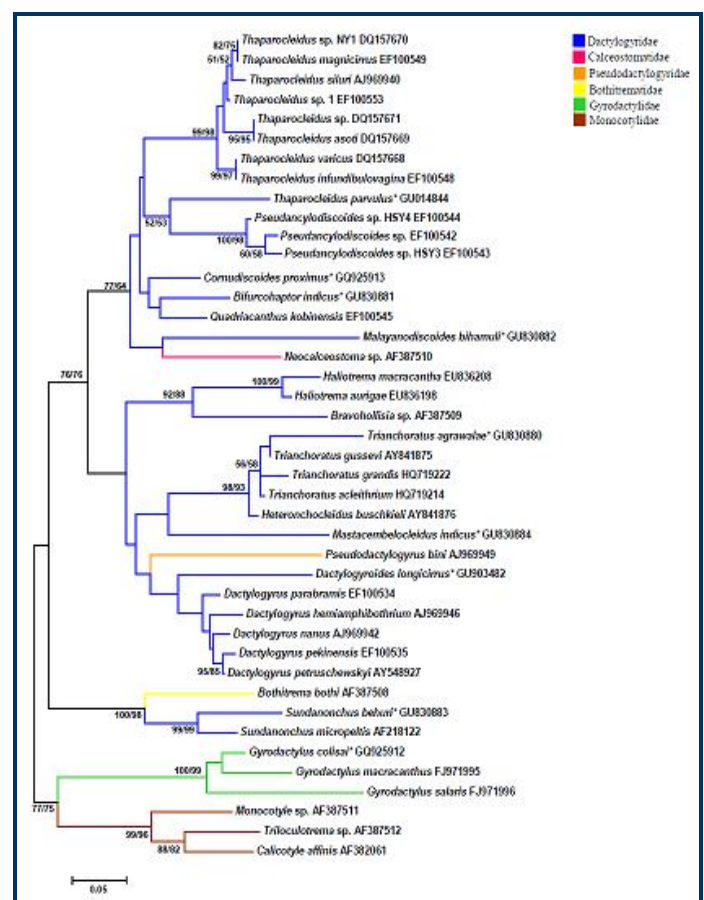
Methodology:
Sample Collection
Monogeneans were collected from infected freshwater piscine hosts from Hastinapur Meerut, U.P., India (29^{0}01'N and 77^{0}45'E). Parasites were isolated from the gill filaments and skin of the host as per method suggested by Malmberg [15]. Monogeneans were examined under a dissecting stereoscopic microscope. Morphological study of the monogeneans was made as suggested by Malmberg [15]. Collected monogeneans were cleaned with water, transferred to microcentrifuge tubes in 95% ethanol for DNA analysis and stored at -20°C until further use. List of the monogenean species used in this study, with their host species, site, voucher details and GenBank accession numbers are given in the **Table 1 (see supplementary material)**. Mounted voucher specimens of each sequenced monogenean species were deposited in the Museum of Department of Zoology, Ch. C.S. University, Meerut (U.P.), India.

DNA and phylogenetic analysis
DNA was extracted from individual parasites by using a Qiagen DNeasy Tissue Kit (Qiagen, Germany) as per the manufacturer's instructions. 28S rDNA was amplified by PCR using forward (5'-ACCCGCTGAATTTAAGCAT-3') and reverse primers (5'-CTCTTCAGAGTACTTTTCAAC-3') [16]. PCR amplification was performed using the following protocol in a final volume of 25 µl PCR reaction. Each amplification reaction contained 10X PCR buffer, 0.4 mM dNTP, 1 U Taq polymerase (Biotools, Spain) and 10 pM of each primer. PCR was carried out with the following steps: an initial denaturation at 94°C for 3 min, 35 cycles of 94°C for 30 s, 56°C for 45 s, and 72°C for 1 min, and a final extension at 72°C for 10 min. PCR products were separated by electrophoresis through 1.5% agarose gels in TBE (Tris–borate–EDTA) buffer, stained with ethidium bromide, transilluminated under ultraviolet light.

PCR products were purified using Chromous PCR clean up Kit (#PCR 10, Chromous Biotech) and sequenced in both directions using PCR primers on an automated sequencer using a Big Dye Terminator vr.3.1 cycle sequencing kit in an ABI 3130 Genetic Analyzer (Applied Biosystems). In addition to the sequences generated from this study, all nucleotide sequences were initially aligned by ClustalW and then manually adjusted. Phylogenetic analyses based on the maximum parsimony (MP), neighbour- joining (NJ) and maximum-likelihood (ML) algorithms. The phylogenetic analysis was performed using a distance method with the MEGA 5 [17]. The distance matrix and the NJ tree were based on the Kimura's 2-Parameter (K2P) model and gaps were treated as missing data. Subsequently, the most-parsimonious tree was obtained using the Close-Neighbor-Interchange algorithm and branch robustness was estimated through bootstrap (BP) analyses of 1000.
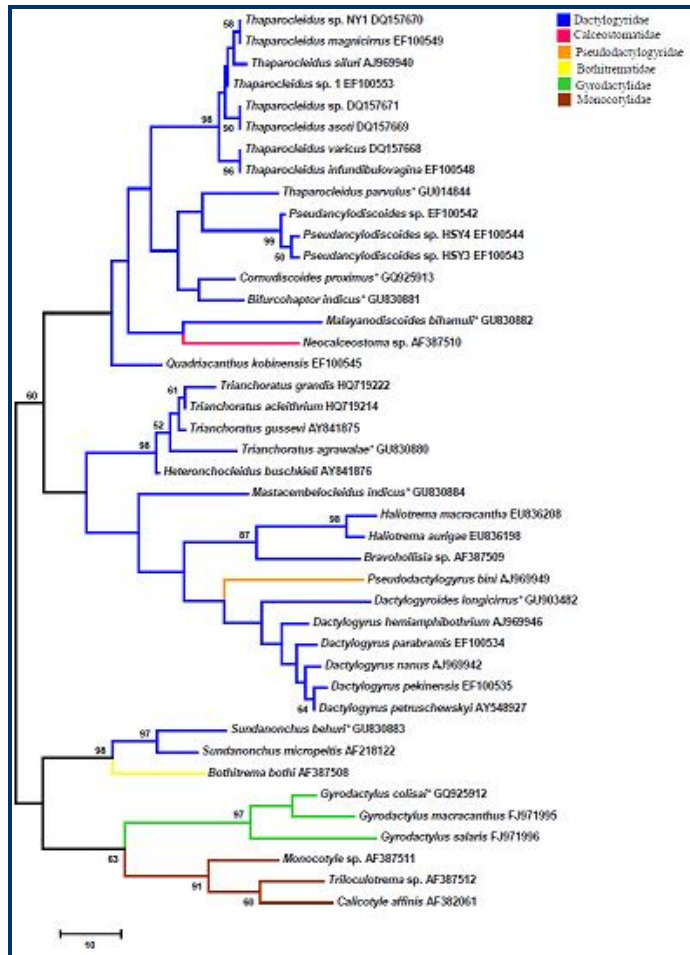

Figure 1: Neighbor joining (NJ) tree of monogenean species from different geographical regions. The tree was identical to that obtained using maximum likelihood (ML) and the numbers along branches represent bootstrap values given as NJ/ML. The asterisk indicates the species sequenced from India.

Predicted 28S RNA secondary structures and analyses
Secondary structures of 28S sequences of monogenean species were predicted by the online MFold package [18]. MFold is the most widely used algorithm for RNA secondary structure prediction that is based on a search for the minimal free energy state. Since, GC content is known to influence structural energy GC percentage was determined using a GC calculator (http://www.genomicsplace.com/gc_calc.html). Prediction of 28S consensus secondary for nine different monogeneans was

made by using the MARNA web server [19] based on both the primary and secondary structures. Default settings were used including base deletion, was scored 2.0, base mismatch 1.0, arc removing 1.5, arc breaking 1.75, and mismatch 2 with ensemble of shaped structures.

Figure 2: Maximum parsimony (MP) tree of monogenean species from different geographical regions. The asterisk indicates the species sequenced from India.
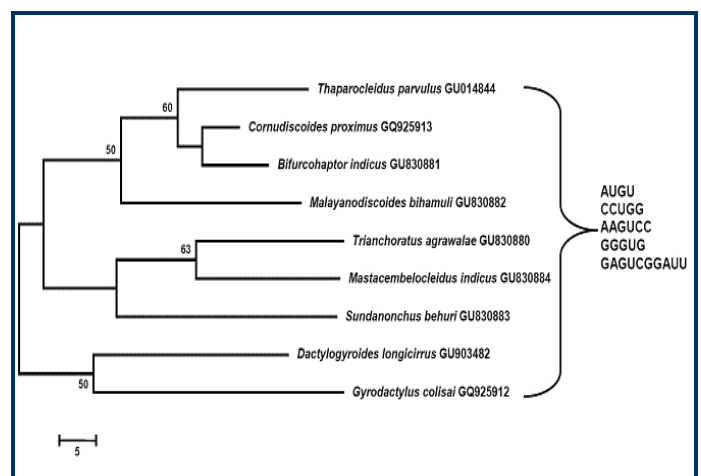
Motif identification, testing and validation

28S sequence motifs were identified from the aligned sequences by using PRATT software (http://www.ebi.ac.uk/pratt/). The C% parameter was adjusted to report pattern matching at 100% of the sequences input. The motifs were expressed by using the DNA alphabet (A, T, C, G) in PROSITE language [20]. Validation of the motifs was performed for each monogenean species using a "PATTERN MATCHING" web application (http://genoweb.univrennes1.fr/Serveur-GPO/outils_acces.php3?id_syndic=175). In order to test for additional reliability of monogenean species identification, the motifs were evaluated using the Basic Local Alignment Search Tool (BLAST) against the GenBank database of the National Centre for Biotechnology Information. BLAST outputs were then analysed to find only exact or perfect matches showing significantly high score and low *E*-values for species. The BLAST analysis investigated motifs that exhibited conserved sites obtained from the species. A motif was considered highly specific to a monogenean species if it matched most or all the 28S sequences available for that species.

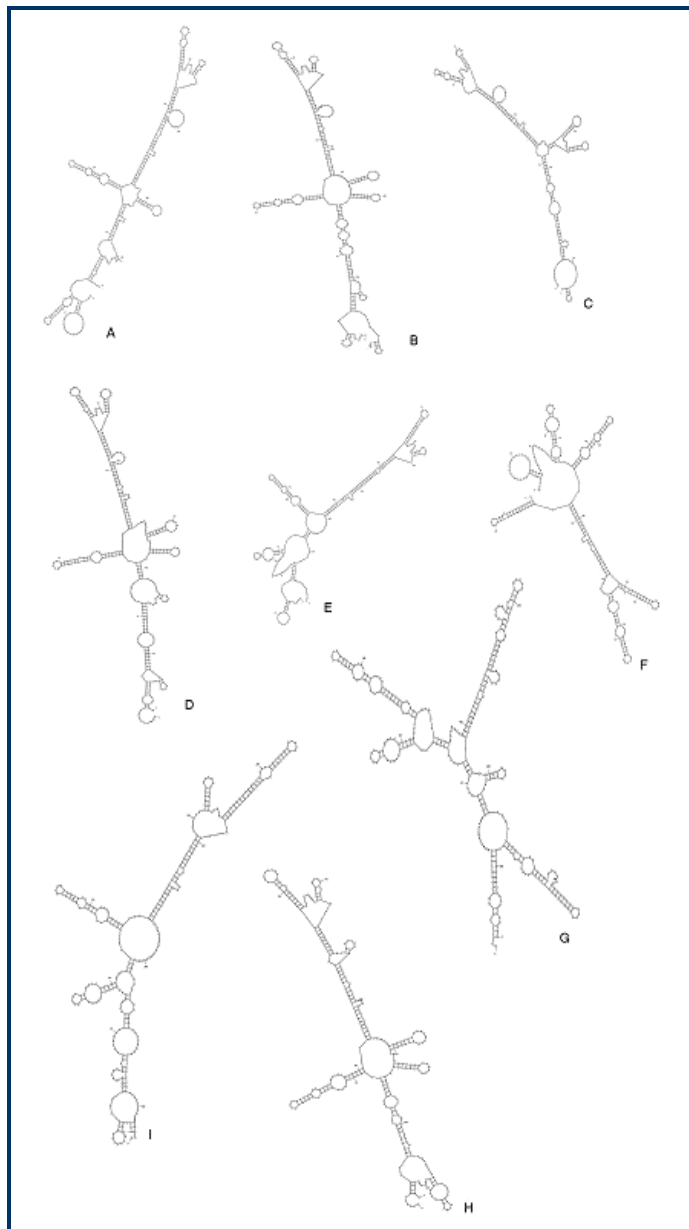### Results:
Construction of phylogenetic trees

Phylogenetic trees were made by comparing the 28S sequence of monogenean species from India with other species of different geographical isolates. Estimates of evolutionary divergence between sequences were conducted using the K-2-parameter model. Bootstrap results from NJ and ML analyses **(Figure 1)** indicate the phylogenetic position of these monogeneans. The MP tree **(Figure 2)** reveals a similar topology as observed in NJ and ML, but with bootstrap values lower than the NJ and ML trees. Among the monogeneans, phylogenetic relationships indicate that *T. parvulus* constitutes a clade with *Pseudancylodiscoides*, *Cornudiscoides*, *Bifurcohaptor* and *Quadriacanthus*. The species *T. parvulus*, *C. proximus* and *B. indicus* were found in a close molecular biological relationship. The species *T. parvulus*, *C. proximus* and *B. indicus* are both molecularly and morphologically closely related. *M. bihamuli* showed its validity as it is a monotypic genus and formed a separate clade with a *Neocalceostoma* species. Phylogenetic relationship of another group *Trianchoratus* and especially *T. agrawalae* is also molecularly supplemented that constitute a clade with *Heteronchocleidus* and *Mastacembelocleidus indicus* together with other species of genus *Trianchoratus*. The genera *Mastacembelocleidus* is found to be closely related to the genus *Trianchoratus* and established its validity by forming a different clade. *Dactylogyroides i.e.*, *D. longicirrus* also confirmed its position by forming a clade with its sister genus, *Dactylogyrus* from which it was originally differentiated. Another species of genus *Sundanonchus*, *S. behuri* validated itself by placing with its sister species *S. micropeltis* having high bootstrap values in NJ, ML and MP analysis along with genus *Bothitrema*. Recently, the generic placement of the species *S. behuri* was changed from genus *Urocleidus* to *Sundanonchus* on the basis of morphology [21]. Finally, within monogenea, another very species rich group *Gyrodactylus* includes *G. colisai* with a well-supported bootstrap value and supports its status by placing it with other species of this genus. The MP tree was obtained using the Close-Neighbor-Interchange algorithm [22] in which the initial trees were obtained with the random addition of sequences (10 replicates). Besides this, a separate phylogenetic analysis for nine different species from the Indian region were also obtained **(Figure 3)** based on MP method.

Figure 3: Topology obtained from maximum parsimony (MP) analysis for nine species of monogeneans reported from Indian region**.**

BIOINFORMATION

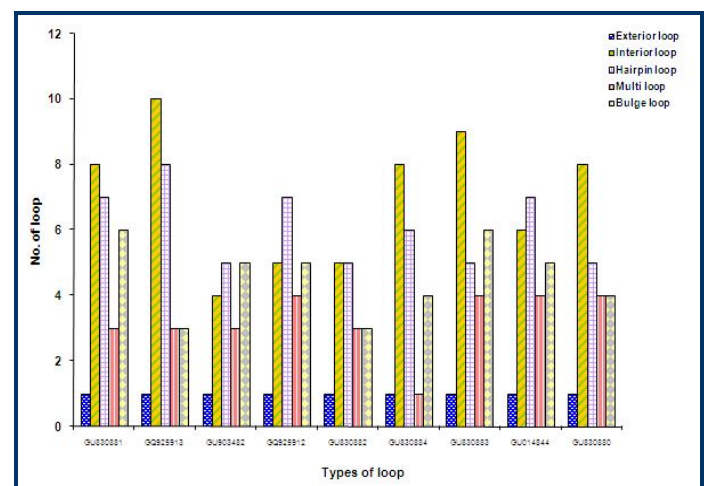## Secondary structure analysis

Predicted 28S rDNA secondary structural features from the different isolates were reconstructed **(Figure 4A-I)** with the highest negative free energy which provides the basic information for phylogenetic analysis. The free energy value of all these parasites and the characteristics of sequences for the 28S rDNA region shown in **Table 2 (see supplementary material)**. Length of 28S sequences of nine selected monogeneans ranged from 298 to 362 bp. G+C content for all isolates ranged from 45% to 53%.

Figure 4: Predicted 28S rDNA secondary structures for nine species of Monogenea from the Indian region. **(A)** *Bifurcohaptor indicus* **(B)** *Cornudiscoides proximus* **(C)** *Dactylogyroides longicirrus* **(D)** *Gyrodactylus colisai* **(E)** *Malayanodiscoides bihamuli* **(F)** *Mastacembelocleidus indicus* **(G)** *Sundanonchus behuri* **(H)** *Thaparocleidus parvulus* **(I)** *Trianchoratus agrawalae*

The RNA secondary structures of the 28S gene regions were analyzed on the basis of conserved stems and loops. The observed structural similarities in the predicted secondary

structure are further reflected at the energy level. Highest negative structural energy was observed in the *Gyrodactylus colisai* (-117.80 Kcal) followed by *Thaparocleidus parvulus* (-116.60 Kcal) while *Malayanodiscoides bihamuli* (-98.30 Kcal) had lowest negative free energy. Structural energy was found to be nearly similar in *Cornudiscoides proximus* and *Trianchoratus agrawalae viz.*, (-114.90 Kcal) and (-114.70 Kcal) respectively. Major common features found to be conserved in all nine species **(Figure 3)** are (1) AUGU; (2) CCUGG; (3) AAGUCC; (4) GGGUG; (5) GAGUCGGAUU. The secondary structure analysis reveals presence of external loop, multi loop, bulge loop, hairpin loop and interior loop. In order of preference for the conservation, it is found to be in the case of external loop and multi loop followed by bulge loop and hairpin loop with considerable variations was found in the interior loop **(Figure 5)**. External loop remained constant in all nine species. In the present work, we applied a more objective approach for the reconstruction of best alignment using secondary structure **(Figure 6)**. The figure shows alignment of nine monogenean species which evaluates 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 as the predicted structure **(Figure 7)**. The consensus structure is printed as a string of dots and brackets on top of the alignment. The string is well bracketed, such that each base pair in the structure is shown by corresponding opening and closing brackets. Furthermore, compatible base pairs are dark grey, 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. The saturation decreases with the number of incompatible base pairs; thus, it shows the structural conservation of the base pair. Prediction of the consensus structure is much higher in accuracy than the secondary structure prediction from single sequences.
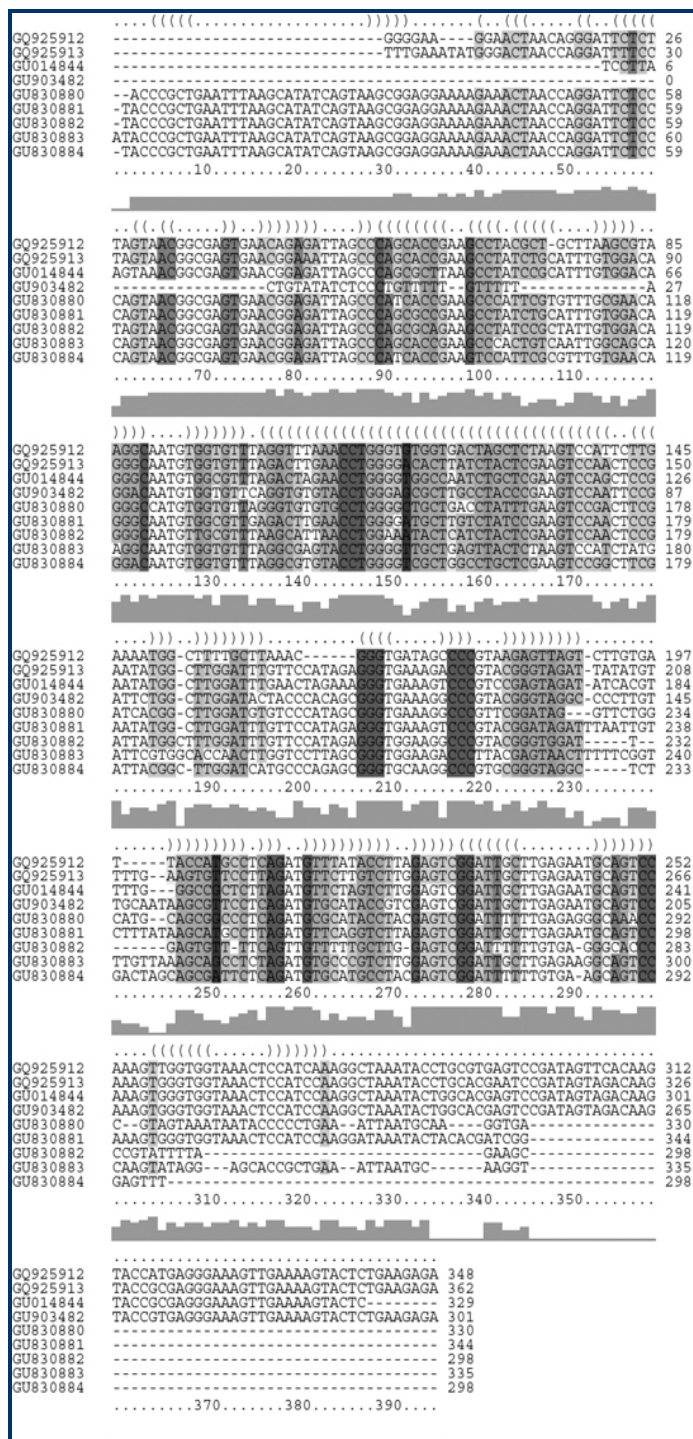

Figure 5: Distribution of various types of loops in 28S region in monogenean parasites of nine species from Indian region.

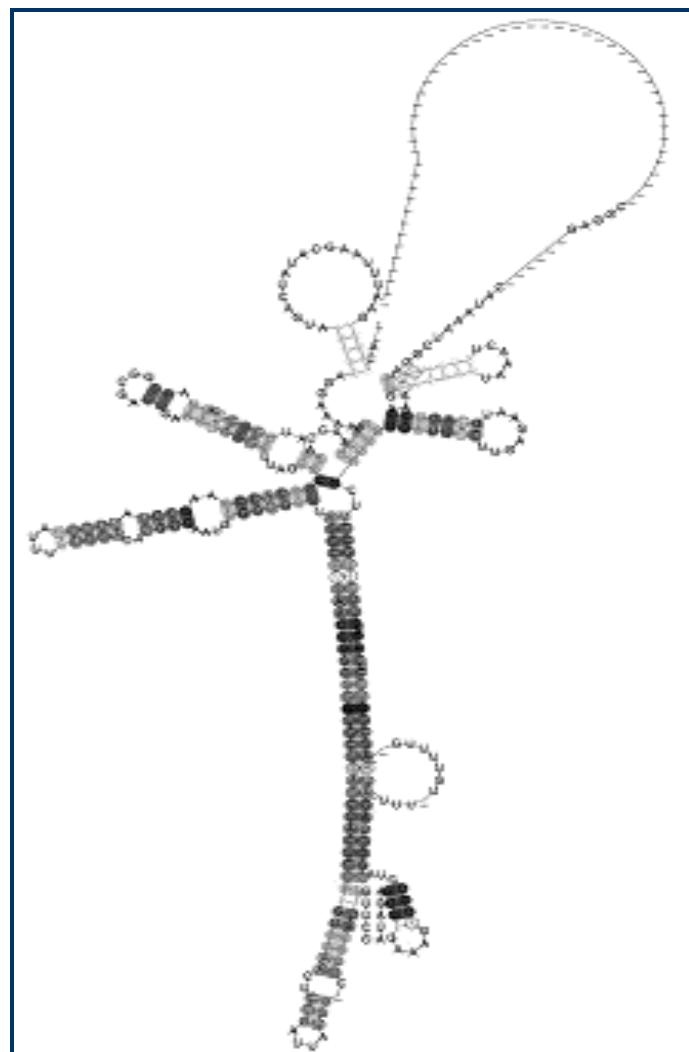## In-silico identification of monogenean species based on 28S motifs

During the study, we identified sequences of motifs from 28S rDNA region of the nine monogenean species. These motifs were screened, validated and as a final choice, six representative short sequence motifs of sizes inferior to 35

nucleotides **Table 3 (see supplementary material)** were selected. Total motifs of 28S sequences were tested by BLAST analysis against the generalized GenBank database. All the motifs showed exact or perfect BLAST matches with the monogenean sequences (best hits, 100; 100% of identity; E values, 3e^{-08} to 1e^{-07}). All motifs did not match any distantly related non-monogenean species available in the GenBank database. This proves that this tool can also be successfully used for phylogenetic analysis.

Figure 6: Nine monogenean species sequence alignment shows a consensus secondary structure. The structure is shown in the dot bracket format above the alignment and each corresponding bracket represents consensus base pairs of the alignment columns beneath. A sequence conservation profile is also shown in light grey bars below the alignment.

Figure 7: Consensus putative secondary structure of the 28S region of nine species of Monogenea from the Indian region. The consensus predicted secondary structure from Figure 6, light and dark grey according to the different types of base pairs in the corresponding alignment columns.

Discussion:

The taxonomy of monogeneans is based mainly on morphological data but DNA based methods work as supplementary tools for more authentic and accurate identification. In the case of monogeneans, sequences of 28S rDNA have been successfully used to study phylogenetic relationships at higher levels i.e., family and subfamily **[17, 23]** and generic levels **[24-27]**. These studies of 28S sequences from monogenean species indicate that there exists a high specific homogeneity. During this study, *T. parvulus* was found to be closely related to *Pseudancylodiscoides* and *Cornudiscoides* species because these genera also exhibit morphological similarities **[3]**. *Cornudiscoides* was similar to *Thaparocleidus* but differs in having a ventral bar that is divided into two parts and a long pair of modified marginal hooks which are usually situated close to the ventral anchors **[3]**. Besides this, in *Thaparocleidus* species they have complete ventral bar and lacks the long pair of marginal hooks found in *Cornudiscoides* species **[3]**. However,

Thaparocleidus and *Pseudancylodiscoides* exhibit close morphological resemblance but can be differentiated on the basis of having a divided ventral bar, the parts of which are well separated, and presence of larval type of marginal hooks [3]. Although, *Pseudancylodiscoides* has been considered a synonym of *Thaparocleidus* [28] but now this genus is considered valid [3]. Validity of *Bifurcohaptor* species is also erroneous in India because many species are probably synonymous and are in the category of *species inquirendae*. About 14 species of *Bifurcohaptor* reported by different workers from India were found to be synonym of *B. indicus* [29]. On the evidence of molecular phylogenetic analysis in this study by different methods, *B. indicus* is valid and closely related to *C. proximus* in having a divided ventral bar and ventral anchors disposed on separate lobes of the haptor. Another monogenean of Dactylogyridae studied *M. bihamuli* which was considered as a monotypic genus and therefore no other species can provide for comparative morphological or molecular information. Methods of phylogenetic reconstruction unequivocally place *T. agrawalae* with its sister species from different geographical regions amongst a clade with members of *Heteronchocleidus* with high bootstrap values. *Trianchoratus* and *Heteronchocleidus* showed close morphological and molecular similarity [30, 31, 32] so both genera have been assigned to the subfamily Heteronchocleidinae [33] (Ancyrocephalidae Bychowsky 1937).

A new genus *Mastacembelocleidus* within Monogenea under the family Dactylogyridae was proposed to include all freshwater monogeneans on Mastacembelid fish [34]. In this analysis, different tree methods exhibited that *M. indicus* forms a separate clade which shows their closeness to members of another clade including *Trianchoratus* and *Heteronchocleidus*. This study also shows the validation of *D. longicirrus*, an indigenous monogenean shows close similarity with the genus *Dactylogyrus* from which it was originally differentiated. Genus *Dactylogyroides* [35] was proposed for the worms previously described under the genus *Dactylogyrus*. The tree topologies derived from the phylogenetic analysis inferred from 28S rDNA data depicted that *Dactylogyroides* and *Dactylogyrus* as genetically closely related sister taxa. Therefore, based on our molecular results we propose that the species *D. longicirrus* is correctly accommodated in the genus *Dactylogyroides*. In the Indian monogeneans, the taxonomic position of *Sundanonchus* [36] within the class Monogenea has been unstable since the genus was proposed. In India, the species *S. behuri* was originally described as *Urocleidus behuri* [37] but this generic placement was not suitable because *Urocleidus* [38] is typically a freshwater North American genus. On the basis of the presence of a single intestinal tract, 16 hooks and a coiled copulatory tube, characteristics of *Sundanonchus* [21] placed *U. behuri* in this genus but retained the species valid. Though difficult, the accurate identification and validation of monogeneans species is very crucial for further study of this diverse, economically and ecologically important group. Kritsky et al. [34] considered Indian species of the genus *Urocleidus* [38] as *incertae sedis*. Therefore, Tripathi et al. [21] redescribed *Urocleidus behuri* [37] from *Nandus nandus* in India and transferred it to the genus *Sundanonchus* on the basis of morphological features only. But, the morphological characters among genera and species groups are affected by investigator's personal subjective view. Through this study an attempt has also been made to confirm the validation of this species from India by the use of molecular

characters and secondary structure prediction in taxonomic and phylogenetic studies. The results show that on the basis of molecular similarity with *S. micropeltis*, the *S. behuri* shows its validation and correct generic placement with high bootstrap values. Genus *Gyrodactylus* within monogenea is very species rich, for which the morphological and morphometric diagnosis is evidently difficult [39]. The best evidence of this difficulty is the ratio of described and named species to the estimated global number of species and predicted that the real number of species might be more than 20 000, yet only 470 names are considered valid and available [39]. It is difficult to manage 20 000 species in a morphological archive based on subtle differences in the opisthaptoral hard parts. Thus, additional and more informative characters are needed to understand the evolution and even taxonomy of parasites in this genus. Our analysis show that, beside the morphological features, molecular data further supports that *G. colisai* fits in the *G. neonephrotus anguillae* sp. group [40] with high bootstrap values.

RNA structure and prediction analyses plays an important role in evaluating the evolutionary relations that link all organisms with each other which led Woese to propose the Archaea as a distinct major branch on the "Tree of Life" [41]. RNA structure proves commonly used markers for phylogenetic reconstruction and knowledge of RNA secondary structure can improve alignment quality [42]. Elements of RNA secondary structure themselves can be treated as evolving characters and phylogenetic connections may be traced by changes in structural character states [43]. Predicted rRNA structures of monogeneans reported from India shows similarities in their thermodynamic energy and structural parameters like in different types of loops. Since rRNA forms evolutionarily well conserved secondary structures and these structures are related to the functions of molecules. Thus, computing the consensus structure that is common to several related RNA sequences can drastically improve the prediction [44]. It can be used as an additional source of data incorporating structural parameters of molecules for the study of monogenean phylogeny. Prediction of consensus have been made for nine species of monogeneans using individual sequences which were then combined in a form of sequence alignment for the finding of conserved structure motifs that is needed for the accuracy. This study also investigated the development of short 28S sequence motifs, as DNA oligonucleotide barcodes for unambiguous, correct and easy identification of monogenean species. A critical point in monogenean taxonomy is the identification of these small parasites and this difficulty can be overcome by motifs sequences. The present in-silico identification of nine monogenean species with 28S motifs is consistent with investigations made using traditional approaches like by morphology as well as through molecular phylogenetics and secondary structure predictions. This approach provides a new tool for an accurate identification of monogenean species and DNA barcoding which also offers new ways of understanding their life cycles more clearly. In conclusion, this study reveals that the 28S rDNA gene may prove useful for studies of systematics of parasitic Platyhelminthes. Molecular study of this group along with secondary structure analysis could be a valuable tool to distinguish new species and too strong monogenean systematic because on the basis of morphological studies identification and validation of these parasites is very difficult.

BIOINFORMATION

Conclusion:
This study also reveals that 28S sequences are reliable for sequence analysis and secondary structure prediction to resolve the phylogeny. 28S sequence motifs allowed an accurate in-silico distinction of the monogeneans and indicate that motifs (≤50 bp in size) can be considered a promising tool for monogenean species identification. Further studies are required for more adequate elucidation of the phylogeny of monogenean parasites using other genome sequences.

References:
[1] Boeger W & Kritsky DC, Taylor and Francis, London, UK and New York, USA 2001, pp. 92
[2] Poisot T & Desdevises Y, *Biol J Linn Soc.* 2010 **99**: 559
[3] Lim LH *et al. Syst Parasitol.* 2001 **50**: 159 [PMID: 11590306]
[4] Hadjiolov AA *et al. Nucleic Acids Res.* 1984 **12**: 3677 [PMID: 6328433]
[5] Michot B *et al. Eur J Biochem.* 1990 **188**: 219 [PMID: 2138538]
[6] Schnare MN *et al. J Mol Biol.* 1996 **256**: 701 [PMID: 8642592]
[7] Gillespie JJ *et al. Mol Biol Evol.* 2005 **22**: 1593 [PMID: 12814653]
[8] Olson PD *et al. Int J Parasitol.* 2003 **33**: 733 [PMID: 8845967]
[9] Kjer KM. *Mol Phylogenet Evol.* 1995 **4**: 314 [PMID: 8845967]
[10] Hickson RE *et al. Mol Biol Evol.* 1996 **3**:150
[11] Cunningham CO *et al. J Parasitol.* 2000 **86**: 567 [PMID: 10864256]
[12] Lydeard C *et al. Mol Phylogenet Evol.* 2000 **15**: 83 [PMID: 10764537]
[13] Bychowsky BE, *Series Biology.* 1937 **4**: 353
[14] Dávidová M *et al. Parasitol Res.* 2005 **95**: 398 [PMID: 15739070]
[15] Malmberg G, *Arkiv fu¨r Zoologi Serie.* 1970 **2**: 1
[16] Mollaret I *et al. Int J Parasitol.* 2000 **30**: 171 [PMID: 10704600]
[17] Tamura K *et al. Mol Biol Evol.* 2011 **28**: 2731 [PMID: 21546353]
[18] Zuker M, *Nucleic Acids Res.* 2003 **31**: 3406 [PMID: 12824337]
[19] Siebert S & Backofen R, *Bioinformatics.* 2005 **21**: 3352 [PMID: 15972285]
[20] Bairoch A & Bucher P, *Nucleic Acids Res.* 1994 **22**: 3583 [PMID: 7937064]
[21] Tripathi A *et al. Parasitol Int.* 2009 **58**: 18 [PMID: 18761106]
[22] Nei M & Kumar S 2000, Oxford University Press, New York.
[23] Jovelin R & Justine JL, *Int J Parasitol.* 2001 **31**: 393 [PMID: 11306118]
[24] Mollaret I *et al. Int J Parasitol.* 2000 **30**: 659 [PMID: 10779582]
[25] Wu XY *et al. Parasitol*ogy. 2005 **130**: 669 [PMID: 15977904]
[26] Šimková A et al. *Parasitol.* 2006 **133**: 43 [PMID: 16515727]
[27] Lee SU *et al. Korean J Parasitol.* 2007 **45**: 181 [PMID: 17876163]
[28] Lim LHS, *Syst Parasitol.* 1996 **35**: 207
[29] Pandey KC & Singh HS, *Indian J Helminthol.* 1990 **6**: 39
[30] Lim LHS, *Para¬sitol Hung.* 1986 **19**: 31
[31] Lim LH, *Int J Parasitol.* 1998 **28**: 1495 [PMID: 9801913]
[32] Tan WB *et al. The Raffles Bulletin of Zoology.* 2011 **59**: 127
[33] Price CE, *Quarterly Journal of the Florida Academy Science.* 1968 **29**: 199
[34] Kritsky DC *et al. Folia Parasitol.* 2004 **51**: 291 [PMID: 15729940]
[35] Gusev AV, *Bulletin of the Fisheries Research Station of Ceylon.* 1963 **16**: 53
[36] Lim LHS & Furtado JI, *Folia Parasit.* 1985 **32**: 11
[37] Agrawal N & Singh HS, *Pranikee.* 1982 **3**: 22
[38] Mueller JF, *Roosevelt Wild Life Annals.* 1934 **3**: 335
[39] Bakke TA *et al. Adv Parasitol.* 2007 **64**: 161
[40] Malmberg G, *Czechoslovak Academy of Sciences, Prague, The Czech Republic* 1964, pp. 203-230.
[41] Woese CR *et al. Proceedings of the National Academy of Sciences USA.* 1990 **87**: 4576
[42] Goertzen LR *et al. Mol Phylogenet Evol.* 2003 **29**: 216 [PMID: 13678678]
[43] Knudsen V & Caetano-Anolles G, *Nucleic Acids Res.* 2008 **36**: w85 [PMID: 18448469]
[44] Hofacker IL *et al. J Mol Biol.* 2002 **319**: 1059 [PMID: 12079347]

BIOINFORMATION

Supplementary material:

Table 1: List of monogenean, their hosts, site of infection, voucher and GenBank accession numbers.

Host	Monogenean spp.	Site of infection	Voucher no.	Accession no.
Puntius sophore	*Dactylogyroides longicirrus*	Gills	HS/Monogenea/2009/01	GU903482
Mystus vittatus	*Bifurcohaptor indicus*	Gills	HS/Monogenea/2009/02	GU830881
Mystus vittatus	*Cornudiscoides proximus*	Gills	HS/Monogenea/2009/03	GQ925913
Anabas testudineus	*Trianchoratus agrawalae*	Gills	HS/Monogenea/2009/04	GU830880
Mystus vittatus	*Thaparocleidus parvulus*	Gills	HS/Monogenea/2009/05	GU014844
Macrognathus aculeatus	*Mastacembelocleidus indicus*	Gills	HS/Monogenea/2009/06	GU830884
Colisa fasciata	*Gyrodactylus colisai*	Skin	HS/Monogenea/2009/07	GQ925912
Notopterus chitala	*Malayanodiscoides bihamuli*	Gills	HS/Monogenea/2009/08	GU830882
Nandus nandus	*Sundanonchus behuri*	Gills	HS/Monogenea/2009/09	GU830883

Table 2: Secondary structural features of 28S sequences from monogenean isolates of India from different hosts.

Isolates	Sequence length	G+C%	AU (no.)	GU (no.)	Energy (Kcal)
GU830881	344	46	184	178	-109.80
GQ925913	362	45	199	192	-114.90
GU903482	301	49	152	161	-111.30
GQ925912	348	45	191	184	-117.80
GU830882	298	46	159	163	-98.30
GU830884	298	53	140	160	-103.20
GU830883	335	49	171	172	-107.20
GU014844	329	49	167	173	-116.60
GU830880	330	51	160	169	-114.70

Table 3: BLAST search outputs for Monogenea 28S sequence motifs against NCBI GenBank database.

Monogenean species motif patterns	No. of best hits	Identity (%)	E-value
>Pattern 1 G-T-C-G-G-A-T-T-[GT]-[CT]-T-T-G-[AT]-G-A-x(1,2)-G-C-A-[ACG]-[ACT]-C-C-[ACG]-[AG]-[AGT]-[AGT]-[GT]-[AGT]	100	100	$1e^{-07}$
>Pattern 2 G-T-C-G-G-A-T-T-[GT]-[CT]-T-T-G-[AT]-G-A-x(1,2)- G-C-A-[ACG]-[ACT]-C-C-x(1,3)-A-[GT]-T-[AGT]-[AGT]	100	100	$3e^{-08}$
>Pattern 3 G-T-C-G-G-A-T-T-x(1,2)-T-T-[GT]-[AG]-[AGT]-[AG]-A-[AGT]-G-x(0,1)-C-A-[ACG]-[ACT]-C-C-[ACG]-[AG]-[AGT]-[AGT]-[GT]-[AGT]	100	100	$1e^{-07}$
>Pattern 4 G-T-C-G-G-A-T-T-[GT]-[CT]-T-T-G-[AT]-G-A-x(1,2)-G-x(1,2)-A-[ACG]-[CT]-C-C-[ACG]-[AGT]-[AGT]-[AGT]-T-[AGT]	100	100	$1e^{-07}$
>Pattern 5 G-T-C-G-G-A-T-T-[GT]-[CT]-T-T-G-[AT]-G-A-x(1,2)-G-C-A-[ACG]-[ACT]-C-C-x(1,3)-G-T-[AGT]-[GT]	100	100	$3e^{-08}$
>Pattern 6 T-C-G-G-A-T-T-[GT]-[CT]-T-T-G-[AT]-G-A-x(1,2)-G-C-A-[ACG]-[ACT]-C-C-[ACG]-[AG]-[AGT]-[AGT]-[GT]-[AGT]	100	100	$3e^{-07}$