Exploring unique sequence repeat patterns and secondary structures in rDNA internal transcribed spacers ITS1 and ITS2 for characterization of catfish species

Authors

Mohd Imran[1]*, Sana Nafees[2]

Affiliations

[1]*Department of Zoology, Faculty of Life Sciences, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India.*

Email: mimran@myamu.ac.in

[2]*Department of Biosciences, Jamia Millia Islamia, New Delhi 110025, India.*

Email: sanabiosciences@gmail.com

Abstract

The genetic variabilities in internal transcribed spacers ITS1 and ITS2 of rDNA are explored to a limited extent in vertebrates including fish species. This study explores unique molecular signature in these spacers from multiple fish species from the order siluriformes. Supporting the concerted evolution, both ITS1 and ITS2 were found prevalent with microsatellite and random repeats, with diverse distribution in each species in term of nucleotide composition, position and length, giving a species-specific impression in repeats. Their secondary structures too have shown distinct structural homology with variations in helix length, loops, and non-canonical bps, where ITS2 4-helix carries additional nine conserved motifs while that of ITS1 was found highly branched with multiple sub-branching unique to each species and a long tail helix common in all. This combination of secondary structures and repeats have formed unique molecular signatures distinguishing each species. Since the evolutionary and maturation biology of rDNA is similar in vertebrates, these molecular signatures in internal spacers could effectively contribute in the molecular characterization of other species too.

Keywords: rDNA, ITS, microsatellite, secondary structure

INTRODUCTION

The primary structure of DNA, i.e. linear nucleotide sequence often reveals some striking features when investigated at their secondary structure level. Not only the coding segment but the non-coding DNA especially rDNA also keeps a highly complex structural composition when it comes to the primary and secondary structures of its internal spacers- Internal Transcribed Spacer region 1 (ITS1) and region 2 (ITS2). These ITS1 and ITS2 are located in between 17-18s rDNA and 5.8s rDNA, 5.8s rDNA and 25-28s rDNA, respectively; and hence share all its general features viz. multi-chromosomal origin, tandemly repeated presence (Reed et al. 2000) and concerted evolution (Hillis and Dixon 1991). The peculiar property associated with both the internal spacers is their variable size, which varies not just in between different species but also within the same species (Harris et al. 2000; Kuriiwa et al. 2007). They show this dynamicity in their size due to the prevalence of multiple microsatellites in their sequences. Besides, the multi-chromosomal location of NORs (Nucleolus Organizer Regions) is also suggested a probable factor behind (Ritland et al. 1993). These microsatellites are tandemly repeated sequences of short nucleotide sequence/motif ranges from 1 to 6 bp (Kelkar et al. 2010). Depending on the composition, they could be mononucleotide repeats, di-, tri-, tetra-, penta-, and hexanucleotide repeats. They are basically known for their unique feature, i.e. the variation in the number of tandem repeats, their number may vary between different species or even among individual of the same species. Because of all these associated variables, the ITS regions are found with wide differences in length among different organisms. For example, in loach, the ITS1 was reported just 343 bp while it was reported 1095 bp long in humans (Gonzalez et al. 1990). Similarly, such a wide difference is also observed in ITS2 which is reportedly just 363-369 bp long in *Anopheles nuneztovari* (fritz et al. 1994), 384-490 bp in catfishes (Jansen et al. 2006) and reaches to a length of 1300 bp in Crayfish (Harris et al. 2000).

Although both the ITS1 and ITS2 are non-coding markers, they still play a crucial role in the maturation of the precursor rRNA molecule, where they provide spatial secondary structure geometry crucial for the binding of snoRNA and RAC protein complex necessary for proper ribosomal maturation. These internal spacers carry margin to bear changes in their nucleotide composition and sequence length to the point of non-disruption of their secondary structure which appears as an evolutionary adaptation to conserve their functional aspect (Pleyte et al. 1992; Cote and Peculis 2001; Abeyrathne et al. 2002). The ITS2 also provides better information for taxonomic inferences when used along with its secondary structure (Coleman 2003). The same approach isused in this study for both the internal spacers ITS1 and ITS2.

The presence of extremely conserved rDNA segments at their 5' and 3' end of ITS1 and ITS2 makes them an easy candidate for PCR based study. They have been studied extensively for barcoding fungal species (Schoch et al. 2012; Kapoor et al. 2018), resolving biogeography (16) and species complex in plants (Ritland et al. 1993; Vander et al. 1998). Even within the animal kingdom they are explored more in the invertebrate, viz. phylogenetic relationship in Ciliates (Chen et al. 2021), corals (Aguilar and Reimer 2010), resolving cryptic speciation in nematodes (Pinacho-Pinacho et al. 2018), phylogeography of Mosquito (Fritz et al. 1994), species identification of Molluscs (Hurtado et al. 2011). Several studies have also been conducted in vertebrates for phylogenetic relationship and species identification (Voronov et al. 2006; Coleman 2013) including several reports from fish species as well (Booton et al. 1997; Domanica et al. 1997; Sajdak and Phillips 1997; Presa et al. 2002; Montoya-Burgos 2003; Vinas and Tudela 2009; Vera et al. 2011; Myslowski et al. 2013; Puncher et al. 2015) but still it is not widely explored in any vertebrate group even in pisces.

Since the structural organization of rDNA is conserved across the eukaryotes and the complex process of ribosomal maturations too, is similar (Greelings et al. 2000; Abeyrathne et al. 2002), five fish species from the order silurifomes, class pisces of the vertebrate, were

included in the study to trace the distribution of repeated sequences in the internal transcribed spacer regions ITS1 & ITS2 of rDNA, to predict their secondary structures, and to look for their combined potential in species diagnosis. The catfish species included were *Clarias batrachus* and *C. gariepinus* (family: Clariidae); *Heteropneustes fossilis* (family: Heteropneustidae); *Sperata seenghala* (family: Bagridae), and *Wallago attu* (family: Siluridae). Additionally, no prior study has explored the rDNA spacer structures in vertebrate species in this perspective.

MATERIALS AND METHODS

Sample and DNA Isolation

Multiple specimens of the fish species were collected from three distant locations on the northern riverine system of Ganges i.e., Garhwal (30°24'5.35"N, 78°45'46.11"E), Aligarh (27°58'41.57"N, 78° 8'17.96"E) and Varanasi (25°19′08″N 83°00′46″E). Specimens were identified using standard identification features (Talwar and Jhingran 1992). Fishes were anesthetized prior to blood extraction with MS222 anesthetic. Approximately 0.5 ml of the blood extracted from the heart and caudal vein in EDTA coated vial. Extraction of total genomic content, including both nuclear and mitochondrial genome, was performed according to the high salt method (Montgomery and Sise 1990).

Primer Designing

To deal with the questions under study, ITS1 and ITS2 were targeted separately. Since, no such primers were available that could specifically target these spacer regions in the aforesaid species, new sets of primers were designed from complete rDNA cistron of *Clarias gariepinus* (Genbank accession no. AJ876383.1) using NCBI primer designing tool

(https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (Ye et al. 2012) which include ITS1 forward (ITS.1.F1- TCGATCAAACTTGACTATCTAGAGG; ITS.1.F2- CGATCAAACTTGACTATCTAGAGG A); ITS1 reverse (ITS.1.R1- CAGCTGGCTGCGTTCTTCAT); ITS2 forward (ITS2.F- CAGGACACATTGATCATCGAC AC); ITS2 reverse (ITS2.R- AATTCAGCGGGTCGTCTC). The primers were found compatible with species from 159 genera, where 37 genera were found with compatibility in multiple sister species (Supplementary data S1).

PCR standardization, Gel electrophoresis and Sequencing

The PCR conditions were standardized for all five species using $MgCl_2$ range of 1- 3mM, annealing temperature (Ta) range of 46-56°C for 45-60 seconds depending upon the ITS marker. The amplification was done using peqSTAR 2X thermocycler (PEQLAB Biotechnologie, Germany) in 15µl PCR reaction mixture containing 50-100ng of DNA template, 1.5µl of 10X PCR reaction buffer, 1.5 to 3.5mM $MgCl_2$, 0.3µl of 10mM dNTPs mix (fermentas), 0.25µl each of 10pm forward and reverse primers, 0.25 unit of TaqDNA polymerase (fermentas). The PCR cycle involved an initial denaturation at 95°C for 5 min, followed by 35 cycles as (i) template denaturation at 95°C for 1min, (ii) primer annealing varied with species within a range of 55-62°C, and (iii) extension at 72°C for 1 min. A final extension was also performed at 72°C for 5 minutes at the end of 35 cycles.

The PCR products were subjected to electrophoresis on 1.5% agarose gel. Both the set of primers provided a very specific noise free amplification of ITS1 and ITS2 in all five species, with a clear indication of reproducible results as shown by the identical band positions among the multiple individuals of each species. The difference in the band positions in different species for both the spacers (ITS1 and ITS2) unquestionably depict that neither of them possesses size uniformity and thus have shown considerable size range (Fig. 1). Sanger

sequencing was performed using ABI 3730XL DNA Analyser with BigDyeTM Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific), and the sequences were submitted to NCBI GenBank.

Sequence Analysis

The ITS1 and ITS2 sequences from multiple replicates of all five species were aligned with Clustal W, and required editing viz. labelling was done in 'Mega7: Text File Editor and Format Converter' (Kumar et al. 2016). The repeated sequence tracing was done with MISA- a web based repeat finding tool (https://webblast.ipk-gatersleben.de/misa/) (Beier et al. 2017) and with manual tracing of sequences. The aligned sequences were then compared and analyzed within and between species for their relative differences, to find the overall pattern of repeats and unique microsatellites if any from the point of view of species identity.

Secondary Structure Analysis

For ITS2, the secondary structure analysis was done at ITS2 database. This includes annotation of ITS2 sequences, secondary structure prediction using homology modelling, multiple alignment of secondary structures, conserved motif of secondary structures (Koetschan et al. 2010). The ITS1 secondary structure prediction was done using RNAfold web server with minimum free energy method (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) (Gruber et al. 2008). The LocARNA web server (https://rna.informatik.uni-freiburg.de) (Will et al. 2007, 2012; Raden et al. 2018) was used for the multiple sequence alignment of ITS1 sequences on the basis of their secondary structure. The annotation of both ITS1 and ITS2 secondary structures was done in a similar manner using software- 4SALE (Seibel et al. 2006, 2008) and VARNA (Darty et al. 2009) (Supplementary data S2).

RESULTS

Description of ITS Repeats

The ITS1 and ITS2 sequences of each catfish species were found with the presence of multiple repeats. Some repeats qualify the standard definition of microsatellites (Oliveira et al. 2006) while many don't fit into that but effectively describe the repetitive nature of both ITS1 and ITS2 and placed in the category of the random repeats. The highly variable distribution of distinct microsatellites and random repeats are found affecting the lengths of both ITS1 and ITS2 in a considerable manner.

Microsatellite and Random repeats

The repeat distributions and patterns are found highly diverse in five species ranging from short to long; simple to complex. Repeats of every kind from mononucleotide to tetranucleotide; perfect, imperfect, interrupted and composite (Table 1 & 2, Supplementary data S3). The ITS1 and ITS2 both have shown microsatellite and random repeats except the *H. fossilis* ITS2 which consists of only random repeats. Not all five species were found prevalent with every type of repeat. A considerable degree of intraspecific similarity was observed in repeats, where each species was found containing its own set of repeats differing from that of other species in terms of nucleotide composition, position (locus) and length. From the secondary structural point of view, this distinct composition and distribution of repeats causes wide differences between species in terms of their location on the different helices of secondary structure, and thus providing a species specific architecture to secondary structures.

Table 1. ITS2 nucleotide repeats. Table showing the microsatellite and random repeats in the ITS2 of catfish *C. batrachus*, *C. gareipinus*, *H. fossilis*, *Sperata seenghala* and *W. attu*. The variants of the same microsatellite between different individuals of the same species are separated by (/).

Species	**Repeats**[a]	
C. batrachus	Microsatellite repeat	CTCTCTCTGCTGCCTCT
	Random repeat	GGGGAGACGAGGGAGGAGAGGGG
C. gariepinus	Microsatellite repeats	CCCCCCGCTCTCTCT
		GAGAGAGGGGGGG
		CCTTCTCCTCCTCCTCCACCTC
	Random repeats	GGTTCCCTTCC
		CGGCGGGCGGGGAGGCCGGGGG
		CCCTTCTCCCTTCCT
		GGAGAGGGGAGAA
H. fossilis	microsatellite repeats	-
	Random repeats	TCCTCCCAAA
		TTTTTCCCTTCCACCCC
		TCTCCCCTCC
		AAAACCCCCGGCCGGAAACGGTTCC
		GTCGTCGTGCGCGGCTGTCGGTGGGTCCCTTCC
		GGTGGGTGTGGT
		GGCGGCGGGG
		GGCCCGGGCCGCCGCGTGGCGCCCC
S. seenghala	Microsatellite repeats	CCCCTCTCTCTCTCCCTCCCC/ CCCCTCTCTCTCTCTCTCCCTCCCTCCCC/ CCCCTCTCTCTCTCTCTCTCTCCCCC
		AAAAAGAAA
		AAAAA/ AAAAAAA/ AAAAAA
		AAAAAAAAA/ AAAAAAAAAA
	Random repeats	CCACCCCCCGCCTCCGTCCTCCC
		AAAAAGGGAGGTGGGG/ AAAAAAAGGAGGTGGGG/ AAAAAAGGAGGTGGGG
W. attu	Microsatellite repeats	TTTTTTTTTTT
		AAAAAAGAAA
		GGGGCGGGGGG
		$(A)_6GAGA(A)_6(GA)_6CGGGGAGGGGGG$/ $(A)_6GAGA(A)_6(GA)_7CGGGGGGGGGGG$/ $(A)_6GAGA(A)_6(GA)_7CGGGGAGGGGGG$
		AAAAAAA

	Random repeats	AAAAAAGAAACCCCTCCC
		CCTTTCCCCCAAACC
		CCCTTTGGGGCGGGGGG
		TTCCCCGGAAACCCCGAGGGCGGGGAAGGG
		TTTCCTTCCTTT

[a] supplementary data S3

Table 2. ITS1 nucleotide repeats. Table showing the microsatellite and random repeats in the ITS1 of catfish *C. batrachus, C. gareipinus, H. fossilis, Sperata seenghala* and *W. attu*. The variants of the same microsatellite between different individuals of the same species are separated by (/).

Species	**Repeats[a]**	
C. batrachus	Microsatellite repeat	CCTCCTCCTCCACCTCTT
	Random repeat	AGAGAGCCCCGGGCTCTCT
C. gariepinus	Microsatellite repeats	GGGGGG
		CCTCTCCTCT/ CCTCCTCCTCCTTCTT
		CCCCCAGCCC/ CCCCCGGCCCCCC
		TTTTTTT
		$(GTGC)_3T(GGT)_2GCC(GGT)_2$
		TCCCTCTCTCTCCCTCTCT
		GCGCGCGCGCGCGC
		AAAAAAAACCAAAAA
	Random repeats	GGTGGGGCTGGCT
		CTTCTCCCTCGCCTCTTTCTTCC
		GGCGGGCGCGCGCCGGTCGGGCTGGGAGGG
		CCTCCCTTCTGTTCCACCCCC
		CCCTCCCCTAAGGTTT
		CCTCTCCTTCCT
		CCCCGCGCACGACCGCCCCCCGCCTCGCC
		CTCTCCCGCCCACCCC
		CGCGGGGGCGGG
		TTCCGGGGCGCCCTCCCCT
		GGTTTTTAACACACACTGTGTT
H. fossilis	Microsatellite repeats	GTGTGTGTGTGAGAGA
		CTCTCTCCTTCTCT
		CACACACGCGCGCGCG/ $CACACA(CG)_6$
		CCCCCC/ CCCCCCC
		TTTTTT
		AAAAAAA
	Random repeats	CCGCCGGGGGTGCGCCGCC
		GAGAGAGCCCC
		AAACCCTTCCGG
		GCGCCGCGGGTGGGGGCGCG

		CCGGGGCGCCCTCCCCGG
S. seenghala	Microsatellite repeats	AAAGAAAGAAAGAAAGAAAG/ AAAGAAAGAAAG/ AAAGAAAGAAAGAAAG
		GGGGGTAGGGGGG/ GGGGGTAGGGGG
		GACG/ GACGGACG/ GACGGACGGACGGACG
		$(CCTT)_3(CCCT)_3CTCTCCCTCT$/ $(CCTT)_4(CCCT)_4CTCTCCCTCT$/ $(CCTT)_2(CCCT)CGCT(CCCT)CGCT(CCCT)_2CT$
		CCCCTCCCCT
		GGGGTGGGG
		CCCCCACCCCGACCCCC/ CCCCCACCCCAACCCCC/ CCCCCACCCCACACCCCC
		TCTCTCTCGTCTCCCCC
		TTTTTTT
		AAAAAAA/ AAAAAAAA
	Random repeats	TCTTTCCCC/TTTTTTCCCC
		CCCTTTT
		GGGGGAAAAAGAGAAAGGGGACGAGAAAAGGGAA
		TCTTTTCTTT
W. attu	Microsatellite repeats	$A_6(GAT)A_7GA_4(CGAGG)A_{11}GA_6GA_3GA_5$/ $A_6(GAT)A_6GA_4(CGAGGG)A_{26}$
		CCCCCCC CCCCCCCCCCCC
		AAAACAAAAA
		$(T)_7C(T)_6GAC(T)_9CTTTCTTA(TTC)_4TCTCCCCTTC(T)_6G(T)_4$
		AAAACAAAACCAAAAAAAAAAAA
	Random repeats	CCCGGGGG
		AAAAACCTTTGGG
		GGGTACCTACCGGCCGG
		TTTTTATTTCCTCC
		CCCCGGCCGACCCGGGGCGCCCTT
		TATATTTTGTTGTTTCTCTTT

[a] supplementary data S3

Secondary Structure of ITS2

Structure prediction

All the species were found possessing similar architecture of secondary structure with all consist of 4 helices branching off from the central ring which is one of the hallmarks of eukaryotic ITS2 structure with helix III is the longest among the four. The central ring is not usually a perfect circular ring rather quite variable in shape which continues into a stalk through which the ITS2 structure remains attached to the parent rRNA gene transcript. Helix I, II, IV and ring stalk are closely situated at one side of the central ring while the helix III is oriented in a direction opposite to the rest (Fig. 2 & 3, Supplementary data S4).

Motif search

The eukaryotic ITS2 possess some specific nucleotide features known as hallmark of the eukaryotic ITS2 sequences and its secondary structure as they are observed ubiquitously in different eukaryotic species. The motif search using ITS2 database server has found these hallmark features in the ITS2 sequences of all five catfish species, too, which include - i) A conserved UG motif in different variants of UGGU/UGGGU/UGGGG string located in the helix III of the secondary structure near its terminal loop (Fig. 4). ii) Pyrimidine-Pyrimidine mismatch in the helix II of the ITS2 secondary structure which consists of placement of a pyrimidine base against a pyrimidine base instead of a purine base. It is found to be cytosine-uracil in *C. batrachus, C. gareipinus, H. fossilis, W. attu*, and cytosine-cytosine in *S. seenghala*. It leads to a slight bulge in helix II of the secondary structure because of the absence of any bonding between the bases of the same group (Fig. 3). The exact position of these motifs varies from species to species depending upon the length of the ITS2 which ranges over a considerable size as previously described. The deviations from these hallmarks of secondary structure are also reported (48).

In addition to the eukaryotic conserved motifs, the secondary structure based multiple sequence alignment of the ITS2 have shown the presence of nine motifs which are conserved between these catfish species (Fig. 4).

Loops

The helices are interrupted with loops of different sizes at multiple positions which are called as interior loops. Along with intermittent loops all four helices ending into terminal loops as well (Fig. 3). The helix III possesses the maximum number of loops while the helix I possess the lowest number of loops; the position, number and the pattern of the loops in helices varied with species and together showing a species specific pattern. The four helices have shown a considerably wider range of loops between different species (Supplementary Table S1).

Non-canonical base-pairs

Non-canonical base pairing goes against the standard Watson & Crick base pairing pattern of DNA and RNA, where Adenine(A) could bind with Cytosine(C), and Guanine(G) could bind with Uracil(U) (Fig. 3). All the species have a shown a distinct pattern of non-canonical base pairing where only one type of base pairing i.e. G-U is found in all the secondary structure with only a single adenine with cytosine (A-C) in the helix IV of all the individuals of *Wallago attu*. All four helices in all the species are found prevalent with only G-U that appeared as a very common non-canonical base pair in ITS2 secondary structure. It ranged from a lowest of 05 to a highest of 15 G-U bp in *H. fossilis* and *C. gariepinus*, respectively (Supplementary Table S2).

Length of helices

The ITS2 secondary structure possessed considerable differences among its four helices which was clearly evident in all the species (Supplementary Table S3, Fig. 3). The helix III

was found the longest helix in all species while the helix I was the shortest one with the exception of *S. seenghala* where helix IV was the shortest. Besides this huge difference between different helices within the same species, each helix was found with notable length range between different species. The helix I varied over a range of 06 nucleotides in *W. attu* to 33 nucleotides in *C. batrachus*, Helix II ranged from 44 nucleotides in *S. seenghala* to 105 nucleotides again in *C. batrachus*, the helix III was found varied over a comparatively higher range from 114 to 288 nucleotides in *C. batrachus* and *C. gariepinus*, respectively, and the helix IV also varied over a wider range with merely 13 nucleotides in *S. seenghala* to a count of 137 nucleotides in *C. gariepinus*. These wider length differences are because ITS2 was found to vary over a wider nucleotide range as previously discussed. Despite these major differences, no species has shown any deviation from the general attribute of ITS2 secondary structure, thus emphasizing the general structural conservation as reported in other eukaryotes (Schultz et al. 2005). Helix branching is also reported in several species in eukaryotes, and here among catfishes, single branching was found in helix II of *C. batrachus*.

Secondary Structure of ITS1

The secondary structure of ITS1 in five catfish species has shown a structural analogy with that of ITS2 with a resemblance to the overall contour but highly distinct structural attributes. All the secondary structures, in general, possesses a long helix similar in appearance to helix III of ITS2 and rest all helices are gathered very near to each other on the opposite end of the structure in a more or less similar manner to the secondary structure of ITS2 (Fig. 5), although the origin of ITS1 structure from the parent rRNA gene transcript was appeared different from that of ITS2 where the 5’ and 3’ points of attachment were located on one side of the structure unlike the presence of central ring and stalk in ITS2 (Fig. 6). Additionally, unlike ITS2, no eukaryotic structural hallmarks have been reported in ITS1. The helix branching is very extensive in ITS1, and the number of helices were also found to be quite variable between the

five species, unlike the fixed number of helices in ITS2. Keeping in view the high degree of variability in ITS1 secondary structures, and no previous related literature on catfish species, the structure has been divided into two parts for ease of understanding, i.e. upper branched region which is attached directly to the parent rRNA gene transcript and the lower tailed region which consist of long outward helix. The pattern of helix branching having major differences between species has given a unique secondary structure to the ITS1 of each catfish species. Additionally, unlike ITS2, no conserved sequence motif was found in multiple sequence alignment of ITS1 secondary structures. (supplementary data S4).

Loops and Non-canonical basepairs

The loops and non-canonical bps in ITS1 have shown a trend similar to that of ITS2 with a species-specific pattern. The internal and terminal loops dispersed with variable positions, numbers and patterns in different species. The non-canonical G-U basepair in the only form abundant in ITS1 with different distribution pattern in each species. The minimum 12 G-U bp and maximum 21 G-U bp were present in *W. attu* and *C. gariepinus*, respectively. (supplementary Table S4).

DISCUSSION

The structural attributes associated with the rDNA internal spacers found exclusive to the point that they provided visible distinction between them in different species in terms of their repeat distribution and secondary structures which together formed a unique rDNA signature for each species. At the primary level, their nucleotide sequences found abundant with the one highly variable attribute i.e. repeated sequences. The repeats are also well reported from diverse taxa including corals (Grajales et al. 2007), nematodes (Conole et al. 2001), beetles (Gomez-Zurita et al. 2000), and fishes (Sajdak and Phillips 1997; Chow et al. 2006). But there was no previous report of repeat distribution as observed in these catfish species, even

the type and distribution of nucleotide repeats was similar between multiple individuals of each species in terms of no. of repeats and position of repeats, and forming a pattern similar in trend but unique in each species in both the internal spacers. The presence of diverse combination of microsatellite and random repeats in spacer sequences have shown their diversity of sequence motifs within the same nucleotide string which possess the loci of different sequence combination that could be as perfect as the mononucleotide repeat of thymine (T_{12}) to the highly complex microsatellite repeat as $A_6(GA)_2A_6(GA)_7CG_{11}$ in *W. attu* along with strings of complex random repeats such as $T_2C_4G_2A_3C_4GAG_3CG_4A_2G_3$ which is without any basic repeating unit like microsatellite repeats. Although exception was found in *H. fossilis* ITS2 which was abundant with only randomly repeated sequences but that too was conserved within species.

The possible factor behind this uniformity in variability in the multi-chromosomal rRNA gene is the pattern of its evolution i.e. the concerted evolution, which involves homogenization of ribosomal gene from different chromosomes (Elder and Turner 1995), preventing intragenomic ITS variant due to multichromosomal origin, allowing only the accumulation of interspecific difference (Hillis and Dixon 1991). Although the concerted evolution is reported as the general trend in eukaryotes emphasizing on the complete homogenization process, it is not universal, several cases of incomplete and non-concerted evolution are also reported reflecting towards the incomplete homogenization of ribosomal gene (Eickbush and Eickbush 2007). In evolutionary perspective, the rDNA shows a range of non-concerted, incomplete concerted to concerted pattern of evolution depending upon the execution of the process of homogenization which is influenced by the genetic and epigenetic factors (Wang et al. 2023). Since we got only single band in multiple individuals of each species, it shows the absence of intragenomic length variation which in turn refers to the complete homogenization necessary for concerted evolution, thus emphasizing on the

presence of concerted mode of evolution of rDNA in catfish species. But as already said that homogenization process is influenced by several factor, so at this point the wide compatibility of the primers designed in this study with 159 genera of fish species further adds to the prospect of exploring these dynamic rDNA internal spacer for the better evolutionary understanding of non-coding rDNA not just in fish species but for the vertebrates in general.

Although both ITS1 and ITS2 spacers showed a similar trend in their intra- and interspecific variabilities, both the spacers found highly diverse to each other. This diversity of ITS concealed in the nucleotide sequences was impossible to visualize without secondary structures which defined their unique sequences composition hidden in the plain sight among the sequence repeats. The spatial geometry of secondary structures and its functional relevance associated with evolutionary conserved structural attributes (Pleyte et al. 1992; Cote and Peculis 2001; Abeyrathne et al. 2002) appears as the factor responsible for the degree of uniformity observed in both the internal spacers between different catfish species. But within the limits of uniformity both ITS1 and ITS2 possess considerable differences between all five species in terms of different secondary structure attributes viz. helix length, no. of loops, non-canonical bp, distribution of repeats and conserved motif on helices which together contributed distinguished contour to each species, thus provided a relatively good amount of distinguishing information for each species both at their primary and secondary structure levels.

Out of the two spacers, ITS2 has been well reported for interspecies differentiation in eukaryotes and recommended for better taxonomic inferences when used along with its 4-helix secondary structure which is evolutionary conserved across the eukaryotes (Mai and Coleman 1997; Joseph et al. 1999; Chen et al. 2004; Young and Coleman 2004; Schultz et al. 2005; Schultz and Wolf 2009), hence found conserved between these catfish species as well. It allows the proper multiple alignment between species even when the sequences are highly

diverged and could not be aligned at primary structural level (13), which is clearly observed in this study where the multi-species alignment of ITS2 secondary structures showed the presence of nine more conserved sequence motifs between five catfish species in addition to the standard eukaryotic conserved motifs at the same time when their primary structure failed to align between species. Moreover, within this conserved structural contour each species carries its unique attributes viz. helix I was longest in *C. batrachus* with branched helix II, *C. gariepinus* was found consisting of helix IV with consecutive looping, *H. fossilis* with a pattern in helix I and IV similar to that of *C. gariepinus* but unlike other species it has a unique pattern of three consecutive loop with just four bonded pairs. Unique to *S. seenghala,* the central ring which was not a perfect ring structure even in other catfish species was also present as bi-lobed structure with a comparatively long stalk, and also unlike other species helix IV was shortest in *S. seenghala*. Likewise, the unique structural characteristic in *W. attu* was the excessive looping specially in the helix III, near to the central ring. Hence this diversity in uniformity is the key to successful utilization of ITS2 for the differentiation of catfish species.

The structural uniformity was also traced in the secondary structure of ITS1, although contrary to ITS2, absence of any standard eukaryotic ITS1 secondary structure makes it more diverse which was quite apparently having no interspecific conserved region between five catfish species. But at the same time structural conservation was well obvious in ITS1 as well with upper branched region and a long tail helix, though the branching is diverged so much between species that it formed a unique pattern of helices and sub-helices in each species, thus providing species diversity among the structural uniformity in ITS1 on a pattern similar to ITS2.

Conclusively, the diverse structures of both the internal spacers have differentiated all five catfish species and also hold the potential to contribute in the molecular genetics and

systematics of not only catfish species but other fish species in general, and could also contribute for the other vertebrate species as they share similar process of ribosomal maturation where these ITS regions play a very crucial role by offering their secondary structures as potential binding sites for several maturation factors such as snoRNA and RAC protein. The functional dynamics of diverse repeat patterns in internal spacers in the conserved maturation process is a thing to be explored further.

Although there are several reports of ITS from fish species, but overall, there are limited studies in vertebrates from pisces to mammals, and this is the first time such a detailed description has been given in any vertebrate species about its composition of short sequence repeats. Such a dynamic sequence composition in internal spacers reflect towards their possible sequence diversity in the vertebrate groups. Since the pisces are the earliest in vertebrate evolution followed by amphibia, reptilia, aves and mammalia, the comparative repeats distributions among the higher vertebrates could give new insights into the evolution of rDNA as well as non-coding DNA in general which is known by the abundance of repeated sequences. Although the methodology is quite very extensive, the picture which come out in terms of repeat distribution and secondary structure makes ITS a very promising DNA segment which hold the key to species delimitation and discovery in vertebrates.

Accession numbers

"ITS1 GenBank accession nos. MT105349, MT105350, MT105351, MT105352, MT105353, MT105354, MT105355, MT105356, MT105357, MT105358, MT105359, MT105360, MT105361, MT105362, MT105363, MT105364, MT105365.

ITS2 GenBank accession nos. MT234369, MT234372, MT234373, MT234374, MT234376, MT234377, MT234378, MT234380, MT234381, MT234382, MT234383, MT234384, MT234385, MT234386, MT234387.

Authors Contribution

Mohd Imran: Conceptualization; Methodology; Formal analysis; Funding acquisition; Writing -original draft; review & editing. **Sana Nafees:** Conceptualization; Methodology, Formal analysis; review & editing.

Declaration of Interest

The authors declare no competing Interest.

Acknowledgments

The Authors are especially grateful to Prof. Mohammad Afzal for his constructive insights into this work. The authors also acknowledge the Department of Zoology, Aligarh Muslim University, for providing necessary laboratory facilities. The Council of Scientific and Industrial Research (CSIR), New Delhi, India is also gratefully acknowledged for providing funding support (09/112(0603)/2018-EMR-I).

REFERENCES

Abeyrathne PD, Lalev AI, Nazar RN (2002) A RAC Protein-binding Site in the Internal Transcribed Spacer 2 of Pre-rRNA Transcripts from Schizosaccharomyces pombe. Journal of Biological Chemistry 277:21291-21299. https://doi.org/10.1074/jbc.M201751200

Aguilar C, Reimer JD (2010) Molecular phylogenetic hypotheses of *Zoanthus* species (Anthozoa: Hexacorallia) using RNA secondary structure of the internal transcribed spacer 2 (ITS2). Mar Biodiv 40:195–204. https://doi.org/10.1007/s12526-010-0043-2

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Booton GC, Kaufman L, Chandler M, Oguto-Ohwayo R, Duan W, Fuerst PA (1999) Evolution of the Ribosomal RNA Internal Transcribed Spacer One (ITS-1) in Cichlid Fishes of the Lake Victoria Region. Mol Phylogenet Evol 11:273–282. https://doi.org/10.1006/mpev.1998.0576.

Chen CA, Chang CC, Wei NV, Chen CH, Lein YT, Lin HE, Dai CF, Wallace CC (2004) Secondary structure and phylogenetic utility of the ribosomal internal transcribed spacer 2 (ITS2) in scleractinian corals. Zool Stud 43:759–771

Chen X, Li J, Xu K (2021) Insights into the phylogeny of three systematically controversial subfamilies of urostylid ciliates based on rDNA. Zool Scr 50:383-395. https://doi.org/10.1111/zsc.12473

Chow S, Nakagawa T, Suzuki N, Takeyama H, Matsunaga T (2006) Phylogenetic relationships among *Thunnus* species inferred from rDNA ITS1 sequence. J Fish Biol 68:24-35. https://doi.org/10.1111/j.0022-1112.2006.00945.x

Coleman AW (2003) ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet 19:370-375. https://doi.org/10.1016/S0168-9525(03)00118-5

Coleman AW (2013) Analysis of Mammalian rDNA Internal Transcribed Spacers. PLoS ONE 8(11):e79122. https://doi.org/10.1371/journal.pone.0079122

Conole JC, Chilton NB, Jarvis T, Gasser RB (2001) Mutation scanning analysis of microsatellite variability in the second internal transcribed spacer (precursor ribosomal RNA) for three species of *Metastrongylus* (Strongylida: Metastrongyloidea). Parasitology 122:195-206. https://doi.org/10.1017/s0031182001007223

Côté CA, Peculis BA (2001) Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. Nucleic Acids Res 29:2106-2116. https://doi.org/10.1093/nar/29.10.2106

Darty K, Denise A, Ponty Y (2009) VARNA Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25(15):1974–1975. https://doi.org/10.1093/bioinformatics/btp250

Domanico MJ, Phillips RB, Oakley TH (1997) Phylogenetic analysis of Pacific salmon (genus *Oncorhynchus*) using nuclear and mitochondrial DNA sequences. Can J Fish Aquat Sci 54:1865-1872. https://doi.org/10.1139/cjfas-54-8-1865.

Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175:477-485. https://doi.org/10.1534/genetics.107.071399

Elder JFJr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70 297-320. https://doi.org/10.1086/419073

Fritz GN, Conn J, Cockburn A, Seawright J (1994) Sequence analysis of the ribosomal DNA internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera Culicidae). Mol Biol Evol 11:406-416. https://doi.org/10.1093/oxfordjournals.molbev.a040122.

Geerlings TH, Vos JC, Raué HA (2000) The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. RNA 6:1698-1703. https://doi.org/10.1017/s1355838200001540

Gómez-Zurita J, Juan C, Petitpierre E (2000) Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). Insect Molecular Biology 9:591-604. https://doi.org/10.1046/j.1365-2583.2000.00223.x

Gonzalez IL, Sylvester JE, Smith TF, Stambolian D, Schmickel RD (1990) Ribosomal RNA gene sequences and hominoid phylogeny. Mol Biol Evol 7:203–219. https://doi.org/10.1093/oxfordjournals.molbev.a040600.

Grajales A, Aguilar C, Sánchez JA (2007) Phylogenetic reconstruction using secondary structures of Internal Transcribed Spacer 2 (ITS2 rDNA) finding the molecular and morphological gap in Caribbean gorgonian corals. BMC Evol Biol 7:90. https://doi.org/10.1186/1471-2148-7-90

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA websuite. Nucleic Acids Res 36:W70-W74. https://doi.org/10.1093/nar/gkn188

Harris DJ, Crandall KA (2000) Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies. Mol Biol Evol 17:284-291. https://doi.org/10.1093/oxfordjournals.molbev.a026308

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Q Rev Biol 66:411-453. https://doi.org/10.1086/417338

Hlinka O, Murrell A, Barker S (2002) Evolution of the secondary structure of the rRNA internal transcribed spacer 2 (ITS2) in hard ticks (Ixodidae Arthropoda). Heredity 88:275-279. https://doi.org/10.1038/sj.hdy.6800040

Hurtado NS, Pérez-García C, Morán P, Pasantes JJ (2011) Genetic and cytological evidence of hybridization between native *Ruditapes decussatus* and introduced *Ruditapes philippinarum* (Mollusca Bivalvia Veneridae) in NW Spain. Aquaculture 311:123-128. https://doi.org/10.1016/j.aquaculture.2010.12.015

Jansen G, Devaere S, Weekers PHH, Adriaens D (2006) Phylogenetic relationships and divergence time estimate of African anguilliform catfish (Siluriformes: Clariidae) inferred from ribosomal gene and spacer sequences. Mol Phylogenet Evol 38:65-78. https://doi.org/10.1016/j.ympev.2005.09.011

Joseph N, Krauskopf E, Vera MI, Michot B (1999) Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. Nucleic Acids Res 27:4533-4540. https://doi.org/10.1093/nar/27.23.4533

Kapoor N, Gambhir L, Saxena S (2018) Secondary structure prediction of ITS rRNA region and molecular phylogeny an integrated approach for the precise speciation of *Muscodor* species. Ann Microbiol 68:763–772. https://doi.org/10.1007/s13213-018-1381-8

Kelkar YD, Strubczewski N, Hile SE, Chiaromonte F, Eckert KA, Makova KD (2010) What is a microsatellite A computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biol Evol 2:620-635. https://doi.org/10.1093/gbe/evq046.

Koetschan C, Förster F, Keller A, Schleicher T, Ruderisch B, Schwarz R, Müller T, Wolf M, Schultz J (2010) The ITS2 Database III - sequences and structures for phylogeny. Nucleic Acids Res 38:D275– D279. https://doi.org/10.1093/nar/gkp966

Kumar S, Stecher G, Tamura K (2016) MEGA7 Molecular evolutionary genetics analysis version 70 for bigger datasets. Mol Biol Evol 33:1870-1874. https://doi.org/10.1093/molbev/msw054

Kuriiwa K, Hanzawa N, Yoshino T, Kimura S, Nishida M (2007) Phylogenetic relationships and natural hybridization in rabbitfishes (Teleostei: Siganidae) inferred from mitochondrial and nuclear DNA analyses. Mol Phylogenet Evol 45:69–80. https://doi.org/10.1016/j.ympev.2007.04.018.

Mai JC, Coleman AW (1997) The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. J Mol Evol 44:258-271. https://doi.org/10.1007/PL00006143

Montgomery GW, Sise JA (1990) Extraction of DNA from sheep white blood cells. New Zealand Journal of Agricultural Research 33:437-441. https://doi.org/10.1080/00288233.1990.10428440

Montoya-Burgos JI (2003) Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae) with implications on the diversification of Neotropical ichthyofauna Mol Ecol 12:1855–1867. https://doi.org/10.1046/j.1365-294X.2003.01857.x

Mysłowski B, Panicz R, Sadowski J, Hofsoe P (2013) Genetic structure of the whitefish (*Coregonus lavaretus*) population inhabiting the Miedwie Lake Poland based on partial ND-1 and ITS-1 gene sequences Aquac Int. 21:739-750. https://doi.org/10.1007/s10499-012-9575-8

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira MLC (2006) Origin evolution and genome distribution of microsatellites. Genet Mol Biol 29:294–307. https://doi.org/10.1590/S1415-47572006000200018

Pinacho-Pinacho CD, García-Varela M, Sereno-Uribe AL, Pérez-Ponce de León G (2018) A hyper-diverse genus of acanthocephalans revealed by tree-based and non-tree-based species delimitation methods: Ten cryptic species of *Neoechinorhynchus* in Middle American freshwater fishes. Mol Phylogenet Evol 127:30–45. https://doi.org/10.1016/j.ympev.2018.05.023

Pleyte KA, Duncan SD, Phillips RB (1992) Evolutionary relationships of the salmonid fish genus *Salvelinus* inferred from DNA sequences of the first internal transcribed spacer (ITS1) of ribosomal DNA. Mol Phylogenet Evol 1:223–230. https://doi.org/10.1016/1055-7903(92)90019-D

Presa P, Pardo BG, Martínez P, Bernatchez L (2002) Phylogeographic congruence between mtDNA and rDNA its markers in brown trout. Mol Biol Evol 19:2161–2175. https://doi.org/10.1093/oxfordjournals.molbev.a004041

Puncher GN, Arrizabalaga H, Alemany F, Cariani A, Oray IK, Karakulak FS, Basilone G, Cuttitta A, Mazzola S, Tinti F (2015) Molecular identification of Atlantic bluefin tuna (*Thunnus thynnus*, Scombridae) larvae and development of a DNA character-based identification key for Mediterranean scombrids. PLoS One 10:e0130407. https://doi.org/10.1371/journal.pone.0130407

Raden M, Ali SM, Alkhnbashi OS, Busch A, Costa F, Davis JA, Eggenhofer F, Gelhausen R, Georg J, Heyne S et al (2018) Freiburg RNA tools: a central online resource for RNA-focused research and teaching. Nucleic Acids Res 2:W25-W29. https://doi.org/10.1093/nar/gky329

Reed KM, Hackett JD, Phillips RB (2000) Comparative analysis of intra-individual and inter-species DNA sequence variation in salmonid ribosomal DNA cistrons. Gene 249:115-125. https://doi.org/10.1016/S0378-1119(00)00156-6.

Ritland CE, Ritland K, Straus NA (1993) Variation in the ribosomal internal transcribed spacers (ITS1 and ITS2) among eight taxa of the *Mimulus guttatus* species complex. Mol Biol Evol 10:1273–1288. https://doi.org/10.1093/oxfordjournals.molbev.a040067.

Sajdak SL, Phillips RB (1997) Phylogenetic relationships among Coregonus species inferred from the DNA sequence of the first internal transcribed spacer (ITS1) of ribosomal DNA. Canadian Journal of Fisheries and Aquatic Sciences 54:1494-1503. https://doi.org/10.1139/cjfas-54-7-1494.

Schoch CL, Seifert KA ,Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W et al (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109 6241–6246. https://doi.org/10.1073/pnas.1117018109.

Schultz J, Maisel S, Gerlach D, Müller T, Wolf M (2005) A common core of secondary structure of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. RNA 11:361-364. https://doi.org/10.1261/rna.7204505

Schultz J, Wolf M (2009) ITS2 sequence–structure analysis in phylogenetics A how-to manual for molecular systematic. Mol Phylogenet Evol 52:520-523. https://doi.org/10.1016/j.ympev.2009.01.008

Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M (2006) 4SALE - A tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics 7:498. https://doi.org/10.1186/1471-2105-7-498

Seibel PN, Müller T, Dandekar T, Wolf M (2008) Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Res Notes 1:91. https://doi.org/10.1186/1756-0500-1-91

Talwar PK, Jhingran AG (1991) Inland fishes of India and adjacent countries. Volume 2. A.A. Balkema, Rotterdam

Tremetsberger K, Weiss-Schneeweiss H, Stuessy T, Samuel R, Kadlec G, Ortiz MA, Talavera S (2005) Nuclear ribosomal DNA and karyotypes indicate a NW African origin of South

American *Hypochaeris* (Asteraceae Cichorieae). Mol Phylogenet Evol 35:102–116. https://doi.org/10.1016/j.ympev.2004.12.022.

Vander Stappen J, Van Campenhout S, Gama Lopez S, Volckaert G (1998) Sequencing of the internal transcribed spacer region ITS1 as a molecular tool detecting variation in the *Stylosanthes guianensis* species complex. Theor Appl Genet 96:869–877. https://doi.org/10.1007/s001220050814.

Vera M, Sourinejad I, Bouza C, Vilas R, Pino-Querido A, Kalbassi MR, Martínez P (2011) Phylogeography genetic structure and conservation of the endangered Caspian brown trout Salmo trutta caspius (Kessler 1877) from Iran. Hydrobiologia 664:51–67. https://doi.org/10.1007/s10750-010-0581-4

Vinas J, Tudela S (2009) A validated methodology for genetic identification of Tuna species (Genus *Thunnus*). PLoS One 4:e7606. https://doiorg/101371/journalpone0007606

Voronov AS, Shibalev DV, Ryskov AP, Kupriyanova NS (2006) Evolutionary divergence of ribosomal internal transcribed spacer 2 in lizards. Mol Biol 40:37-42. https://doi.org/10.1134/S0026893306010079

Wang W, Zhang X, Garcia S, Leitch AR, Kovařík A (2023) Intragenomic rDNA variation - the product of concerted evolution mutation or something in between? Heredity 131:179–188. https://doi.org/101038/s41437-023-00634-5

Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R (2012) LocARNA-P accurate boundary prediction and improved detection of structural RNAs. RNA 18:900-914. https://doi.org/doi:10.1261/rna.029041.111

Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R (2007) Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLOS Computational Biology 3:e65. https://doi.org/10.1371/journal.pcbi.0030065

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13:134. https://doi.org/10.1186/1471-2105-13-134

Young I, Coleman AW (2004) The advantages of the ITS2 region of the nuclear rDNA cistron for analysis of phylogenetic relationships of insects: a Drosophila example. Mol Phylogenet Evol 30:236–242. https://doi.org/10.1016/S1055-7903(03)00178-7

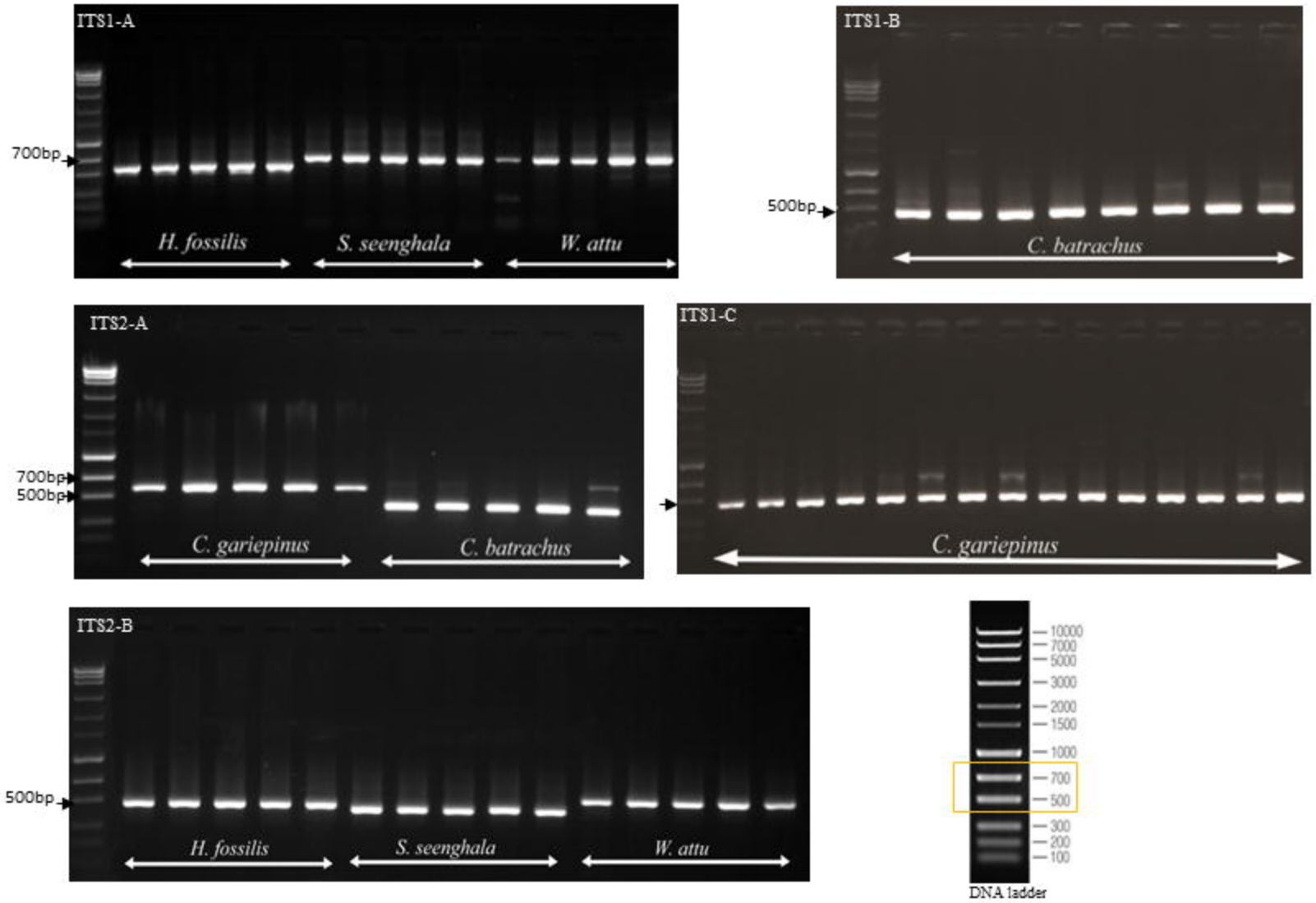

Figure 1. Agarose gel images of ITS1 and ITS2 amplicons. The ➤ indicate the amplicon size of ITS1 and ITS2 against the standard DNA ladder in the range of 500-700bp.

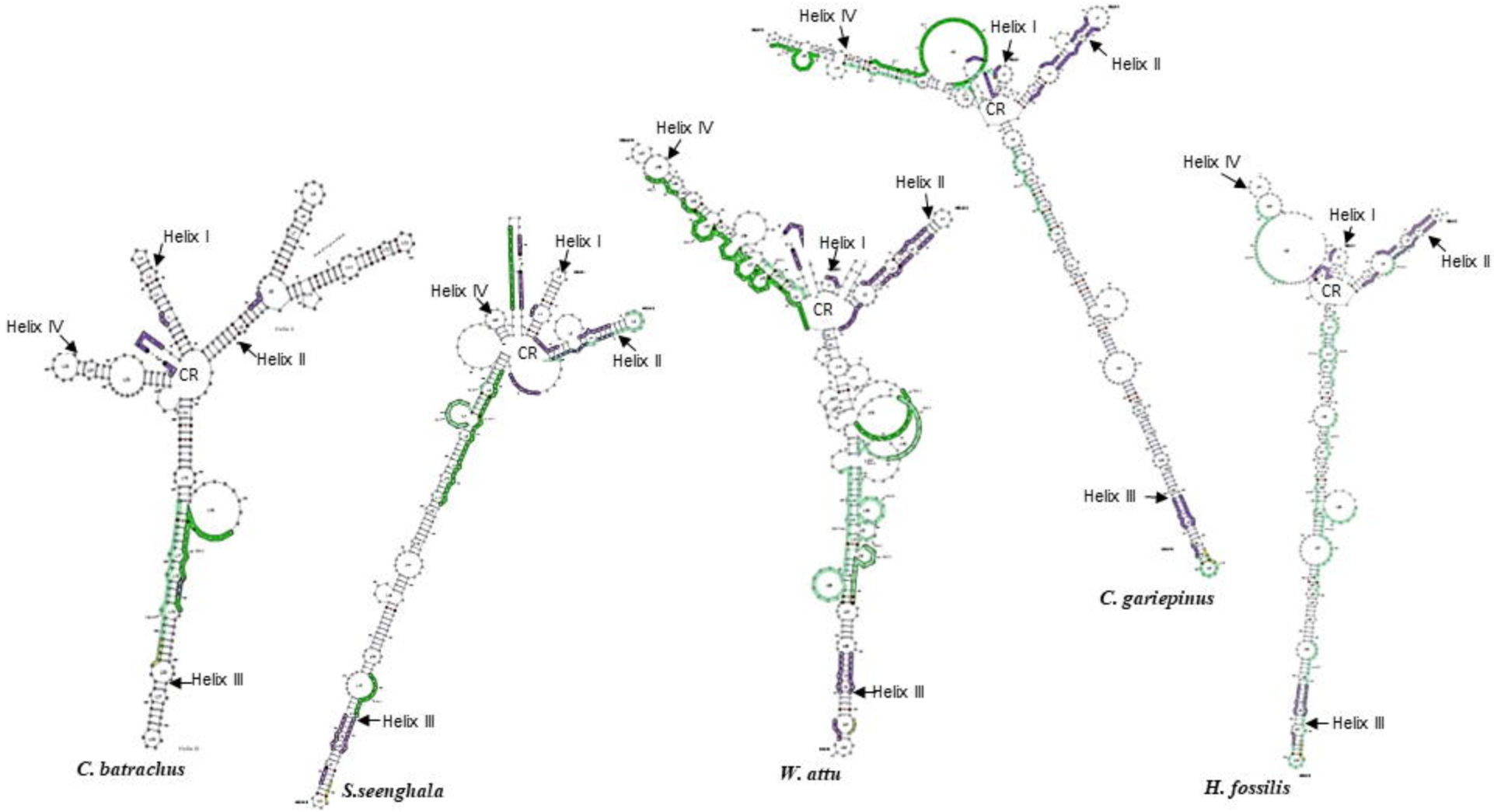

Figure 2. Secondary structure of ITS2. The figure showing the 4-helix secondary structure of ITS2 in *C. batrachus, S.seenghala, W. attu, C. gariepinus and H. fossilis.*, with a clockwise arrangemet of helix I to IV round a central ring (CR).

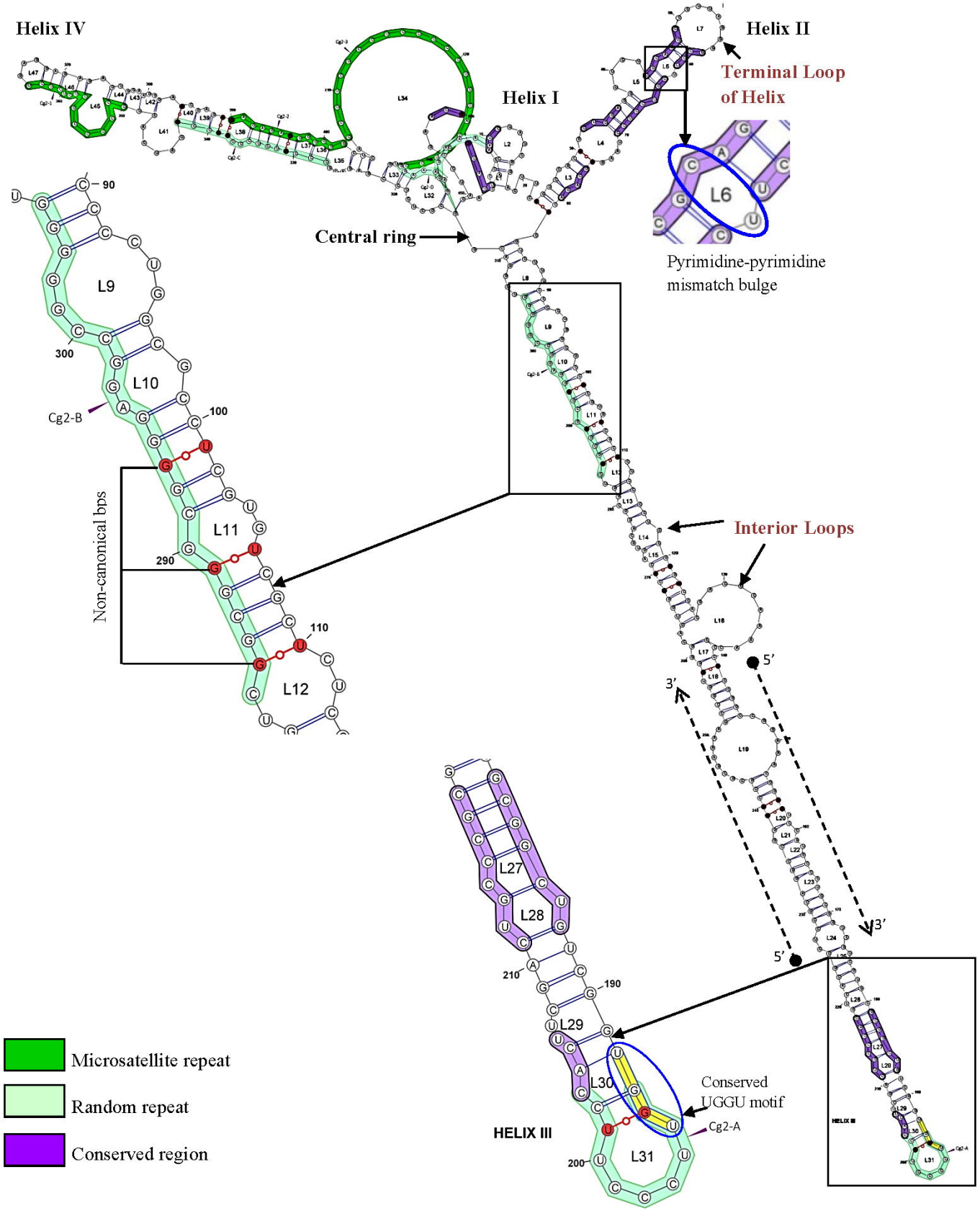

Figure 3: Structural organisation of ITS2 secondary structure (*Clarias gariepinus*). The 4-helix secondary structure showing the central ring with its four helices from Helix I to Helix IV projecting outward in clockwise direction. The helices are interrupted with internal loops of variable size with each helix ends into a terminal loop. The annotation was done with java-based software VARNA (Visualization Applet for RNA). Microsatellite repeats, Random repeats, eukaryotic conserved motifs, catfish conserved motifs, non-canonical bps and pyrimidine-pyrimidine mismatch are shown in magnified view round the secondary structure.

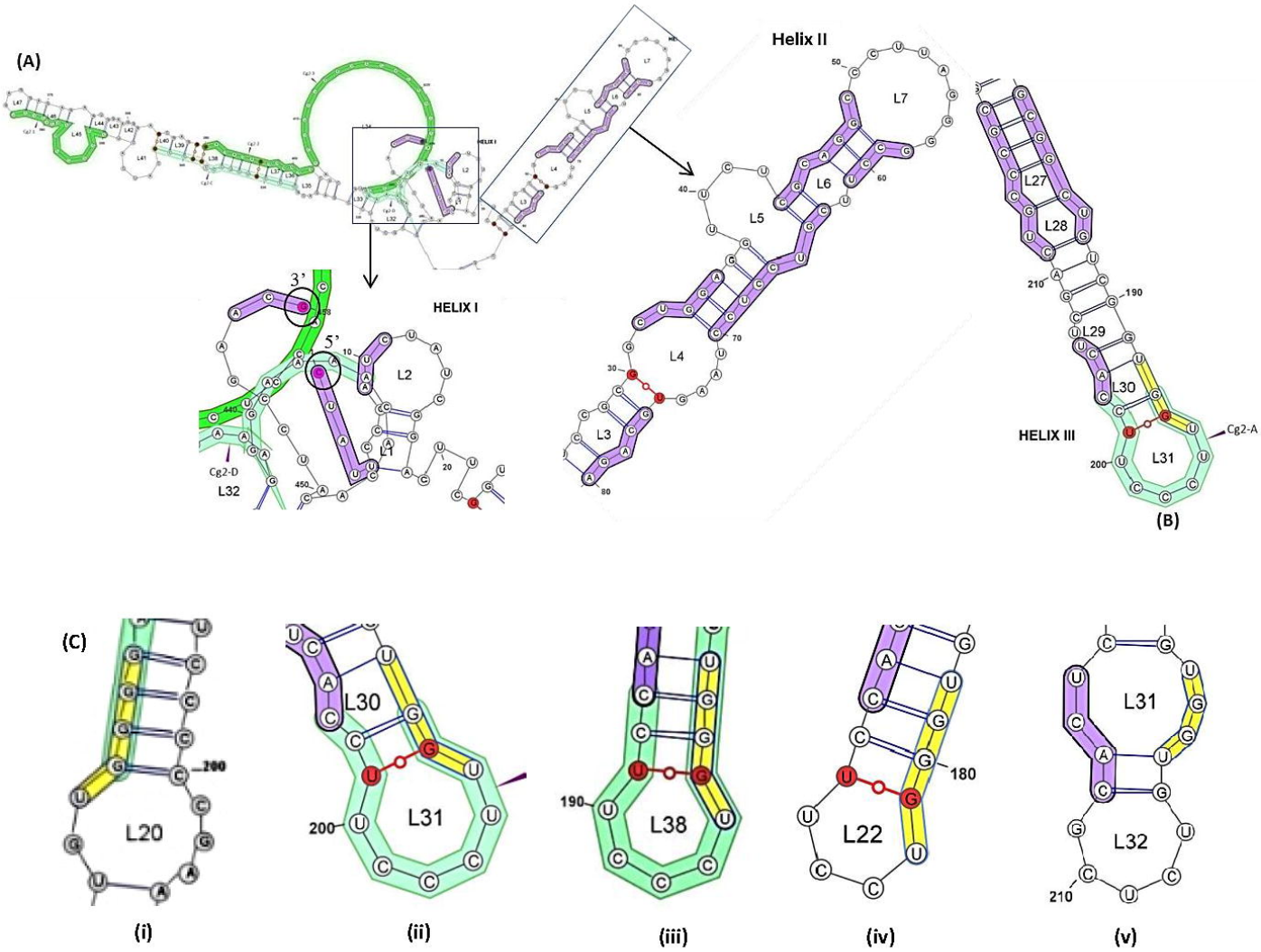

Figure 4. Conserved nucleotide sequences on ITS2 secondary structure. Catfish conserved regions are highlighted in blue where fig. (A) showing the conserved sequences on 5' (CUAUU) and 3' end (ACG), helix I (AAUC) and II (CUGGA, CGCAGGC, GCCU, CGUCCUCC, GCAGA); fig. (B) conserved sequences on the helix III (GCGGCUG, CACU, CUGCCCGC); and fig. (C) showing the eukaryotic conserved UG motif highlighted in yellow: (i) UGGGG in *C. batrachus,* (ii) UGGU in *C. gareipinus*, (iii) UGGGU in *H. fossilis*, (iv) UGGGU in *S. seenghala*, (v) UGGU in *W. attu*.

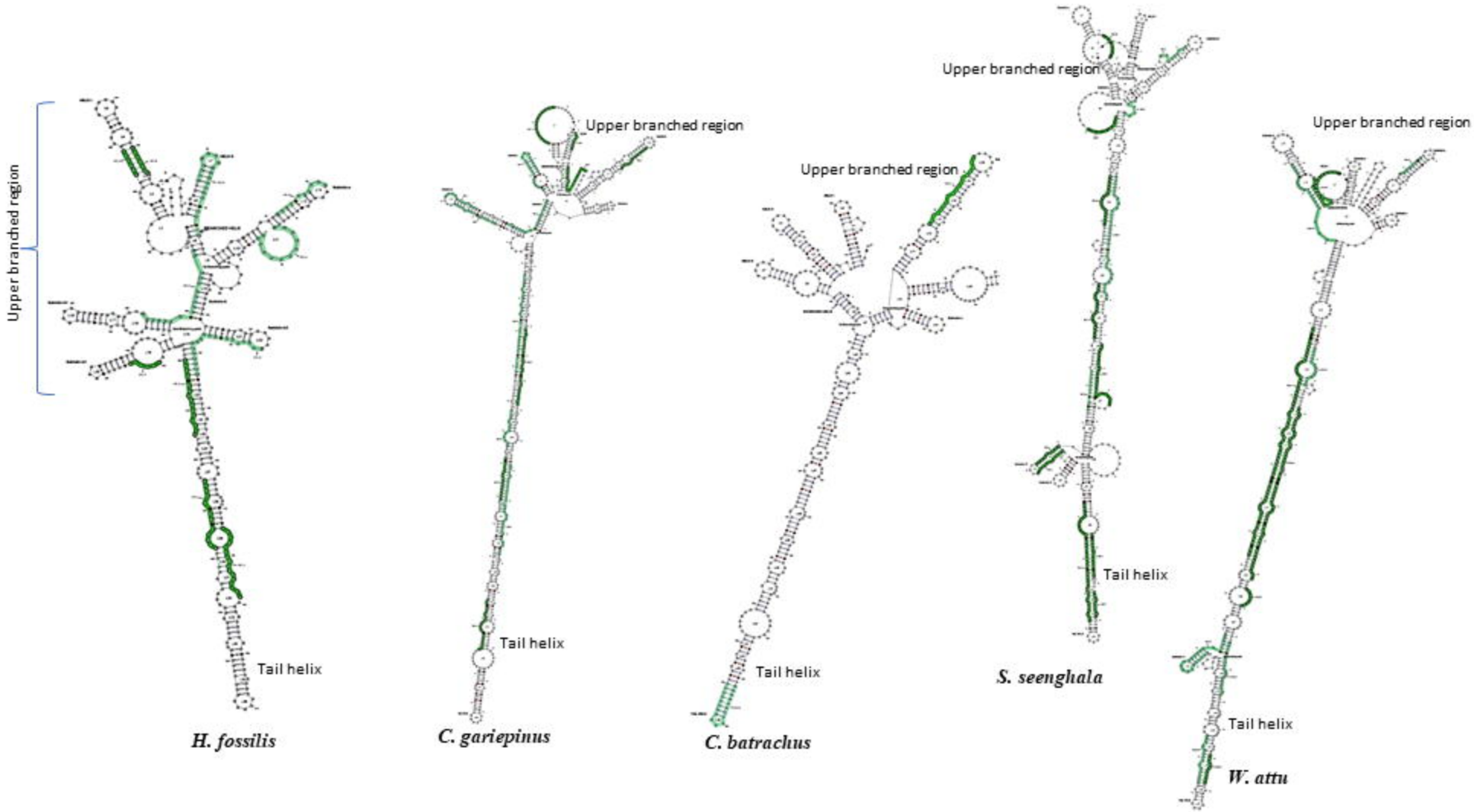

Figure 5. Secondary structure of ITS1. The ITS1 of *H. fossilis, C. gariepinus, C. batrchus, S. seenghala, and W. attu* showing a similar secondary structure without any fix number of helices unlike ITS2, with a upper branched region and a long tail helix.

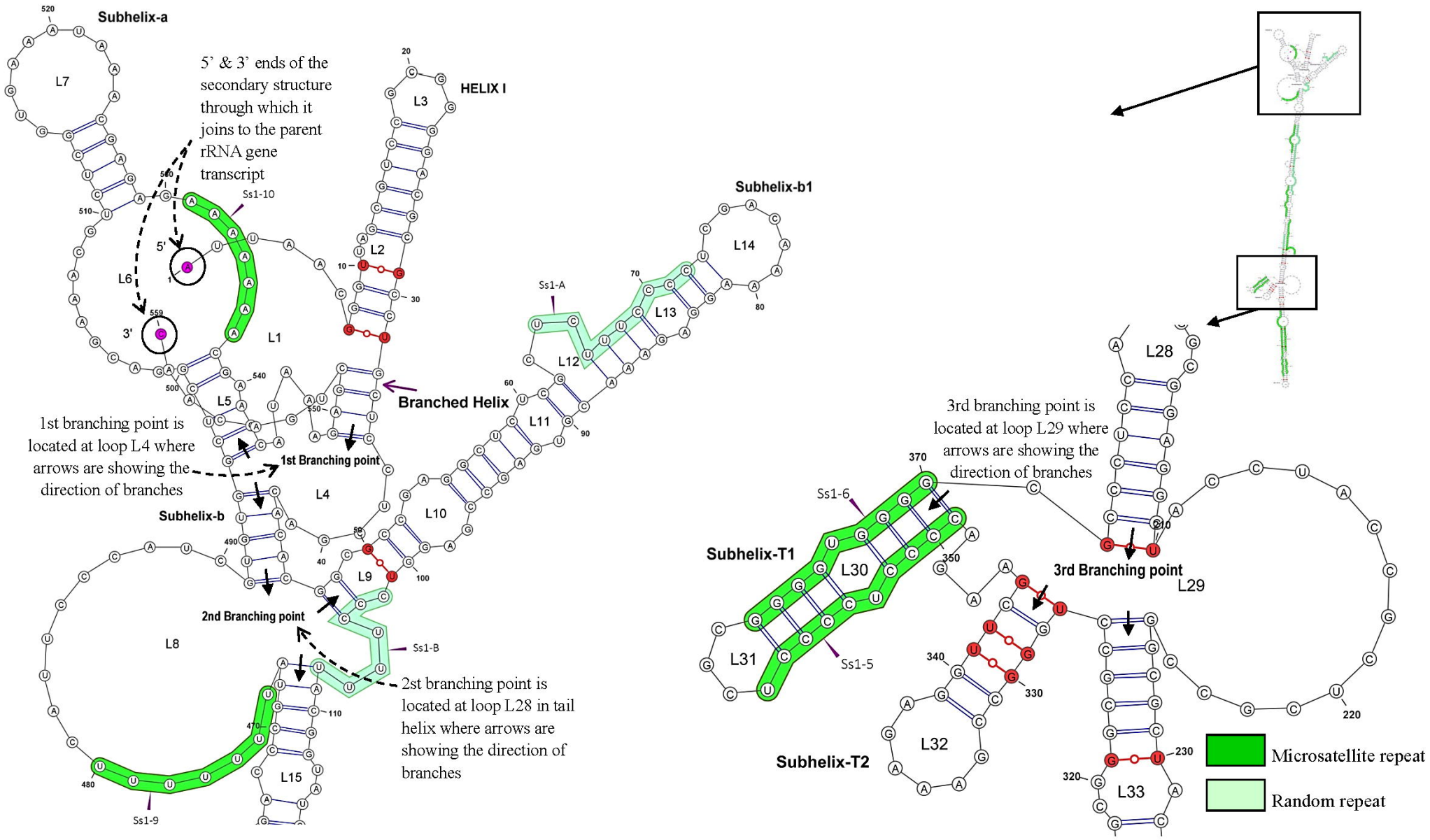

Figure 6. Structural organisation of ITS1 secondary structure (*Sperata seenghala).* The ITS1 secondary structure is composed of multiple helices with one large branched helix. The helices are interrupted with loops (L) of variable size. The arrow → indicates the direction of the branching. The annotation was done with java-based software VARNA. Microsatellite repeat, Random repeat, branching points, non-canonical bp are shown in magnified view round the secondary structure.