Use of sequence motifs as barcodes and secondary structures of Internal Transcribed spacer 2 (ITS2, rDNA) for identification of the Indian liver fluke, *Fasciola* (Trematoda: Fasciolidae)

Pramod Kumar Prasad[1], Veena Tandon[1], Devendra Kumar Biswal[2], Lalit Mohan Goswami[1] and Anupam Chatterjee[3]

[1]Department of Zoology, North Eastern Hill University, Shillong, Meghalaya, 793022, India; [2]Bioinformatics Centre, North Eastern Hill University, Shillong, Meghalaya, 793022, India; [3]Department of Biotechnology and Bioinformatics, North Eastern Hill University, Shillong, Meghalaya, 793033, India. Pramod Kumar Prasad - E-mail: pramod812@gmail.com; Tel: (+91) 0364 2722312; Fax: (+91) 0364 2722301; *Corresponding author

Abstract:
Most phylogenetic studies using current methods have focused on primary DNA sequence information. However, RNA secondary structures are particularly useful in systematics because they include characteristics that give "morphological" information which is not found in the primary sequence. Also DNA sequence motifs from the internal transcribed spacer (ITS) of the nuclear rRNA repeat are useful for identification of trematodes. The species of liver flukes of the genus *Fasciola* (Platyhelminthes: Digenea: Fasciolidae) are obligate parasitic trematodes residing in the large biliary ducts of herbivorous mammals. While *Fasciola hepatica* has a cosmopolitan distribution, the other major species, i.e., *F. gigantica* is reportedly prevalent in the tropical and subtropical regions of Africa and Asia. To determine the *Fasciola* sp. of Assam (India) origin based on rDNA molecular data, ribosomal ITS2 region was sequenced (EF027103) and analysed. NCBI databases were used for sequence homology analysis and the phylogenetic trees were constructed based upon the ITS2 using MEGA and a Bayesian analysis of the combined data. The latter approach allowed us to include both primary sequence and RNA molecular morphometrics and revealed a close relationship with isolates of *F. gigantica* from China, Indonesia and Japan, the isolate from China with significant bootstrap values being the closest. ITS2 sequence motifs allowed an accurate in silico distinction of liver flukes. The data indicate that ITS2 motifs (≤ 50 bp in size) can be considered promising tool for trematode species identification. Using the novel approach of molecular morphometrics that is based on ITS2 secondary structure homologies, phylogenetic relationships of the various isolates of fasciolid species have been discussed.

Keywords: *Fasciola hepatica*; *Fasciola gigantica*; secondary structure; internal transcribed spacer; bar coding; motifs

Background:
The identification of closely related species based on morphological characters can be difficult. This is particularly the case for soft-bodied animals such as digenean trematodes. PCR-based techniques utilizing the rDNA ITS2 sequences, which occur between the 5.8S and 28S coding regions, have proven to be a reliable tool to identify the helminth species and their phylogenetic relationships [1]. The nuclear ribosomal DNA second internal transcribed spacer (ITS2) sequences, which occur between the 5.8S and 28S coding regions, have proven useful for diagnostic purposes at the level of species. *Fasciola* spp and isolates of *Fascioloides magna*, another member of the same family, from different geographical regions have been discriminated on the basis of ITS sequences [2]. The ITS2 sequences have also been used to characterize the liver flukes from mainland China, which include *F. hepatica*, *F. gigantica* and an intermediate genotype, including polymorphism among ITS2 copies within the same fluke individual [3]. The emergence of sequence-based identification with a BLAST similarity search connected to public databases has resolved several experimental and taxonomic constraints. In addition, BLAST outcomes give no information about species delimitation for closely related species. The identification of unknown ITS sequences based on these approaches therefore needs to be supported by phylogenetic analysis [4].

The trematode flukes of the genus *Fasciola* (the sheep liver fluke) are parasites of herbivores and infect humans accidentally causing fascioliasis worldwide. The parasite is very cosmopolitan in distribution being found throughout all regions of the world, both temperate and tropical. *F. hepatica* is the causative agent of fascioliasis or 'liver rot' in ruminants, where it may be an important pathogen. Human infections with *F. hepatica* are found in areas where sheep and cattle are raised, and where humans consume raw watercress, including Europe, the Middle East and Asia [5]. Infection with *F.gigantica* on the other hand is found more commonly in tropical regions of the world, areas affected including Africa, Asia, many Pacific islands including Hawaii (where it has been reported in man), the Middle East and Southern Europe and the south of the USA [6]. Both fasciolids follow a similar two-host life cycle, in which freshwater snails of the family Lymnaeidae act as intermediate or vector hosts and a broad spectrum of mammals including humans, mainly herbivorous large-sized species, act as definitive hosts [7].

314

Most phylogenetic studies using current methods have focused on primary DNA sequence information. However, RNA secondary structures are particularly useful in systematics because they include characteristics, not found in the primary sequence, that give 'morphological' information [8]. 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 ITS2 secondary structure homologies (geometrical features, bond energies, base composition etc.) is recently being used to study the phylogenetic relationships of various species [9]. This method allows one to take into account the regions where multiple alignments are barely reliable because of a large number of insertions and deletions. This method is based on the assumption that secondary structure can be phylogenetically as significant as primary sequence. It is well known that rRNA is highly conserved throughout evolution. Thus, the secondary-structure elements of the RNA molecule, i.e., the helices, loops, bulges, and separating single-stranded portions, can be considered phylogenetic characters [10, 11, 12].

At the same time, the ITS offers sequence motifs that are useful for the development of DNA barcoding. This is a technique that uses short DNA sequences from a standardized region of the genome as a diagnostic "biomarker" for species identification [13, 14] as testified to by the creation of the consortium for the Barcode of life (http://www.barcoding.si.edu/). Different species have different barcodes, which allow them to be used to (i) identify specimens, (ii) discover possible new species, and (iii) make taxonomy more effective for science and society. In the present study, our main objective was to identify the Indian liver fluke (the Assam isolates) collected from ruminant hosts by the design and testing of ITS sequence motifs for oligonucleotide bar codes. We also aimed at determining the species prevalent in the northeastern region of the country, by comparing these sequences by primary sequence analysis and molecular morphometrics data.

Methodology:
Parasite material and DNA isolation
Adult *Fasciola* were obtained in live form from hepatic biliary ducts of freshly slaughtered cow, *Bos indicus*. The worms recovered from these hosts represented the geographical isolates from Assam, Northeast India and morphologically resembled *Fasciola gigantica* (deposition number of paratypes at Zoological Survey of India, Kolkata = W7787/1). Eggs were obtained from mature adult flukes by squeezing between two glass slides. For the purpose of DNA extraction, adult flukes were processed singly; eggs recovered from each of these specimens were also processed separately. The adult flukes were first immersed in digestion extraction buffer (containing 1% SDS, 25 mg Proteinase K) at 37° C overnight. DNA was then extracted from lysed individual worms by standard ethanol precipitation technique [15] and also extracted on FTA cards using Whatman's FTA Purification Reagent as described elsewhere [16]. DNA from the eggs was extracted only with the FTA card technique.

### DNA amplification, Sequencing and its Analysis
The rDNA region spanning the ITS2 region was amplified from DNA obtained from the fluke by PCR. We used the universal primers, considered to be the general primers for trematodes and are designed based on conserved ITS sequences of *Schistosoma* species following [17] as given in **Figure 1**. The PCR amplification was performed following the standard protocol [18] with minor modifications. The resultant PCR products were separated by electrophoresis through 1.6% (w/v) agarose gels in TAE buffer, stained with ethidium bromide, transilluminated under ultraviolet light and then photographed. The known size fragments of Phi X 174 DNA/ Hae III Digest in agarose gel were used as marker. For DNA sequencing, the PCR products were purified using Genei Quick PCR purification Kit, and sequenced in both directions using PCR primers on an automated sequencer by DNA sequencing services of TCGA, New Delhi and Bangalore Genei Ltd. The DNA sequences were put to further analysis by using various Bioinformatics tools including similarity search BLAST (URL http://www.ncbi.nlm.nih.gov/blast) and phylogenetic prediction by CLUSTALW (URL http://www.ebi.ac.uk/clustalw) for each query DNA sequence.

Molecular Phylogenetic Analysis
Initially, the sequences were aligned using ClustalW multiple alignment [19] with the default gap and extension penalties used by this program. Phylogenetic tree-building methods presume particular evolutionary models. Therefore, while interpreting the results obtained, we considered different tree building models to entertain possible explanations. Only unique sequences were used in tree construction. ITS sequences were entered in the MEGA for construction of the phylogenetic trees using Maximum Parsimony and distance methods namely the Neighbor-Joining, UPGMA and Minimum Evolution. Branch support was given using 1000 bootstrap replicates in MEGA [20].

ITS2 region- 3S (forward): 5'-GGTACCGGTGGATCACTCGGCTCGTG-3'
               A28 (reverse): 5'-GGGATCCTGGTTAGTTTCTTTTCCTCCGC-3'

Figure 1: Primer design for ITS2
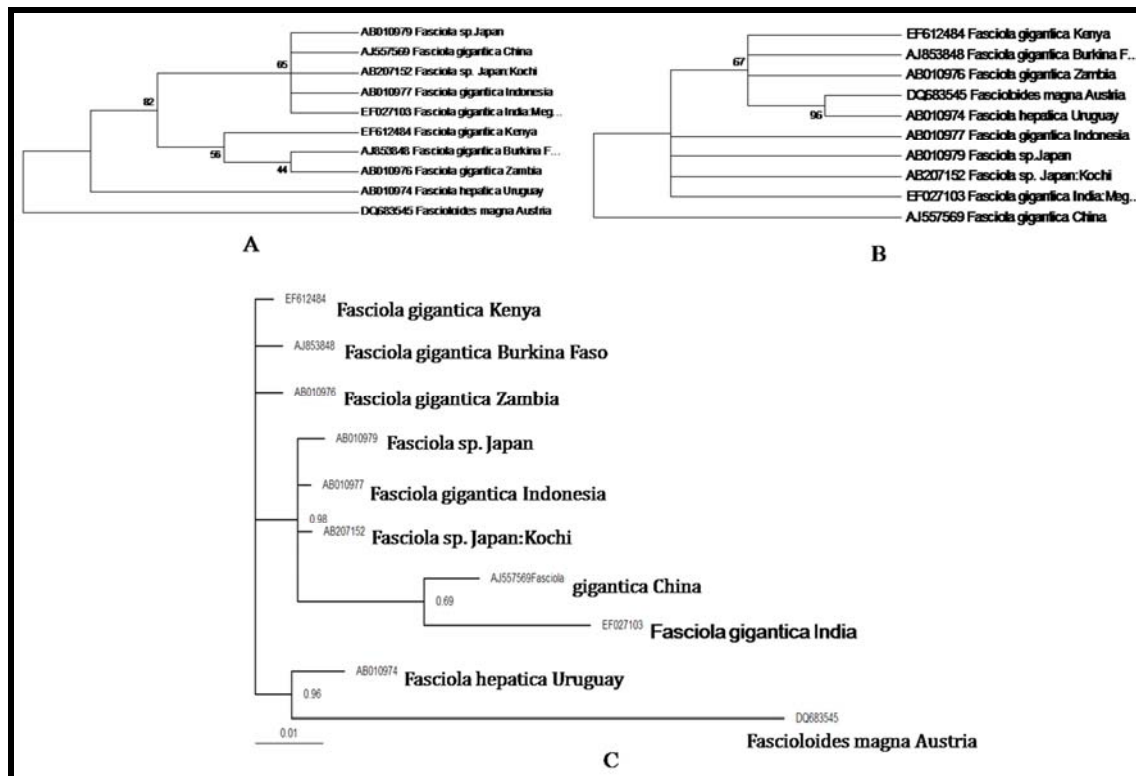
Figure 2: Phylogenetic trees of ITS2 sequences of fasciolid species showing bootstrap values. (A) Neighbor Joining; (B) Maximum Parsimony; (C) Hypothetical Bayesian analysis.

Bayesian phylogenetic analysis

DNA sequences were aligned using ClustalX 2.0.7 [21]. The interleaved NEXUS file was edited manually in order for it to be recognized by Mr. Bayes V3.1.2 programme. Phylogenetic analysis was carried out using the Bayesian approach with combined datasets using Mrbayes 3.1 [22], wherein each data partition is allowed to have a different evolution rate. The model of evolution and prior settings for individual loci was set to the GTR model with gamma-distributed rate variation across sites and a proportion of invariable sites. Metropolis-coupled Markov chain Monte Carlo (MCMCMC) sampling was performed with four incrementally heated chains that were combinatorial run for 40,000 generations. The convergence of MCMCMC was then monitored by examining the value of the marginal likelihood through generations. Coalescence of substitution rate and rate model parameters were also examined. Average standard deviation of split frequencies was checked and the generations were kept on adding until the standard deviation value was below 0.01. The values slightly differed because of stochastic effects. The sample of substitution model parameters and samples of trees and branch lengths were summarized by the "sump burnin" and "sumt burnin" commands respectively. The values in the following commands were adjusted as per the 25% of our samples. The cladogram with the posterior probabilities for each split and a phylogram with mean branch lengths were generated and subsequently read by the tree drawing program Tree view V1.6.6 [23].

Motif identification, testing and validation

The ITS sequence motifs were identified from aligned sequences of the data set for the species using PRATT software (http://genoweb1.irisa.fr/Serveur-GPO/outils_acces.php3?id_syndic=70). The minimum percentage of sequences to match (C %) parameter was adjusted to report pattern matching at 100% of the sequence input. The motifs were expressed using the DNA alphabet (A, T, C, G) in PROSITE language. The validation of the motifs was performed for each species using a "PATTERN MATCHING" Web application (http://genoweb.univ-rennes1.fr/Serveur-GPO/outils_acces.php3?id_syndic=186).We considered that a motif was highly specific to a *Fasciola* species if it matched most or all the ITS sequences of this species but no other ITS of any other trematodes species.

Evaluation of the ITS motif specificity through BLAST analysis

The *Fasciola* ITS sequence motifs (patterns 1-15) were subjected to BLAST algorithm against the non-redundant GenBank database of the National Center for Biotechnology Information (nr at NCBI). The BLAST output were then analyzed to find only the exact or perfect pairwise matches showing significantly high scores in terms of percent identity and low expect (E) values for each species.

316
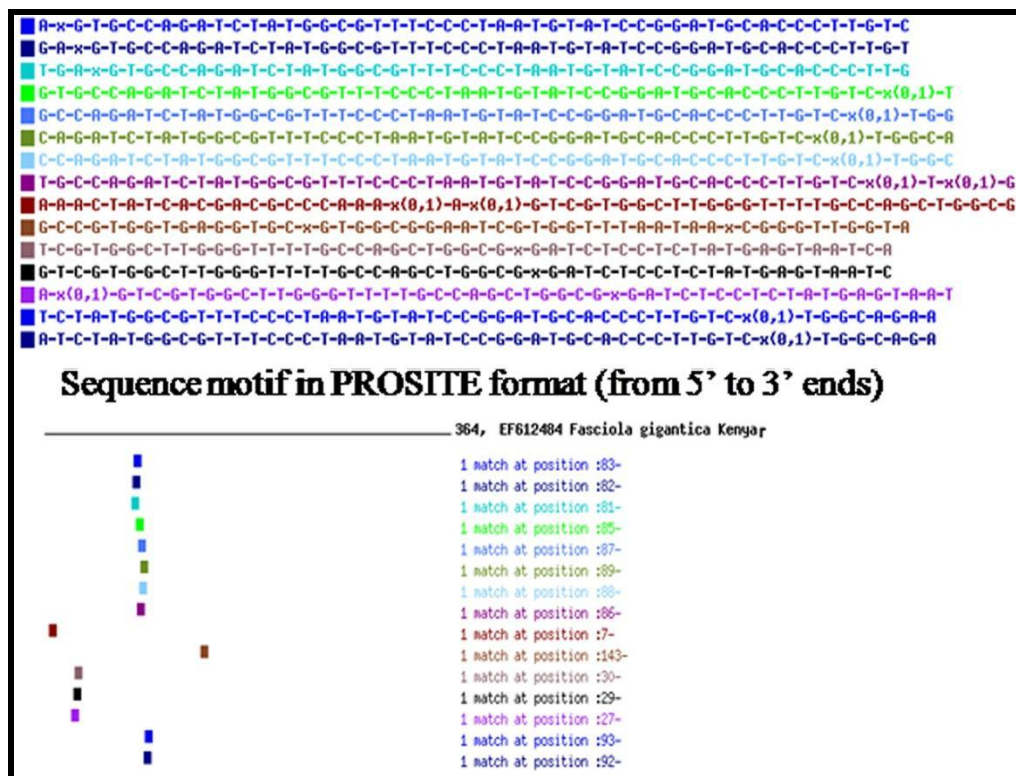
Figure 3: Sample of sequence motif in PROSITE format (from 5' to 3' ends) and pattern matching of ITS motifs of different geographical isolates of fasciolid species.

Predicted ITS2 RNA secondary structures and analyses
Secondary structures of ITS2 sequences of various fasciolid species were reconstructed by aligning their sequences using Bioedit [24]. The acquired structures with restrictions and constrains were submitted in MFOLD [25]. RNA was folded at a fixed temperature of 37° C, and the structure chosen from different output files was the desired 6-helicoidal ring or the one with the highest negative free energy if various similar structures were obtained.

Results:
PCR amplification of ITS regions and its analysis
The PCR-amplified products were successfully obtained using the primers as mentioned above. The nucleotide sequences were obtained for ITS2 of rDNA and were compared with sequences of other fasciolid species obtained from Genbank. The fragments of amplified DNA were estimated to be 480-550bp long. For comparative purpose, the ITS2 sequences of fasciolids from various geographical regions were obtained from GenBank (Table 1). The Blast hit results show that the query ITS sequences were more similar to the sequences of various geographical isolates of *Fasciola* sp., *F. hepatica* and *F. gigantica* besides *Fascioloides magna* (belongs to the same family, i.e., Fasciolidae).

Phylogenetic trees
Phylogenetic trees were obtained by comparing the sequences of *F. gigantica* and available ITS sequences for other fasciolid species. Phylogenetic analyses using the various distance methods and character method like Maximum Parsimony showed that the topology is similar among the trees obtained with significant bootstrap support for the clades. The values of 70% and above in the bootstrap test of phylogenetic accuracy indicates reliable grouping among different members of fasciolids. The phylogenetic trees constructed in MEGA revealed a close relationship with isolates of *F. gigantica* from China, Indonesia and Japan (**Figure 2**).

***In silico* identification of *Fasciola* sp. based on pattern matching ITS motifs**
A total of 15 ITS2 PROSITE motifs were tested by BLAST analysis against the generalized GenBank database at NCBI (**Figure 3**). The *Fasciola* motifs exhibited exact or perfect matches with *Fasciola* from different geographical isolates with significant E-value scores. Eight (pattern 1-8) out of 15 *Fasciola* motifs exactly matched the sequences of *Fasciola* isolates from GenBank (best hits = more than 100; 100% of identity; E-values = $2e^{-16}$ to $1e^{-17}$). Five (pattern 11-15) showed (best hits = more than 100; 95% identity and significant E-value ranging from $5e^{-14}$ to $1e^{-14}$). The remaining two motifs (pattern 9-10) exhibited more than 100 best hits with 92% identity and an E-value score ranging from $4e^{-11}$ to $1e^{-11}$ (**Tables 2 & 3 in supplementary material**).
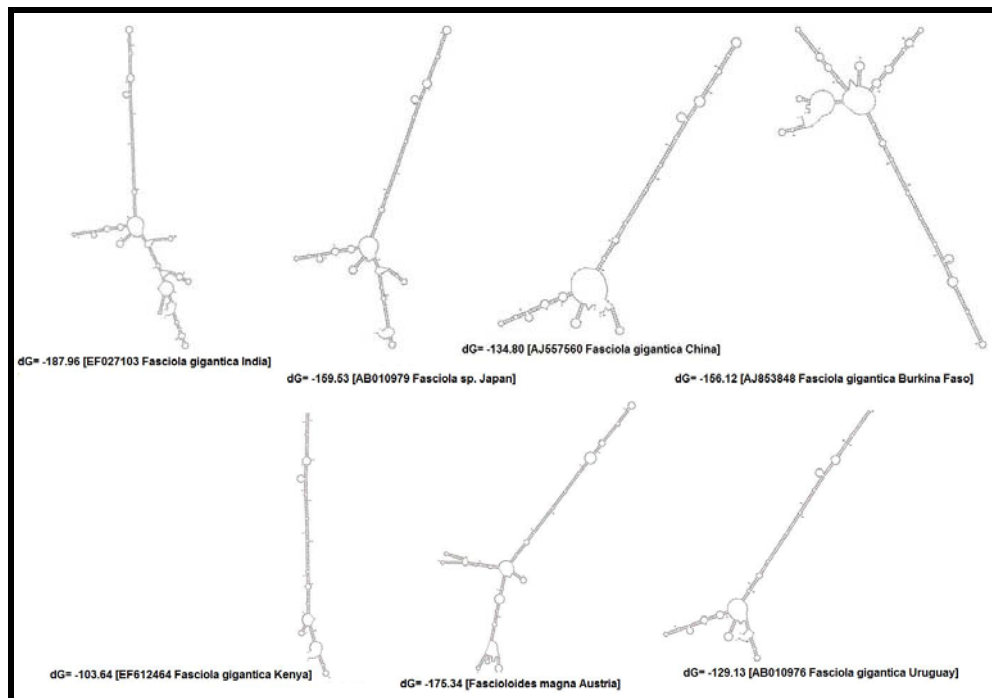
Figures 4: Predicted ITS2 RNA secondary structures and their structure formation enthalpies according to MFOLD.

Secondary structure analysis

Six predicted RNA secondary structures were reconstructed from the unique sequences with highest negative free energy of *F. gigantica* to provide the basic information for phylogenetic analysis **(Figure 4)**. The ITS2 plus flanking regions of nuclear region ranged from 606bp in *F. gigantica* India to a minimum length of 361bp in *F. gigantica* China. The secondary structural features of ITS2 regions as shown in the figure were analysed based on conserved stems and loops. *F.* sp. Japan, *F. gigantica* Indonesia, *F. gigantica* Zambia and *F. hepatica* Uruguay exhibited same type of secondary structures. *F. gigantica* isolates from India and China show overall similarity in the ITS2 rRNA folding and have identical secondary structure. Secondary structures of the remaining species are weakly variable. The observed similarities at the secondary structural level are further reflected at the energy level. The only difference in their topology is due to differences in nucleotide sequences. These secondary structure predictions indicate that the domains basepair to form a core region central to several stem features implying that conservedness is more important for the proper rRNA folding pattern. Moreover the observed phylogenetic trend was identified with respect to the target accessibility sites for the seven different isolates. The orders of preference were interior loop, bulge loop, multiple branch loop, hairpin loop and exterior loop in all the isolates. The topology based only on the predicted RNA secondary structure of the ITS2 region resolved most relationships among the species studied. Bayesian analysis of the alignment retained the same topology and supported the same branches as the primary sequences **(Figure 4)**. ITS2 sequence of the Indian isolate revealed closest similarity with the Chinese isolate (*Fasciola gigantica*) with significant bootstrap value.

Discussion:

Species identification and evolutionary inference generally use molecular and phylogenetic approaches on the ITS as target genomic loci. But taxonomy of *Fasciola* spp has been based mainly on morphological data complemented with ecological, cytological and pathological results as well as clinical manifestations. Morphological differences found on stained and mounted adult specimens have been widely used to discriminate between platyhelminth species. It is possible to distinguish between adult *F. hepatica* and *F. gigantica* on the basis of morphology, but much variation exists. Differentiating between two species is not possible on the basis of clinical, pathological, or immunological findings and their eggs are morphologically very similar **[26]**. Consequently, where both species occur concurrently or in overlapping geographical distribution, it is not possible to be certain as to which species is responsible for the disease. The low number of records of infection with *F. gigantica* may well be due to the lack of good tools to distinguish this species from *F. hepatica* **[27]**.

The comparison of ITS sequences from worms of different hosts and of different countries indicates that there exists a high species-specific homogeneity. In the present study, primary sequence analysis revealed a close relationship between the query sequence (from Northeastern region of India) and isolates of *F. gigantica* from China, Indonesia and Japan. Phylogenetic trees constructed showed that the groups of multiple closely related genotypes of *F. gigantica* from Asia are broadly sympatric. Such a pattern

318

is expected for species with high gene flow whose populations have not been sundered by long term biogeographic barriers.

We present here a new approach of molecular morphometrics, in which the measurable structural parameters of the molecules 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 **[28]**. Several patterns of predicted secondary structures of RNA were constructed from unique ITS sequences from different geographical isolates of *F. gigantica*, which provided us with the additional information for correct identification of the species prevalent in the region. Molecular morphometrics appears to be complimentary to classical primary sequence analysis in phylogenetic studies as it takes into consideration only the size variations of homologous structural segments and this choice implies that the overall architecture of the molecule remains same among the observed taxa. This method helps in taking into account the regions where multiple alignments are barely reliable because of large number of insertion/deletion operations. In the present study the secondary structure analysis of the same data also confirmed the results mentioned for primary sequence analysis. Differences in their topology are only due to the fact that there are variable lengths of the sequences. However, there are difficulties in defining a distance between two related structures with variable topologies **[29]**. Nevertheless, because there were inconsistencies in the placement of a few *Fasciola* species, this study needs to be extended, in order to gain a better understanding of the systematics of this group as well as the evolution of their predicted ITS2 RNA secondary structures.

Conclusion:
The present in silico identification of the *Fasciola* spp. with ITS sequence motifs and secondary structure is consistent with investigations made using traditional approaches (i.e., by morphology). The results also corroborate that the *Fasciola* species prevalent in Assam, Northeast India is in fact *F. gigantica* and not *F. hepatica*, which otherwise is the most common liver fluke throughout the globe. Lack of data on genotypic diversity of *Fasciola* species in Africa and India does not allow the origin of regional populations to be unambiguously determined. Further studies with additional molecular markers and bar coding will contribute towards

determining the population structure and divergence between the two closely related species of this genus.

References:
[1] D. Blair *et al., Adv Parasitol* 42: 113 (1999)
[2] R. D. Adlard *et al., Int J Parasitol* 23: 423 (1993)
[3] W. Y. Huang *et al., Vet Parasitol* 120:75 (2004)
[4] M. Holder & P.O. Lewis *Nat Rev Genet* 4: 275 (2003)
[5] S. Mas-Coma *et al., Bull WHO* 77: 340 (1999)
[6] J. A. Hammond *Trans R Soc Trop Med Hyg* 68: 253 (1974)
[7] S. Mas-Coma *et al., Int J Parasitol* 35:1255 (2005)
[8] G. Caetano-Anolles *Nucleic Acids Res* 30: 2575 (2002)
[9] B. Billoud *et al., Mol Biol Evol* 17: 1435 (2000)
[10] C. Zwieb *et al., Nucleic Acids Res* 9: 3621 (1981)
[11] J. Schultz *et al., RNA* 11: 361 (2005)
[12] A. Grajales *et al., BMC Evol Biol* 7: 90 (2007)
[13] M.C. Ebach & C. Holdrege *Nature* 434: 697 (2005)
[14] D. E. Schindel & S.E. Miller *Nature* 435: 17 (2005)
[15] J. Sambrooke *et al., Molecular cloning: a laboratory manual* (1989)
[16] P. K. Prasad *et al., Parasitol Res* 101:1581 (2007)
[17] J. Bowles *et al., Mol Phylogenet Evol* 4: 103 (1995)
[18] B.A. White *PCR Protocols, current methods and applications* (1993)
[19] J. D. Thompson *et al., Nucleic Acids Res* 22:4673 (1994)
[20] D. M. Hillis & J.J *Syst Biol* 42: 182 (1993)
[21] M. A. Larkin *et al., Bioinformation* 23: 2947 (2007)
[22] J. P. Huelsenbeck & F. Ronquist *Bioinformatics* 17: 754 (2001)
[23] R. D.M. Page *Comp Appl Biosci* 12: 357 (1996)
[24] T. A. Hall *Nucl Acids Symp* 41: 95 (1999)
[25] M. Zuker *Nucl Acids Res* 31: 3406 (2003)
[26] W. M. Lotfy & G.V. Hillyer *Exp Pathol Parasitol* 11:9 (2003)
[27] A. Marcilla *et al., Mol Cell Probes* 16:327 (2002)
[28] M. Zuker *Methods Mol Biol* 25: 267 (1994)
[29] B. A. Shapiro *Comput Appl Biosci* 4: 387 (1998)

Supplementary material

Table 1: Fasciolidae species and their various geographical isolates used in this study with the respective GenBank accession numbers for corresponding ITS2 sequences. A= Sequence generated as part of the present study.

Species	GenBank accession	Sequence length (bp)	Classification
Fasciola gigantica India (A)	EF027103	606	*Digenea: Fasciolidae*
Fasciola sp. Japan: Kochi	AB207152	505	*Digenea: Fasciolidae*
Fasciola gigantica Indonesia	AB010977	505	*Digenea: Fasciolidae*
Fasciola sp. Japan	AB010979	505	*Digenea: Fasciolidae*
Fasciola gigantica Zambia	AB010976	505	*Digenea: Fasciolidae*
Fasciola hepatica Uruguay	AB207148	506	*Digenea: Fasciolidae*
Fasciola gigantica Burkina Faso	AJ853848	588	*Digenea: Fasciolidae*
Fasciola gigantica China	AJ557569	361	*Digenea: Fasciolidae*
Fasciola gigantica Kenya	EF612484	364	*Digenea: Fasciolidae*
Fascioloides magna Austria	DQ683545	538	*Digenea: Fasciolidae*

Table 2: BLAST outputs of *Fasciola* ITS sequence motifs against NCBI GenBank database (nr db)

Fasciola species motif patterns	Length (bp)	No. of best Hits	Identity (%)	E-value
>Pattern 1 A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C	50	> than 100	100	$1e^{-17}$
>Pattern 2 G-A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T	50	> than 100	100	$2e^{-16}$
>Pattern 3 T-G-A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G	50	> than 100	100	$2e^{-16}$
>Pattern 4 G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T	50	> than 100	100	$1e^{-17}$
>Pattern 5 G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G	50	> than 100	100	$2e^{-16}$
>Pattern 6 C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G-C-A	50	> than 100	100	$3e^{-15}$
>Pattern 7 C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G-C	50	> than 100	100	$3e^{-15}$

Table 3: BLAST outputs of Fasciola ITS sequence motifs against NCBI GenBank database (nr db)

Fasciola species motif patterns	Length (bp)	No. of best Hits	Identity (%)	E-value
>Pattern 1 A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C	50	> than 100	100	$1e^{-17}$
>Pattern 2 G-A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T	50	> than 100	100	$2e^{-16}$
>Pattern 3 T-G-A-x-G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G	50	> than 100	100	$2e^{-16}$
>Pattern 4 G-T-G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T	50	> than 100	100	$1e^{-17}$
>Pattern 5 G-C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G	50	> than 100	100	$2e^{-16}$
>Pattern 6 C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G-C-A	50	> than 100	100	$3e^{-15}$
>Pattern 7 C-C-A-G-A-T-C-T-A-T-G-G-C-G-T-T-T-C-C-C-T-A-A-T-G-T-A-T-C-C-G-G-A-T-G-C-A-C-C-C-T-T-G-T-C-x(0,1)-T-G-G-C	50	> than 100	100	$3e^{-15}$