Title

Molecular signatures in rDNA internal transcribed spacers ITS1 and ITS2 for species characterization with novel primer sets

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 (Fig. 1); and hence share all its general features viz. multi-chromosomal origin, tandemly repeated presence [1] and concerted evolution [2]. 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 [3,4]. 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 [5]. These microsatellites are tandemly repeated sequences of short nucleotide sequence/motif ranges from 1 to 6 bp [6]. 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 [7]. Similarly, such a wide difference is also observed in ITS2 which is reportedly just 363-369 bp long in *Anopheles nuneztovari* [8], 384-490 bp in catfishes [9] and reaches to a length of 1300 bp in Crayfish [3].

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 [10-12]. The ITS2 also provides better information for taxonomic inferences when used along with its secondary structure [13]. 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 [14,15], resolving biogeography [16] and species complex in plants [5,17]. Even within the animal kingdom they are explored more in the invertebrate, viz. phylogenetic relationship in Ciliates [18], corals[19], resolving cryptic speciation in nematodes [20], phylogeography of Mosquito [8], species identification of Molluscs [21]. Several studies have been conducted in vertebrates also for phylogenetic relationship and species identification [22,23] including several reports from fish species as well [24-32] 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 [12,33], 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 (Fig. 2) 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 [34]. 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 [35].

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/) [36] (Table 1). The primers were found compatible with species from 159 genera, where 37 genera were found with compatibility in multiple sister species (Check Supplementary: Primer compatibility).

Table 1: Newly designed primers for ITS1 & ITS2 using NCBI primer designing tool.

Primer	Sequence (5'-3')	Tm (°C)	GC(%)
ITS.1.F1 (ITS1 forward 1)	TCG ATC AAA CTT GAC TAT CTA GAG G	53.6	40.0
ITS.1.F2 (ITS1 forward 2)	CGA TCA AAC TTG ACT ATC TAG AGG A	53.6	40.0
ITS.1.R1 (ITS1 reverse)	CAG CTG GCT GCG TTC TTC AT	58.3	55.0
ITS2.F (ITS2 forward)	CAG GAC ACA TTG ATC ATC GAC AC	55.8	47.8
ITS2.R (ITS2 reverse)	AAT TCA GCG GGT CGT CTC	54.9	55.6

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 (Table 2). 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, $MgCl_2$ (varied with species), 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. 3, 4; Table 2). 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.

Table 2: Standardized PCR conditions for ITS1& ITS2, PCR products and Target regions

Primer	Species	MgCl2 (mM)	Ta (°C)/Sec.	PCR Product[¶] (bp)
	C. batrachus	3.5	53°C /60sec	~500
ITS.1.F1/F2	*C. gariepinus*	1.5	53°C /60sec	~500
ITS.1.R1	*H. fossilis*	2.5	49°C /60sec	500-600
	S. seenghala	2.5	49°C /60sec	~700
	W. attu	2.5	49°C /60sec	~700
	C. batrachus	1.5	50°C/45sec	400-500
ITS2.F	*C. gariepinus*	1.5	50°C/45sec	~600
ITS2.R	*H. fossilis*	1.5	50°C/45sec	~500
	S. seenghala	1.5	50°C/45sec	~500
	W. attu	1.5	50°C/45sec	~500

[¶] include small rDNA segment at 5'and 3' ends

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' [37]. The repeated sequence tracing was done with MISA- a web based repeat finding tool (https://webblast.ipk-gatersleben.de/misa/) [38]. 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, multiple alignment of secondary structures, conserved motif of secondary structures [39]. The ITS1 secondary structure prediction was done using RNAfold web server (http://rna.tbi.univie.ac.at/cgi-

bin/RNAWebSuite/RNAfold.cgi) [40]. The LocARNA web server (https://rna.informatik.uni-freiburg.de) [41-43] 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 [44,45] and VARNA [46].

RESULTS

Description of ITS Repeats

The ITS1 and ITS2 sequences of each catfish species were found with the presence of multiple repeats (Fig. 5– 14). Some repeats qualify the standard definition of microsatellites [47] 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 (Table 2).

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 (Fig.5-14). The ITS1 and ITS2 both have shown microsatellite and random repeats except the *H. fossilis* ITS2 which consists of only random repeats (fig. 7). 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.

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. 15– 19).

Motif search

The eukaryotic ITS2 possess some specific nucleotide features known as hallmark of the eukaryotic ITS2 sequences and its secondary structure. The motif search using ITS2 database server has found these hallmark features in the ITS2 sequences of all five catspecies, too. The two important features which are observed ubiquitously in different eukaryotic species are- i) A conserved UG motif in different variants viz. UGGU/UGGGU string located in the helix III of the secondary structure towards the 5' end. ii) Pyrimidine-Pyrimidine mismatch- this mismatch consists of placement of a pyrimidine base against a pyrimidine base instead of a purine base in the helix II of the secondary structure. It could be a cytosine & cytosine, thymine/uracil & thymine/uracil or thymine/uracil & cytosine. It leads to a slight bulge in helix II of the secondary structure because of non bonding between the bases of the same group. 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 (Supplementary Fig. 1).

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. 20). 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 1).

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. 21). All the species have a shown a distinct pattern of non-canonical base pairing though only one type of base pairing i.e. G-U is found in all the secondary structure. The presence of another one i.e. adenine with cytosine (A-C) found to be exceptionally rare, just a single one 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 2).

Length of helices

The ITS2 secondary structure possessed considerable differences among its four helices which was clearly evident in all the species (Supplementary Table 3, Fig. 22). 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 [49]. Helix branching is also reported in several species in eukaryotes, and here among catfishes, single branching is found in helix II of *C. batrachus* at loop L6 which has resulted in two subhelices IIa & IIb (fig. 15).

Secondary Structure of ITS1

The secondary structure analysis of ITS1 in five catfish species has shown a structural resemblance with that of ITS2. Though, unlike ITS2, no hallmarks have been reported for ITS1, but still an analogy to ITS2 in the overall appearance is observed in the ITS1 secondary structures.

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 structural contour of the ITS2, although there was no central ring, and the 5' and 3' end of ITS1 were differently located from species to species.

The origin of ITS1 structure from the parent rRNA gene transcript was appeared different as the point of attachment i.e. the 5' and 3' end of ITS1 secondary structure were located on one side of the structure unlike the presence of central ring and stalk in ITS2, and from there originates the highly branched ITS1 secondary structure with multiple helices and sub-helices along with a long helix pointing outward which resembles the helix III of ITS2. Although within this analogy the structural attributes of ITS1 were entirely different from that of ITS2, the features of any helix were found to be conserved in the secondary structure of ITS1 too, as observed in ITS2 helices viz. presence of internal and terminal loops, non-canonical base-pairs, variable length of helices. Helix branching is comparatively more common than ITS2, and the number of helices in ITS1 is also quite variable between the five species, unlike ITS2 which possesses a fixed number of helices. 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 (Fig. 23-27). The pattern of helix branching and sub-branching having major differences between species has given a unique secondary structure to the ITS1 in each species. Additionally, unlike ITS2, no conserved sequence motif was found in multiple sequence alignment of ITS1 secondary structures (Supplementary Fig. 2).

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 4).

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 [50], nematodes [51], beetles [52], fishes [24, 53]. But there was no previous report of repeat distribution as observed in these catfish species, even the type and distribution of nucleotide repeats wassimilar 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 [54], preventing intragenomic ITS variant due to multichromosomal origin, allowing only the accumulation of interspecific difference [2]. 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 [55]. 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 [56]. 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 [10-12] 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 [49, 57, 58,59-61], 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**.

Abbreviations

rDNA= ribosomal DNA; ITS = Internal transcribed spacer; rRNA= ribosomal RNA; MEGA= Molecular Evolutionary Genetic Analysis; MISA= MIcroSAtellite identification tool; VARNA= Visualization Applet for RNA; LocARNA= Local Alignment of RNA; 4SALE= Secondary Structure Alignment and Editing.

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

1. Reed, K. M., Hackett, J. D., Phillips, R. B., 2000. Comparative analysis of intra-individual and inter-species DNA sequence variation in salmonid ribosomal DNA cistrons. Gene 249, 115-125.
2. Hillis, D.M., Dixon, M.T., 1991. Ribosomal DNA, molecular evolution and phylogenetic inference. Q. Rev. Biol. 66, 411-453.
3. Harris, D.J., Crandall, K.A., 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): Implications for phylogenetic and microsatellite studies. Mol. Biol. Evol. 17, 284–291.
4. 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.
5. Ritland, C. E., Ritland, K., Straus, N. A., 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.
6. Kelkar, Y.D., Strubczewski, N., Hile, S.E., Chiaromonte, F., Eckert, K.A., Makova, K.D., 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.
7. Gonzalez, I.L., Sylvester, J.E., Smith, T.F., Stambolian, D., Schmickel, R.D., 1990.

Ribosomal RNA gene sequences and hominoid phylogeny. Mol. Biol. Evol. 7, 203–219.

8. Fritz, G.N., 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.
9. Jansen, G., Devaere, S., Weekers, P.H.H., 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.
10. Pleyte, K.A., Duncan, S.D., Phillips, R.B., 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.
11. Côté, C.A., Peculis, B.A., 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.
12. Abeyrathne, P.D., Lalev, A.I., Nazar, R.N., 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.
13. Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends. Genet. 19, 370-375.
14. Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., et al., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. U.S.A 109, 6241–6246.
15. 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.
16. Tremetsberger, K., Weiss-Schneeweiss, H., Stuessy, T., Samuel, R., Kadlec, G., Ortiz,

M.A., 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.

17. 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.
18. 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.
19. Aguilar, C., Reimer, J.D., 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.
20. Pinacho-Pinacho, C.D., García-Varela, M., Sereno-Uribe, A.L. 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.
21. Hurtado, N.S., Pérez-García, C., Morán, P., Pasantes, J.J., 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.
22. Coleman, A.W., 2013. Analysis of Mammalian rDNA Internal Transcribed Spacers. PLoS ONE 8, e79122.
23. Voronov, A.S., Shibalev, D.V., Ryskov, A.P., Kupriyanova, N.S., 2006. Evolutionary divergence of ribosomal internal transcribed spacer 2 in lizards. Mol. Biol. 40, 37–42.
24. Sajdak, S.L., Phillips, R.B., 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.

25. Domanico, M.J., Phillips, R.B., Oakley, T.H., 1997. Phylogenetic analysis of Pacific salmon (genus *Oncorhynchus*) using nuclear and mitochondrial DNA sequences. Can. J. Fish. Aquat. Sci. 54, 1865–1872.
26. Booton, G.C., Kaufman, L., Chandler, M., Oguto-Ohwayo, R., Duan, W., Fuerst, P. A., 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.
27. Presa, P., Pardo, B.G., Martínez, P., Bernatchez, L., 2002. Phylogeographic congruence between mtDNA and rDNA its markers in brown trout. Mol. Biol. Evol. 19, 2161–2175.
28. Montoya-Burgos, J.I., 2003. Historical biogeography of the catfish genus *Hypostomus* (Siluriformes: Loricariidae), with implications on the diversification of Neotropical ichthyofauna. Mol. Ecol. 12, 1855–1867.
29. Vera, M., Sourinejad, I., Bouza, C., Vilas, R., Pino-Querido, A., Kalbassi, M.R., 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.
30. Vinas, J., Tudela, S., 2009, A Validated Methodology for Genetic Identification of Tuna Species (Genus *Thunnus*). PLoS One 4, e7606. https://doi.org/10.1371/journal.pone.0007606.
31. 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.
32. Puncher, G.N., Arrizabalaga, H., Alemany, F., Cariani, A., Oray, I.K., Karakulak, F.S., 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.

33. Geerlings, T.H., Vos, J.C., Raué, H.A., 2000. The final step in the formation of 25S rRNA in Saccharomyces cerevisiae is performed by 5'-->3' exonucleases. RNA 6, 1698-1703.
34. Talwar, P.K., Jhingran, A.G., 1992. Inland fishes of India and adjacent countries, vol. 2. Balkema Publishers, Rotterdam.
35. Montgomery, G.W., Sise, J.A., 1990. Extraction of DNA from sheep white blood cells. New Zealand Journal of Agricultural Research 33, 437-441.
36. Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 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.
37. Kumar. S., Stecher, G., Tamura, K., 2016. MEGA7, Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870-1874.
38. Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., 2017. MISA-web, a web server for microsatellite prediction. Bioinformatics 33, 2583–2585.
39. 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.
40. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., Hofacker, I.L., 2008. The Vienna RNA websuite. Nucleic Acids Res. 36, W70-W74.
41. Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F., 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.

42. Will, S., Joshi, T., Hofacker, I.L., Stadler, P.F., Backofen, R., 2012. LocARNA-P, accurate boundary prediction and improved detection of structural RNAs. RNA 18, 900-914.

43. Raden, M., Ali, S.M., Alkhnbashi, O.S., Busch, A., Costa, F., Davis, J.A., 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.

44. Seibel, P.N., 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.

45. Seibel, P.N., 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

46. Darty, K., Denise, A. & Ponty, Y. 2009. VARNA, Interactive drawing and editing of the RNA secondary structure, Bioinformatics 25(15), 1974–1975.

47. Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R., Vieira, M.L.C., 2006. Origin, evolution and genome distribution of microsatellites. Genet. Mol. Biol. 29, 294–307.

48. 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.

49. 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.

50. Grajales, A., Aguilar, C., Sánchez, J.A., 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.

51. Conole, J.C., Chilton, N.B., Jarvis, T., Gasser, R.B., 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.

52. 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.

53. 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.

54. Elder, J.F.Jr., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. 70, 297-320.

55. Eickbush, T.H., Eickbush, D.G., 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175, 477–485.

56. Wang, W., Zhang, X., Garcia, S., Leitch, A.R., Kovařík, A., 2023. Intragenomic rDNA variation - the product of concerted evolution, mutation, or something in between? Heredity 131, 179–188. https://doi.org/10.1038/s41437-023-00634-5

57. Schultz, J., Wolf, M., 2009. ITS2 sequence–structure analysis in phylogenetics, A how-to manual for molecular systematic. Mol. Phylogenet. Evol. 52, 520-523.

58. Chen, C.A., Chang, C.C., Wei, N.V., Chen, C.H., Lein, Y.T., Lin, H.E., Dai, C.F., Wallace C.C., 2004. Secondary structure and phylogenetic utility of the ribosomal internal transcribed spacer 2 (ITS2) in scleractinian corals. Zool. Stud. 43, 759–771.

59. Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. J. Mol. Evol. 44, 258-271.

60. Joseph, N., Krauskopf, E., Vera, M.I., 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.

61. Young, I., Coleman, A.W., 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.

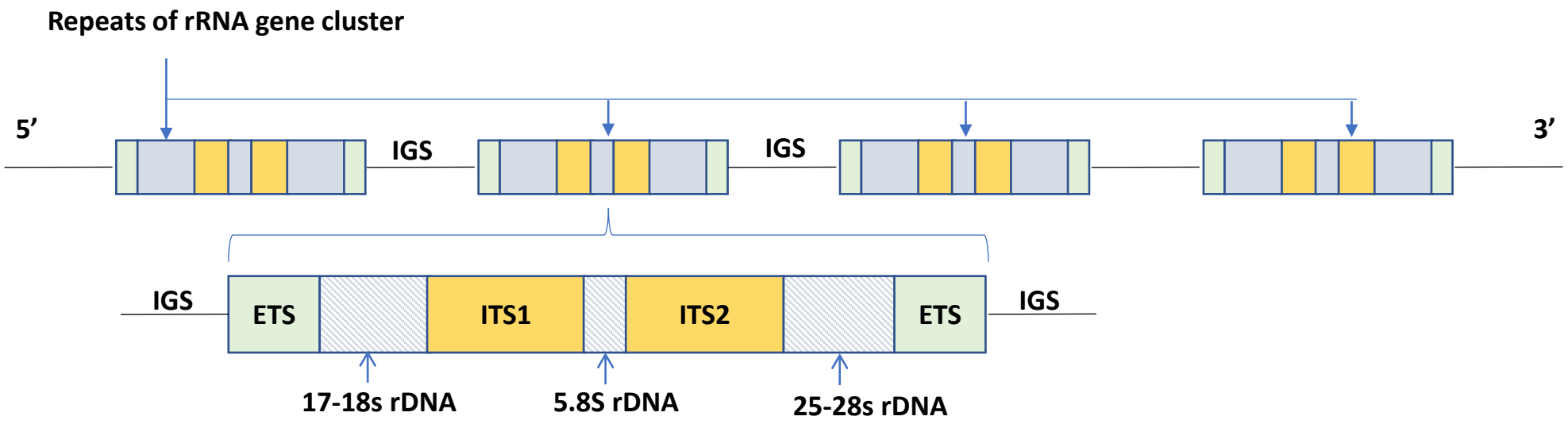


Figure1. rRNA gene showing the repeating clusters of ribosomal subunit genes and spacer regions in the order of 5' to 3' as ETS- External Transcribed Spacer, 17s-18s rDNA, ITS1- Internal Transcribed Spacer 1, 5.8s rDNA, ITS2- Internal Transcribed Spacer 2, 25-28s rDNA, IGS- Intergenic spacer.

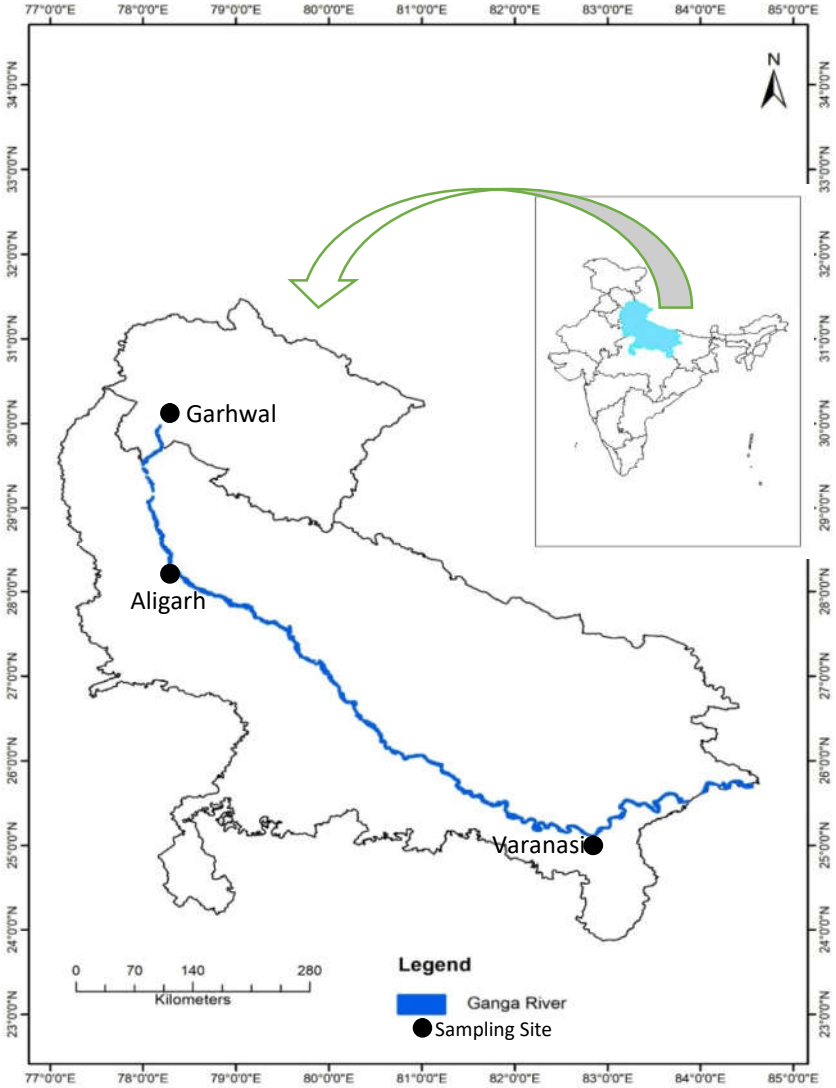
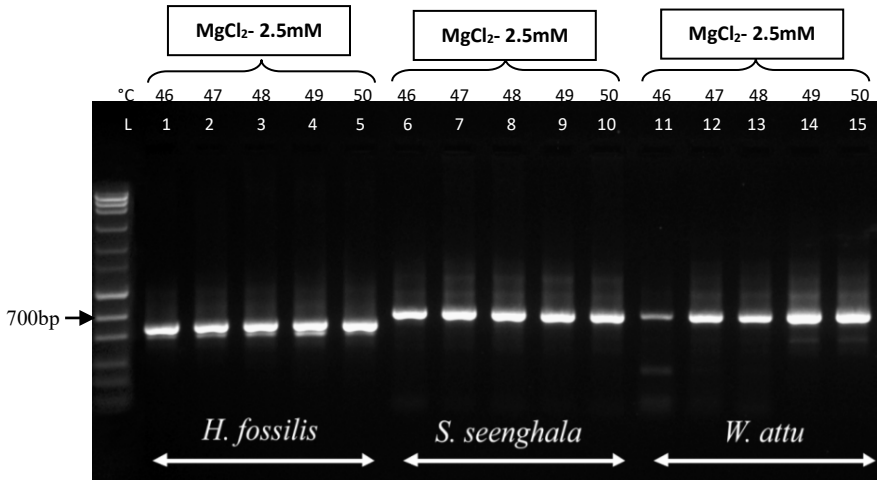
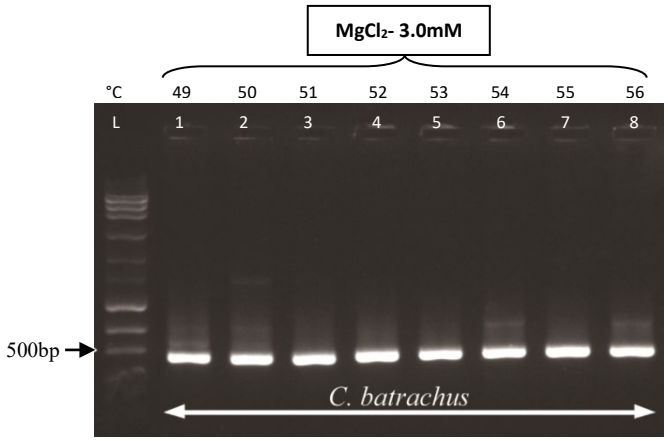


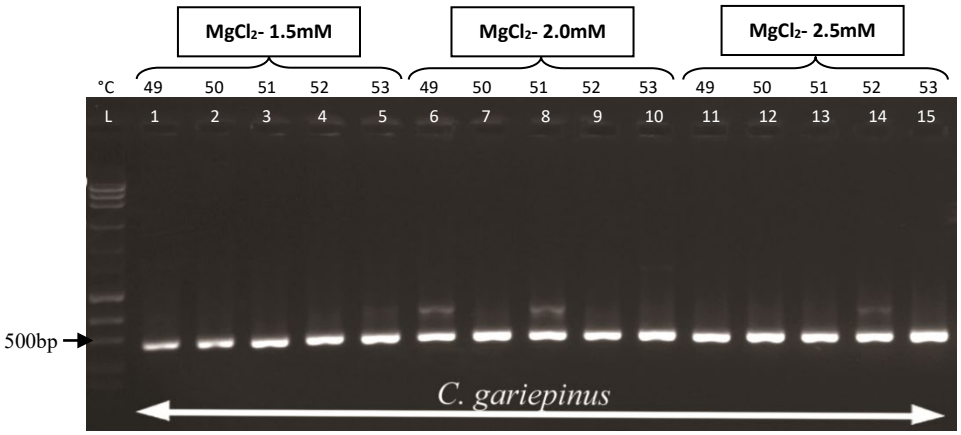
Figure 2. Sampling sites: The three sampling sites- Garhwal, Aligarh and Varanasi on the stretch of river Ganga in the north India (shown in inset).



ITS1-A

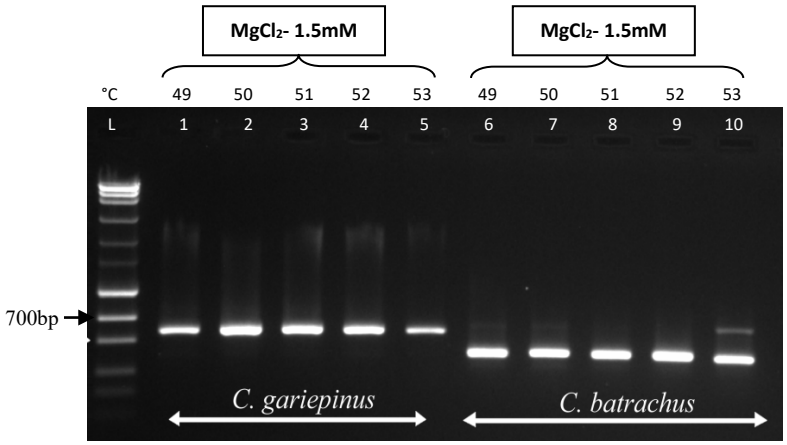


ITS1-B

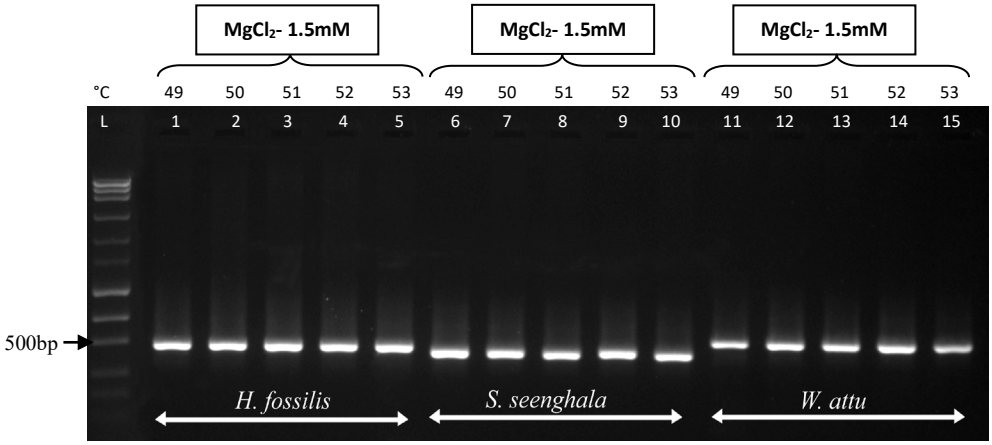


ITS1-C

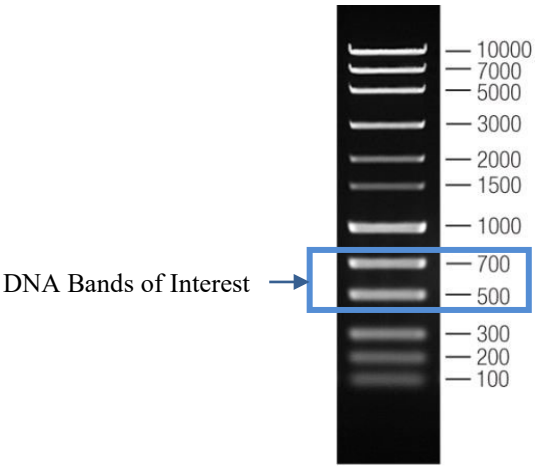
Figure 3. Agarose gel images of ITS1. Figure ITS1-A: *H. fossilis*, *S. seenghala*, & *W. attu*. ITS1-B: *C. batrachus*, and ITS1-C: C*. gariepinus*.. The band size is compared with standard DNA mass ruler from thermo Scientific with a band range of 100bp to 10000bp (indicated as L).



ITS2-A



ITS2-B



L – Standard DNA mass ruler (Thermo Scientific), band range of 100bp to 10000bp

Figure 4. Agarose gel images of ITS2. ITS2-A: *C. gariepinus* and *C. batrachus*; ITS2-B: *H. fossilis*, S. *seenghala*, & *W. attu*. The band size is compared with standard DNA mass ruler from thermo Scientific with a band range of 100bp to 10000bp (indicated as L).

Cb: *Clarias batrachus*

```
                                                                  5.8s rRNA  ITS2
Cb32 ATTAAATTGCGGCCCGAGTCCCTCCTGGGGCTACGTCTGTCTGACTGCTATTCCCATCGATCAGACTTCG 70
Cb35 ----ATTGNNGGCCCGAGTCCCTCCTGGGGCTACGTCTGTCTGACTGCTATTCCCATCGATCAGACTTCG 70
         * *   ************************************************************

Cb32 GTCCGTGGCTGTCGGTGGTTCCCTTCCACTTCGACTGGCCGCTTCACACCTCTGCAGGGCCGGTGGTCGC 140
Cb35 GTCCGTGGCTGTCGGTGGTTCCCTTCCACTTCGACTGGCCGCTTCACACCTCTGCAGGGCCGGTGGTCGC 140
     ************************************************************************

Cb32 GCGAAAAGCTAGTGGGTGCGGTTTGGCGCGTGTGAACGTGTGGGCGGGGAACAGCTGGGGGTCTGGCGGT 210
Cb35 GCGAAAAGCTAGTGGGTGCGGTTTGGCGCGTGTGAACCTGTGGGCGGGGAACAGCTGGGGGTC------- 210
     ************************************* ************************

                                                Cb2-1
Cb32 GCTCTCTCCGAGACGCTTTTCACTTCTCTCTCTGCTGCCTCTATCTCCCCCGAACTTCTCTGCGGAGAGA 280
Cb35 ------TCCGAGACGCTTTTCACTTCTCTCTCTGCTGCCTCTATCTCCCCCGAACTTCTCTGCGGAGAGA 280
            ************************************************************

                    Cb2-A
Cb32 CGTGTGGGGAGACGAGGGAGGAGAGGGGTCAGGGAGGCTCACCTTTCTCTCGAAAGATGCATCCGATCAC 350
Cb35 CGTGTGGGGAGACGAGGGAGGAGAGGGGTCAGGGAGGCTCACCTTTCTCTCGAAAGATGCATCCGATCAC 350
     ************************************************************************

                                  28s rRNA
Cb32 AACTTCAGATTAAACGAGACGACCCGCTGAATT 383
Cb35 AACTTCAGATTAAACGAGACGACCCGCTGAATT 383
     *********************************
```

Microsatellite repeats: Cb2-1

Random repeats: Cb2-A

Figure 5. Multiple sequence alignment of C. batrachus ITS2 (Cb2) showing the position of microsatellite repeat Cb2-1 and random repeat Cb2-A.

Cg: *Clarias gariepinus*

```
                                                                                  5.8s rRNA  ITS2
Cg07 ATTAAATTGCGGCCCGGGTCC-TCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCCCATCTATCGGACT 70
Cg11 AATANATTGCGGCCCGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCCCATCTATCGGACT 70
     * ** **************** ************************************************

Cg07 TCGGTCCGCGGCTGGAGGTTCTCGCAGGCCCTTAGGGGCCTTCGTCCTCCTAAGTGCAGACTCGTCTTTC 140
Cg11 TCGGTCCGCGGCTGGAGGTTCTCGCAGGCCCTTAGGGGCCTTCGTCCTCCTAAGTGCAGACTCGTCTTTC 140
     ************************************************************************

Cg07 CCCTGGCGCCTCGTGTCGCTCTCCCCTCCGTGTCCACGACTCAAAACCCCTCCGGCCGGAAACGGTTTCC 210
Cg11 CCCTGGCGCCTCGTGTCGCTCTCCCCTCCGTGTCCACGACTCAAAACCCCTCCGGCCGGAAACGGTTTCC 210
     ************************************************************************
                                                Cg2-A
Cg07 GTGCCCTGCGGTCGTCGTGCGCGGCTGTCGGTGGTTCCCTTCCACTTCGACTGCCCGCGTCACGCCTCCG 280
Cg11 GTGCCCTGCGGTCGTCGTGCGCGGCTGTCGGTGGTTCCCTTCCACTTCGACTGCCCGCGTCACGCCTCCG 280
     ************************************************************************
                                                                    Cg2-B
Cg07 CAGGCCGGGGCCGCGCGAAAAGCCGGTGGGGCAGTGGGCGCAGTGCAGGCGGTCGGCGGGCGGGGAGGCC 350
Cg11 CAGGCCGGGGCCGCGCGAAAAGCCGGTGGGGCAGTGGGCGCAGTGCAGGCGGTCGGCGGGCGGGGAGGCC 350
     ************************************************************************
                                        Cg2-C                   Cg2-1
Cg07 GGGGGTCGGACGGCGCTCTCCGAGACGCCCTTCTCCCTTCCTCATCACTCCCCCCGCTCTCTCCGAGGAG 420
Cg11 GGGGGTCGGACGGCGCTCTCCGAGACGCCCTTCTCCCTTCCTCATCACTCCCCCCGCTCTCTCTGAGGAG 420
     ******************************************************************* ******
                                Cg2-2                   Cg2-3                   Cg2-D
Cg07 AGAGTGGAAGGGAGCGAGGGAGAGAGGGGGGGCTCACCTTCTCCTCCTCCTCCACCTCCGTGGAGAGGGG 490
Cg11 AGAGTGGAAGGGAGCGAGGGAGAGAGGGGGGGCTCACCTTCTCCTCCTCCTCCACCTCCGTGGAAAAGGG 490
     ******************************************************************* * ***
                          28s rRNA
Cg07 AGAACACACATCCGAACGCGACCTCAGATCAGACGAGACG 530
Cg11 AAAACACACATCCGAACGCGACCTCAGATCAGAAGAGACG 530
     * ******************************** ******
```

Microsatellite repeats: Cg2-1, Cg2-2, Cg2-3

Random repeats: Cg2-A, Cg2-B, Cg2-C, Cg2-D

Figure 6. Multiple sequence alignment of *C. gariepinus* ITS2 (Cg2) showing the position of microsatellite repeats Cg2-1 to Cg2-3 and random repeats Cg2-A to Cg2-D.

Hf: *Heteropneustes fossilis*

```
                                                                                  5.8s rRNA  ITS2
Hf40 ATTANATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCTCATCGATCGGAC 70
Hf44 TTTANATTGCGGCCCGGGGTCC-TCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCTCATCGATCGGAC 70
Hf42 -TTNCATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCTCATCGATCGGAC 70
Hf41 TTTACATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTCTCATCGATCGGAC 70
      **  ***************** ********************************************

                                                   Hf2-A                  Hf2-B
Hf40 TTCGGTCCGCGACTGGAGATTCGCAGGCACCCGCCTTCGTCCTCCCAAACGCAGACTCGTTTTTCCCTTC 140
Hf44 TTCGGTCCGCGACTGGAGATTCGCAGGCACCCGCCTTCGTCCTCCCAAACGCAGACTCGTTTTTCCCTTC 140
Hf42 TTCGGTCCGCGACTGGAGATTCGCAGGCACCCGCCTTCGTCCTCCCAAACGCAGACTCGTTTTTCCCTTC 140
Hf41 TTCGGTCCGCGACTGGAGATTCGCAGGCACCCGCCTTCGTCCTCCCAAACGCAGACTCGTTTTTCCCTTC 140
     ********************************************************************

                       Hf2-C                                Hf2-D
Hf40 CACCCCGTATTGTTCTCCCCTCCGCGTCCACGACTCAAAACCCCC-GGCCGGAAACGGTTCCGTGCCCTG 210
Hf44 CACCCCGTATTGTTCTCCCCTCCGCGTCCACGACTCAAAACCCCC-GGCCGGAAACGGTTCCGTGCCCTG 210
Hf42 CACCCCGTATTGTTCTCCCCTCCGCGTCCACGACTCAAAACCCCCCGGCCGGAAACGGTTCCGTGCCCTG 210
Hf41 CACCCCGTATTGGTCTCCCCTCCGCGTCCACGACTCAAAACCCCC-GGGCGGAAACGGTTCCGTGGCCTG 210
     ************ ******************************** ** *************** ****

                     Hf2-E
Hf40 CGGTCGTCGTGCGCGGCTGTCGGTGGGTCCCTTCCACTTAGACTGCCCGCGTCACGACTCCGCAGGCCTG 280
Hf44 CGGTCGTCGTGCGCGGCTGTCGGTGGGTCCCTTCCACTTAGACTGCCCGCGTCACGACTCCGCAGGCCTG 280
Hf42 CGGTCGTCGTGCGCGGCTGTCGGTGGGTCCCTTCCACTTAGACTGCCCGCGTCACGACTCCGCAGGCCTG 280
Hf41 GGGTCGTCGTGCGCGGCTGGCGGTGGGTCCCTTCCACTTAGACTGGCCGCGTCACGACTCCGCAGGGCTG 280
      ****************** ************************ ******************** ***

                          Hf2-F                                  Hf2-G
Hf40 GGCCGCGTGAGAAGCCGGTGGGTGTGGTCGCGCGTAGCGGGAGAACGGCGGGTAGGCGGCGGGGAGAGGC 350
Hf44 GGCCGCGTGAGAAGCCGGTGGGTGTGGTCGCGCGTAGCGGGAGAACGGCGGGTAGGCGGCGGGGAGAGGC 350
Hf42 GGCCGCGTGAGAAGCCGGTGGGTGTGGTCGCGCGTAGCGGGAGAACGGCGGGTAGGCGGCGGGGAGAGGC 350
Hf41 GGCCGCGTGAGAAGCCGGTGGGTGTGGTCGCGCGTAGCGGGAGAACCGCGGGTAAGCGGCGGGGAAAAGC 350
     ********************************************** ******* ********** * **

                                                                          28s rRNA
                          Hf2-H
Hf40 GGCGCTTTAGGCCCGGGCCGCCGCGTGGCGCCCCACTGCCGAAGCGTTAAGCTTCGGAAATCCGAACGCG 420
Hf44 GGCGCTTTAGGCCCGGGCCGCCGCGTGGCGCCCCACTGCCGAAGCGTTAAGCTTCGGAAATCCGAACGCG 420
Hf42 GGCGCTTTAGGCCCGGGCCGCCGCGTGGCGCCCCACTGCCGAAGCGTTAAGCTTCGGAAATCCGAACGCG 420
Hf41 GGCGCTTTAAGGCCGGGCCGCCGCGTGGCGCCCCACTGGCGAAGCGTTAAGCTTCGGAAATCCGAACGCG 420
     ********* * ************************* ******************************

Hf40 ACCTCAGATCAGACGAGACGACC 443
Hf44 ACCTCAGATCAGACGAGACGACC 443
Hf42 ACCTCAGATCAGACGAGACGACC 443
Hf41 ACCTCCGATCAGACGAGACGACC 443
     ***** *****************
```

Random repeats: Hf2-A, Hf2-B, Hf2-C, Hf2-D, Hf2-E, Hf2-F, Hf2-G, Hf2-H

Figure 7. Multiple sequence alignment of *H. fossilis* ITS2 (Hf2) showing the position of random repeats Hf2-A to Hf2-H.

Ss: *Sperata seenghala*

```
                                                                                  5.8s rRNA  ITS2
Ss08 -TTAANTTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCGTCGATCGGAC 70
Ss11 TTTNCATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCGTCGATCGGAC 70
Ss09 ATTNAATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCGTCGATCGGAC 70
Ss01 TTTAAATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCGTCGATCGGAC 70
      **   ************************************************************

                                                                  Ss2-A
Ss08 GCCCCCGCGTTCCGCGGCTGGAGGTTTTCTTCGCAGGCGGGCCACCCCCCGCCTCCGTCCTCCCAAAAGC 140
Ss11 GCCCCCGCGTTCCGCGGCTGGAGGTTTTCTTCGCAGGCGGGCCACCCCCCGCCTCCGTCCTCCCAAAAGC 140
Ss09 GCCCCCGCGTTCCGCGGCTGGAGGTTTTCTTCGCAGGCGGGCCACCCCCCGCCTCCGTCCTCCCAAAAGC 140
Ss01 GCCCCCGCGTTCCGCGGCTGGAGGTTTTCTTCGCAGGCGGGCCACCCCCCGCCTCCGTCCTCCCAAAAGC 140
     ********************************************************************

                     Ss2-1
Ss08 AGACGCGCCCCTCTCTCTCTC--------CCTCCCCGCGTCGCCGCCTCGCCCAGTTCGGCGCGAGATAA 210
Ss11 AGACGCGCCCCTCTCTCTCTC--------CCTCCCCGCGTCGCCGCCTCGCCCAGTTCGGCGCGAGATAA 210
Ss09 AGACGCGCCCCTCTCTCTCTCTCTCCCTCCCTCCCCGCGTCGCCGCCTCGCCCAGTTCGGCGCGAGATAA 210
Ss01 AGACGCGCCCCTCTCTCTCTCTCTC----TCTCCCCGCGTCGCCGCCTCGCCCAGTTCGGCGCGAGATAA 210
     *********************          ****************************************

      Ss2-2
Ss08 AAAGAAAGCGGCTGCCGGTGGGTCCTTCCACTCTGGCTGCCCGCTTCGAAAAATATAGCGCGCCGAGCGA 280
Ss11 AAAGAAAGCGGCTGCCGGTGGGTCCTTCCACTCTGGCTGCCCGCTTCGAAAAATATAGCGCGCCGAGCGA 280
Ss09 AAAGAAAGCGGCTGCCGGTGGGTCCTTCCACTCTGGCTGCCCGCTTCGAAAAATATAGCGCGCCGAGCGA 280
Ss01 AAAGAAAGCGGCTGCCGGTGGGTCCTTCCACTCTGGCTGCCCGCTTCGAAAAAAATAGCGCGCCGAGCGA 280
     ***************************************************** ****************

                                                Ss2-3          Ss2-B
Ss08 CGGGGCACCCACGCGCGACGTCGTGGATGGGGGAGTGAAAAAG-GGAGGTGGGGCGCACGCACCTCTCTC 350
Ss11 CGGGGCACCCACGCGCGACGTCGTGGATGGGGGAGTGAAAAAG-GGAGGTGGGGCGCACGCACCTCTCTC 350
Ss09 CGGGGCACCCACGCGCGACGTCGTGGATGGGGGAGTGAAAAAAAGGAGGTGGGGCGCACGCACCTCTCTC 350
Ss01 CGGGGCACCCACGCGCGACGTCGTGGATGGGGGAGTGAAAAAA-GGAGGTGGGGCGCACGCACCTCTCTC 350
     ******************************************  **************************

                                                 28s rRNA
                                Ss2-4
Ss08 GA--CCTTACACGTCTCTAAAAAAAAA-CGAACGCGACCTCAGATCAGACGAGACGACCCGCTGAATT 418
Ss11 GA--CCTTACACGTCTCTAAAAAAAAA-CGAACGCGACCTCAGATCAGACGAGACGACCGGCTGAATT 418
Ss09 GA--CCTTACACGTCTCTAAAAAAAAAACGAACGCGACCTCAGATCAGACGAGACGACCCGCTGAATT 418
Ss01 TCTCCCTCACACGTCTCTAAAAAAAAA                                          380
         *** *******************
```

Microsatellite repeats: Ss2-1, Ss2-2, Ss2-3, Ss2-4

Random repeats: Ss2-A, Ss2-B

Figure 8. Multiple sequence alignment of *S. seenghala* ITS2 (Ss2) showing the position of microsatellite repeats Ss2-1 to Ss2-4 and random repeats Ss2-A & Ss2-B.

Wa: *Wallago attu*

```
                                                                                                5.8s rRNA  ITS2
Wa32 TTAAATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCATCGATCGTACG 70
Wa51 TTAAATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCATCGATCGTACG 70
Wa42 TTAAATTGCGGCCCGGGGTCCCTCCCGGGGCTACGCCTGTCTGAGGGTCGCTATTACCATCGATCGTACG 70
     ************************************************************************

Wa32 AAACGCGCGCGACTGGAAGCTCGCAGGCCGGCTGACGGCCTTCGTCCTCCTAAAAGCAGACGCTCCGTTT 140
Wa51 AAACGCGCGCGACTGGAAGCTCGCAGGCCGGCTGACGGCCTTCGTCCTCCTAAAAGCAGACGCTCCGTTT 140
Wa42 AAACGCGCGCGACTGGAAGCTCGCAGGCCGGCTGACGGCCTTCGTCCTCCTAAAAGCAGACGCTCCGTTT 140
     ************************************************************************
                         Wa2-1                  Wa2-2        Wa2-A                      Wa2-B
Wa32 CTCCTTTCCATGCCTTTTTTTTTTTCTCCCAAAAAAGAAACCCCTCCCGAACTCGGCGCCTTTCCCCCAA 210
Wa51 CTCCTTTCCATGCCTTTTTTTTTTTCTCCCAAAAAAGAAACCCCTCCCGAACTCGGCGCCTTTCCCCCAA 210
Wa42 CTCCTTTCCATGCCTTTTTTTTTTTCTCCCAAAAAAGAAACCCCTCCCGAACTCGGCGCCTTTCCCCCAA 210
     ************************************************************************
                Wa2-C          Wa2-3
Wa32 ACCATCCCTTTGGGGCGGGGGGACGCGCGAGCGGCTGCCGGTGGTGTCTCGCACTCTGGCTGCCCGCGAG 280
Wa51 ACCATCCCTTTGGGGCGGGGGGACGCGCGAGCGGCTGCCGGTGGGGTCTCGCACTCTGGCTGCCCGCGAG 280
Wa42 ACCATCCCTTTGGGGCGGGGGGACGCGCGAGCGGCTGCCGGTGGTGTCTCGCACTCTGGCTGCCCGCGAG 280
     ******************************************** *************************
                                Wa2-D
Wa32 CGCGCGTTCCCCGGAAACCCCGAGGGCGGGGAAGGGACAGAGGAAGGGCGCGCCACGGGCGAAAGAGGGG 350
Wa51 CGCGCGTTCCCCGGAAACCCCGAGGGCGGGGAAGGGACAGAGGAAGGGCGCGCCACGGGCGAAAGAGGGG 350
Wa42 CGCGCGTTCCCCGGAAACCCCGAGGGCGGGGAAGGGACAGAGGAAGGGCGCGCCACGGGCGAAAGAGGGG 350
     ************************************************************************
                                   Wa2-4                                Wa2-5
Wa32 GGCGGGAAAAAAGAGAAAAAAAGAGAGAGAGAGACGGGGGGGGGGCTAAAAAAAGAGGGAAACCCAC 420
Wa51 GGCGGGAAAAAAGAGAAAAAAAGAGAGAGAGAGACGGGGAGGGGGCTAAAAAAAGAGGGAAAACCAC 420
Wa42 GGCGGGAAAAAAGAGAAAAAAAGAGAGAGAGAGA--CGGGGAGGGGGCTAAAAAAAGAGGGAGAACCAC 420
     ***********************************  ***** ********** ********** * ****
                                                                28s rRNA
                                      Wa2-E
Wa32 TCCCTCGTTCCGTTGCCCTCTTATCTATTCCCTTCTTTTCAAACGCGACCTCAGATCAGAAGAGACCACC 490
Wa51 CCCCTCGTTCCGGTGCCCTCTTATCTATTTCCTTCCTTTCAAACGCCCCCTCAGATCAAAAGAGACCACC 490
Wa42 TCCCTCGTTCCGTTGCCCTCTTATCTATTTCCTTCCTTTCGAACGCGACCTCAGATCAGACGAGACGACC 490
      *********** *************** ***** **** *****  ********** * ***** ***

Wa32 CGCTGAAATA 500
Wa51 CGCTGAAATA 500
Wa42 CGCTGAATTA 500
     ******* **
```

Microsatellite repeats: Wa2-1, Wa2-2, Wa2-3, Wa2-4, Wa2-5

Random repeats: Wa2-A, Wa2-B, Wa2-C, Wa2-D, Wa2-E

Figure 9. Multiple sequence alignment of *W. attu* ITS2 (Wa2) showing the position of microsatellite repeats Wa2-1 to Wa2-5 and random repeats Wa2-A to Wa2-E.

Cb: *Clarias batrachus*

```
     ITS1                                   Cb1-1
Cb10 ACGGGTAGCCCCCGCAGCGGGGCGCACCTCCTCCTCCACCTCTTGCCCTTGAGATGGACCGAGGCGGCCT 70
Cb16 ACGGGTAGCCCCCGCAGCGGGGCGCACCTCCTCCTCCACCTCTTGCCCTTGAGATGGACCGAGGCGGCCT 70
Cb25 ACGGGTAGCACCCGCAGCGGGGCGCACCTCCTCCTCCACCTCTTGCCCTTGAGATGGACCGAGGCGGCCT 70
     ********* ************************************************************

Cb10 CCGTGACGCTGATGCCTTCCCGAAACCTTTCCGGTGACCGAGAAACAAACGATAGGGCGAGGGTACCTGC 140
Cb16 CCGTGACGCTGATGCCTTCCCGAAACCTTTCCGGTGACCGAGAAACAAACGATAGGGCGAGGGTACCTGC 140
Cb25 CCGTGACGCTGATGCCTTCCCGAAACCTTTCCGGTGACCGAGAAACAAACGATAGGGCGAGGGTACCTGC 140
     ********************************************************************** 140

Cb10 TGCTCGAGTGCGGCGCAACGGGGTTTGGGCTGGCCGTGGTCGGGCGCGTGCGCCGGTCGGGGAGGACATG 210
Cb16 TGCTCGAGTGCGGCGCAACGGGGTTTGGGCTGGCCGTGGTCGGGCGCGTGCGCCGGTCGGGGAGGACATG 210
Cb25 TGCTCGAGTGCGGCGCAACGGGGTTTGGGCTGGCCGTGGTCGGGCGCGTGCGCCGGTCGGGGAGGACATG 210
     ********************************************************************
                                                        Cb1-A
Cb10 TAGGTGTGGGGGTTGGGCATGTGGGTGAAAAGGTGCTACGCAGAGAGCCCCGGGCTCTCTGTGCGGTCGT 280
Cb16 TAGGTGTGGGGGTTGGGCATGTGGGTGAAAAGGTGCTACGCAGAGAGCCCCGGGCTCTCTGTGCGGTCGT 280
Cb25 TAGGTGTGGGGGTTGGGCATGTGGGTGAAAAGGTGCTACGCAGAGAGCCCCGGGCTCTCTGTGCGGTCGT 280
     ********************************************************************

Cb10 CCCGGGATCCCACTCCCTTCTCCCTCGCCTCTGTCTTCCATGCATGCGCGGGCGCTCCTCCCTTTCCCCC 350
Cb16 CCCGGGATCCCACTCCCTTCTCCCTCGCCTCTGTCTTCCATGCATGCGCGGGCGCTCCTCCCTTTCCCCC 350
Cb25 CCCGGGATCCCACTCCCTTCTCCCTCGCCTCTGTCTTCCATGCATGCGCGGGCGCTCCTCCCTTTCCCCC 350
     ********************************************************************

Cb10 CACGCACCCGCCGGCATGGTTCAATGACCCTTGAACGTTCACGGGGAGAGAGTTCTGGGGCGCCCTCCCC 420
Cb16 CACGCACCCGCCGGCATGGTTCAATGACCCTTGAACGTTCACGGGGAGAGAGTTCTGGGGCGCCCTCCCC 420
Cb25 CACGCACCCGCCGGCATGGTTCAATGACCCTTGAACGTTCACGGGGAGAGAGTTCTGGGGCGCCCTCCCC 420
     ********************************************************************
                                                        5.8s rRNA
Cb10 CGAGGTTTTTCCAACACACTGTTGATCTTTTGTGGACTCTTAGCAGTGGATCACTTGCTCGTGTGTCGAT 490
Cb16 CGAGGTTTTTCCAACACACTGTTGATCTTTTGTGGACTCTTAGCAGTGGATCACTTGCTCGTGTGTCGAT 490
Cb25 CGAGGTTTTTCCAACACACTGTTGATCTTTTGTGGACTCTTAGCAGTGGATCACTTGCTCGTGTGTCGAT 490
     ********************************************************************

Cb10 GAAGAACGC 499
Cb16 GAAGAACGC 499
Cb25 GAAGAACGC 499
     *********
```

Microsatellite repeats: Cb1-1

Random repeats: Cb1-A

Figure 10. Multiple sequence alignment of *C. batrachus* ITS1 (Cb1) showing the position of microsatellite repeat Cb1-1 and random repeat Cb1-A.

Cg: *Clarias gariepinus*

```
                                                                  18s rRNA   ITS1
                                                                                                  Cg1-1
Cg06 AAAANTCGTACAGGTTTTCGTAGATAAACCTGCGGAAGGATCATTAACGGGTAGCGT-----GGGGGGCG 70
Cg07 -----NNGTACAGGTTTTCGTAGATAAACCTGCGGAAGGATCATTAACGGGTAGCGT-----GGGGGGCG 70
Cg11 ---AAANGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTAGCGCCCCGCGGGGGGCG 70
            ********** ***** * ******************************      ********

                Cg1-2                                                                  Cg1-3
Cg06 CGCTGCCCTCTCCTCT-AACCCTCGG--------------------------------CCCCCAGCCC-- 140
Cg07 CGCTGCCTTCTCCTCT-AACCCTCGG--------------------------------CCCCCAGCCC-- 140
Cg11 CGCCTCCTCCTCCTCTTAACCCTCGGGACGGAAACCGAGGCTGGCCCGGGGGCTCACACCCCCGGCCCCC 140
     ***  **  ******* *********                                 ***** ****

Cg06 -GCCGAGGCCTTCCCG----CTGACCGGGAAC---GAAAGTTGGGGCGAGGGTANCTGCTGCTCGAGCGT 210
Cg07 -GCCGAGGCCTTCCCG----CTGACCGGGAAC---GAAAGTTGGGGCGAGGGTATGTGCTGCTCGAGCGT 210
Cg11 CGCCGAGGCCTTCCCGAAACCTTTCCGGTGACCGTAAAAGATAGGGCGAGGGTACCTGCTGCTCGTGCGT 210
      ***************    **  ****  **    **** * ***********  ********* ****

Cg06 GGCGCGTTTTTTTGGTGGG---GCTGGC----TCGCTTCTCCC--TCGCCTCTTTCTTCCATGTGCGCAC 280
Cg07 GGCGCGTTTTTTTGGTGGG---GCTGGC----TCGCTTCTCCC--TCGCCTCTTTCTTCCATGTGCGCAC 280
Cg11 GGCGTGGTGTGCGTGCGTGCGTGCTGGTGGTGCCGGTGGTGACGGTAGGTGCTGGCCGTGGCGGGCGCGC 280
     **** * * * Cg1-4 * * *   *****      ** *  *                                Cg1-A

Cg06 GCGCCTC---CCTTCTGTTCCA-------CCCCCATGTTCACTAGGGAGAGTTCCGGNGCGCCCTCCCCT 350
Cg07 GCGCCTC---CCTTCTGTTCCA-------CCCCCATGTTCACTAGGGAGAGTTCCGGGGCGCCCTCCCCT 350
Cg11 GCCGGTCGGGCTGGGAGGGCCGAGCGGAGTCTGGATGGATGGTGCGGAGGGTGAAAAGGCGCTACGCGGA 350
     **   **   *     *  **         *   ***     *  **** **      ****    *

Cg06 AAGG------TTTGTACATACATTGTGTTGCATCTTTGTGGACGATGGCTTTCGAACCTAAAAAAC-AAC 420
Cg07 AAGG------TTTGTACATACATTGTGTTGCATCTTTGTGGACGATGGCTTTCGAACCTAAAAAAC-AAC 420
Cg11 GAAAGTGCTCTCTGTGCGGGCGTCATG--GGATCCCTCTCTCTCCCTCTCTCCGTTCTCACTCGCCTGGC 420
      *        * *** *   * *  **  * ***  * *     Cg1-5     * **  *  *     *   *

Cg06 TGACAACTCTTTGCGGTGGATCACTCAGCTCATGCATCAATGAAGAACGCAGCCAGCTGGGGCGGGCTGT 490
Cg07 TGACAACTCTTTGCGGTGGATCACTCAGCTCATGCATCAATGAAGAACGCAGCCAGCTGAAA-------- 482
Cg11 TGCCCCTTCCGTGCGCGCGCGCGCGCGTGCCTCTCCTTCCTCAGCCACCCATCCCCGCGCACGACCGCCC 490
     ** *   **  **** Cg1-6*  * * *    *Cg1-B* *   * *   ** ** **    *   Cg1-C

Cg06 AGACGGTCGATATCTTTATTGGGCGATGGTAAAATTGAAAC----------------------------- 531
Cg07 ------------------------------------------------------------------------
Cg11 CCCGCCTCGCCGGCGTGGTTCAATGACCCCGAGTTCTCTCCCGCCCACCCCGTTCGCGGGGGCGGGTGAG 560
                                                Cg1-D                Cg1-E
Cg06 ------------------------------------------------------------------------
Cg07 ------------------------------------------------------------------------
Cg11 TAGAGTTCCGGGGCGCCCTCCCCTGAGGTTTTTAACACACACTGTGTTGCATTCTCTGCTGGACGATGGC 630
                  Cg1-F                      Cg1-G

Cg06 -------------------------------------
Cg07 -------------------------------------
Cg11 TTACCTAAAAAAAACCAAAAATGACAACTCTTAGCGG 667
                Cg1-7
                                 5.8s rRNA
```

Microsatellite repeats: Cg1-1, Cg1-2, Cg1-3, Cg1-4, Cg1-5, Cg1-6, Cg1-7

Random repeats: Cg1-A, Cg1-B, Cg1-C, Cg1-D, Cg1-E, Cg1-F, Cg1-G

Figure 11. Multiple sequence alignment of *C. gariepinus* ITS1 (Cg1) showing the position of microsatellite repeat Cg1-1 to Cg1-7 and random repeats Cg1-A to Cg1-G.

Hf: *Heteropneustes fossilis*

```
                                                   18s rRNA  ITS1                                  Hf1-A
Hf40 --AANCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTAGCACCCCGCCGGGGGTGC 70
Hf44 AAANTCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTAGCACCCCGCCGGGGGTGC 70
Hf42 -AAANCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTAGCACCCCGCCGGGGGTGC 70
       *  ************************************************************

                                   Hf1-B                          Hf1-C
Hf40 GCCGCCTCCCATCATTACCCTCGAGAGAGCCCCGTGCCTTCTCGAAACCCTTCCGGTGACCGATGAAACA 140
Hf44 GCCGCCTCCCATCATTACCCTCGAGAGAGCCCCGTGCCTTCTCGAAACCCTTCCGGTGACCGATGAAACA 140
Hf42 GCCGCCTCCCATCATTACCCTCGAGAGAGCCCCGTGCCTTCTCGAAACCCTTCCGGTGACCGATGAAACA 140
     ******************************************************************

                                         Hf1-D
Hf40 ATGGGCGAGGGTACCTGCTGCTCGAGCGCCGCGGGTGGGGGCGCGTGCCGTGCGTAGAGGCCAAGAGTCT 210
Hf44 ATGGGCGAGGGTACCTGCTGCTCGAGCGCCGCGGGTGGGGGCGCGTGCCGTGCGTAGAGGCCAAGAGTCT 210
Hf42 ATGGGCGAGGGTACCTGCTGCTCGAGCGCCGCGGGTGGGGGCGCGTGCCGTGCGTAGAGGCCAAGAGTCT 210
     ******************************************************************

                 Hf1-1
Hf40 GGGGGTGTGTGTGTGAGAGAAAGGCGCCACGCGGAAGGCCCAGTGCCTCCGTGCGGGCGGCTGGATCGAC 280
Hf44 GGGGGTGTGTGTGTGAGAGAAAGGCGCCACGCGGAAGGCCCAGTGCCTCCGTGCGGGCGGCTGGATCGAC 280
Hf42 GGG--TGTGTGTGTGAGAGAAAGGCGCCACGCGGAAGGCCCAGTGCCTCCGTGCGGGCGGCTGGATCGAC 280
     ***  ***************************************************************

                Hf1-2                        Hf1-3                      Hf1-4
Hf40 CACTCTCTCCTTCTCTATCGCCCCTCACACACGCGCGCGCGC--CCACTCAACCCCCC-GCGCTGTCGGC 350
Hf44 CACTCTCTCCTTCTCTATCGCCCCTCACACACGCGCGCGCGC--CCACTCAACCCCCC-GCGCTGCCGGC 350
Hf42 CACTCTCTCCTTCTCTATCGCCCCTCACACACGCGCGCGCGCGCCCACTCAACCCCCCCGCGCTGCCGGC 350
     ****************************************  ************** ****** ****

                                                                Hf1-E
Hf40 GTGGTTCAATGACCCCGAGTTCTATCTCCGGTGACGGGGAGAGTTCCGGGGCGCCCTCCCCGGATGGTTT 420
Hf44 GTGGTTCAATGACCCCGAGTTCTATCCCCGGTGACGGGGAGAGTTCCGGGGCGCCCTCCCCGGATGGTTT 420
Hf42 GTGGTTCAATGACCCCGAGTTCTATCCCCGGTGACGGGGAGAGTTCCGGGGCGCCCTCCCCGGATGGTTT 420
     ************************** ***************************************

                                                                   5.8s rRNA
                        Hf1-5                          Hf1-6
Hf40 ATACACAGTCGATATCTTTTTTGGGCGATGGCAAAATTGAAACAAAAAAACGAAACTGACAACTCTTAGC 490
Hf44 ATACACAGTCGATATCTTTTTTGGGCGATGGCAAAATTGAAACAAAAAAACGAAACTGACAACTCTTAGC 490
Hf42 ATACACAGTCGATATCTTTTTTGGGCGATGGCAAAATTGAAACAAAAAAACGAAACTGACAACTCTTAGC 490
     ******************************************************************

Hf40 GGTGGATCACTCGGCTCGTGCGTCGATGAAAAACC--AGCCCAGCTGAGA- 541
Hf44 GGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGCAGGCCCAGCTGAA-- 541
Hf42 GGTGGATCACTCGGCTCGTGCGTCGATGAAGAACGC--AGCCAGCTGAAGC 541
     ****************************** ***      ********
```

Microsatellite repeats: Hf1-1, Hf1-2, Hf1-3, Hf1-4, Hf1-5, Hf1-6

Random repeats: Hf1-A, Hf1-B, Hf1-C, Hf1-D, Hf1-E

Figure 12. Multiple sequence alignment of *H. fossilis* ITS1 (Hf1) showing the position of microsatellite repeats Hf1-1 to Hf1-6 and random repeats Hf1-A to Hf1-E.

Ss: *Sperata seenghala*

```
                                                                 18s rRNA   ITS1
Ss02 AAANTCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTTAGCGTCCGCGGGGACGCG 70
Ss11 AAANTCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTTAGCGTCCGCGGGGACGCG 70
Ss09 AAAATCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGTTAGCGTCCGCGGGGACGCG 70
Ss01 AAAATCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGGTAGCGTCCGCGGGGACGCG 70
     *** ********************************************** ******************

                                                 Ss1-A
Ss02 CCTGCTCCTCGACACACGGCGCCGAGGCTCTCGCTCTTT-CCCCTCGACAAAAGGAGAAACGTGAGCCGA 140
Ss11 CCTGCTCCTCGACACACGGCGCCGAGGCTCTCGCTCTTT-CCCCTCGACAAAAGGAGAAACGTGAGCCGA 140
Ss09 CCTGCTCCTCGACACACGGCGCCGAGGCTCTCGCTCTTT-CCCCTCGACAAAAGGAGAAACGTGAGCCGA 140
Ss01 CCTGCTCCTCGACACACGGCGCCGAGGCTCGCGCTTTTTTCCCCTTGACAAAAGGAGAAACGTGAGCCGA 140
     ****************************** **** *** ***** ************************

            Ss1-B                                Ss1-C
Ss02 GGTCCCTTTTACGGTATCCCAAGAAGACGGGGGAAAAAGAGAAAGGGGACGAGAAAAGGGAATCGGAAGG 210
Ss11 GGTCCCTTTTACGGTATCCCAAGAAGACGGGGGAAAAAGAGAAAGGGGACGAGAAAAGGGAATCGGAAGG 210
Ss09 GGTCCCTTTTACGGTATCCCAAGAAGACGGGGGAAAAAGAGAAAGGGGACGAGAAAAGGGAATCGGAAGG 210
Ss01 GGTCCCTTTTACGGTATCCCAAGAAGACGGGGGAAAAAGAGAAAGGGGACGAGAAAAGGGAATCGGAAGG 210
     ************************************************************************

                 Ss1-1
Ss02 GGGAAAAGAAAGAAAGAAAGAAAGACGGACGGGGCGGAGGGTACCTACCGCTCGCCGGCGCTACGCGGAG 280
Ss11 GGGAAAAGAAAGAAAGAAAGAAAGACGGACGGGGCGGAGGGTACCTACCGCTCGCCGGCGCTACGCGGAG 280
Ss09 GGGAAAAGAAAGAAAGA--------CGGACGGGGCGGAGGGTACCTACCGCTCGCCGGCGCTACGCGGAG 280
Ss01 GGGAAAAGAAAGAAAGAA----AGACGGACGGGGCGGAGGGTACCTACCGCTCGCCGGCGCTACGCGGAG 280
     *****************        ********************************************

                 Ss1-2                        Ss1-3                   Ss1-4
Ss02 GTCCTTTGGGGGTAGGGGGG-TGA------GACAGACGGGGCTCCTCGTCCCTTCCTTCCTTCC----CT 350
Ss11 GTCCTTTGGGGGTAGGGGGG-TGA------GACGGACGGGGCTCCTCGTCCCTTCCTTCCTTCC----CT 350
Ss09 GTCCTTTGGGGGTAGGGGGGGTGA------GACGGACGGGGCTCCTCGTCCCTTCCTTCCTTCCTTCCCT 350
Ss01 GTCCTTTGGGGGTAGGGGGTGAGACGGACGGACGGACGGGGCTCCTCGTCCCTTCCTTCCCTCG----CT 350
     ******************   **       *** ************************** **     **

                                                                Ss1-5     Ss1-6
Ss02 CCCTCCCTCTCTCCCTCTGTGCGGGCGCCTGGGCCGAAAGAGGTTCGAAGACCCCTCCCCTCGCGGGGTG 420
Ss11 CCCTCCCTCTCTCCCTCTGTGCGGGCGCCTGGGCCGAAAGAGGTTCGAAGACCCCTCCCCTCGCGGGGTG 420
Ss09 CCCTCCCTCTCTCCCTCTGTGCGGGCGCCTGGGCCGAAAGAGGTTCGAAGACCCCTCCCCTCGCGGGGTG 420
Ss01 CCCTCGCTCCCTCCCTCTGTGCGGGCGCCTGGGCCGAAAAAGGTTCGAAGACCCCTCCCCTCGCGGGGTG 420
     ***** *** ***************************** ******************************

                                Ss1-D                   Ss1-7
Ss02 GGGCGCCCTCCACTCGTTCTTTTCTTTATTTCTCTCCCCCACCCCGACCCCCGGATCCTGGTTCTACCCT 490
Ss11 GGGCGCCCTCCACTCGTTCTTTTCTTTATTTCTCTCCCCCACCCCGACCCCCGGATCCTGGTTCTACCCT 490
Ss09 GGGCGCCCTCCACTCGTTCTTTTCTTTATTTCTCTCCCCCACCCCAACCCCCGGATCCTGGTTCT-CCCT 490
Ss01 GGGCGCCCTCCACTCGGTCTTTTCTTTATTTTTCTCCCCCACCCCACACCCCGGATCCTGGTTCT-CCCT 490
     **************** ************** *************   ***************** ****

            Ss1-8                                Ss1-9
Ss02 TTCTCTCTCGTCTCCCCC-GATCCTCCAAGACCCGTATTTTTTTCATTCCCATCGTGTGGCTCGGACGAA 560
Ss11 TTCTCTCTCGTCTCCCCC-GATCCTCCAAGACCCGTATTTTTTTCATTCCCATCGTGTGGCTCGGACGAA 560
Ss09 TTCTCTCTCGTCTCCCCC-GATCCTCCAAGACCCGTATTTTTTTCATTCCCATCGTGTGGCTCGGACGAA 560
Ss01 TTCTCTCTCGTCTCCCCCCGATCCTCCAAGACCCGTATTCTTT--ATTCCCATCGGGTGGCTCGGACAAA 560
     ****************** ******************** ***  ********** *********** **

                                   Ss1-10                      5.8s rRNA
Ss02 CCGTCTCGGTGAAATAAACGAGAGAAAAAAA-CGAACCATACATAG---------------------- 630
Ss11 ACGTCTCGGTGAAATAAACGAGAGAAAAAAA-CGAAGCATAAAGAGCTGACAACTCTTAGCGGTGGATCA 630
Ss09 ACGTCTCGGTGAAATAGACGAGAGAAAAAAAACGAAGCATAAAGAGCTGACAACTCTTAGCGGTGGATCA 630
Ss01 ACGTCTCGGTGAAAGAAACGAGAGAAAAAAA-CCAAGCATAAAGAGCTGACAACTCTTAGCGGTGGATCA 630
      ************* * ********** *** * ** **** * **

Ss02 ------------------------------------- 668
Ss11 CTCGGCTCGTGCGTCGATGAAGAACGCAGGCCAGCTGA 668
Ss09 CTCGGCTCGTGCGTCGATGAAGAACGCAGCCCAGCGGA 668
Ss01 CTCGGCTCGTGCGTCGATGAAGAAGCAGCCCCGGCGGA 668
```

Microsatellite repeats: Ss1-1, Ss1-2, Ss1-3, Ss1-4, Ss1-5, Ss1-6, Ss1-7, Ss1-8, Ss1-9, Ss1-10

Random repeats: Ss1-A, Ss1-B, Ss1-C, Ss1-D

Figure13. Multiple sequence alignment of *S. seenghala* ITS1 (Ss1) showing the position of microsatellite repeats Ss1-1 to Ss1-10 and random repeats Ss1-A to Ss1-D.

Wa: *Wallago attu*

```
                                                  18s rRNA   ITS1
Wa42 -AAAATCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGGCAGCCGCCGTCGGGGCGA 70
Wa32 --AAANCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGGCAGCCGCCGTCGGGGCGA 70
Wa51 --AANTCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGGCAGCCGCCGTCGGGGCGA 70
Wa71 AAAAANCGTACAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAACGGGGCAGCCGCCGTCGGGGCGA 70
       **  ************************************************************

                                                  Wa1-A
Wa42 GCGCCTTCTCACCCTCCTCCTCACGCGACGTGCCGAAGGCCCGGGGGCAGTCGCTCCCGGCCCGAGGCGC 140
Wa32 GCGCCTTCTCACCCTCCTCCTCACGCGACGTGCCGAAGGCCCGGGGGCAGTCGCTCCCGGCCCGAGGCGC 140
Wa51 GCGCCTTCCCACCCTCCTCCTCACGCGACGTGCCGAAGGCCCGGGGGCAGTCGCTCCCGGCCCGAGGCGC 140
Wa71 GCGCCTTCTCACCCTCCTCCTCACGCGACGTGCCGAAGGCCCGGGGGCAGTCGCTCCCGGCCCGAGGCGC 140
     ******** ***********************************************************

                                                                  Wa1-B
Wa42 ATTGCTTTCGGTCTCCTTCGATCAAAAAACAAAGGGAAGGGGAAAGCAAAGCAAGGTCAAAAACCTTTGG 210
Wa32 ATTGCTTTCGGTCTCCTTCGATCAAAAAACAAAGGGAAGGGGAAAGCAAAGCAAGGTCAAAAACCTTTGG 210
Wa51 ATTGCTTTCGGTCTCCTTCGATCAAAAAACAAAGGGAAGGGGAAAGCAAAGCAAGGTCAAAAACCTTTGG 210
Wa71 ATTGCTTTCGGTCTCCTTCGATCAAAAAACAAAGGGAAGGGGAAAGCAAAGCAAGGTCAAAAACCTTTGG 210
     ************************************************************************

Wa42 GAAAGAGAGAAAAAAG                                                      226
                                  Wa1-1                                Wa1-2
Wa32 GAAAGAGAGAAAAAAGATAAAAAAAGAAAACGAGGAAAAAAAAAAAGAAAAAAGAAAGAAAAACACCCCC 280
Wa51 GAAAGAGAGAAAAAAGATAAAAAA-GAAAACGAGGGAAAAAAAAAAAAAAAAAANNAAAAAAAACCCCCC 280
Wa71 GAAAGAGAGAAAAAA                                                      225
     ********************** ********** ********** ******   * *****  *****

Wa42                          Wa1-C                                Wa1-3
Wa32 CCTCCTCCGGCAAGGGTACCTACCGGCCGGACAGAACGGGCTCAGGAATGAAAACAAAAAGTGAATCGAT 350
Wa51 CCCCCC
Wa71
     ** **

Wa42                                                                  Wa1-E
                 Wa1-D
Wa32 TCTCTTTTTATTTCCTCCGTCCCACTGTCGGGCCCAGGTTCAATGACCCCGGCCGACCCGGGGCGCCCTT 420
Wa51
Wa71

Wa42                                                  Wa1-4
Wa32 GCGATGGAGCCGTTGTGATATTTCTTTTTTTCTTTTTTGACTTTTTTTTTCTTTCTTATTCTTCTTCTTC 490
Wa51
Wa71

Wa42                                                                  Wa1-F
Wa32 TCTCCCCTTCTTTTTTGTTTTGCTGCCGGTTTCCCCGCCATTCCCTTTGTTATATTTTGTTGTTTCTCTT 560
Wa51
Wa71

Wa42                                          Wa1-5
Wa32 TACCCTTCTCTGGCTTGGGCAAAGGTAAAACAAAACCAAAAAAAAAAAA 609
Wa51
Wa71
```

Microsatellite repeats: Wa1-1, Wa1-2, Wa1-3, Wa1-4

Random repeats: Wa1-A, Wa1-B, Wa1-C, Wa1-D, Wa1-E, Wa1-F

Figure 14. Multiple sequence alignment of *W. attu* ITS1 (Wa1) showing the position of microsatellite repeats Wa1-1 to Wa1-5 and the random repeats Wa1-A to Wa1-F.

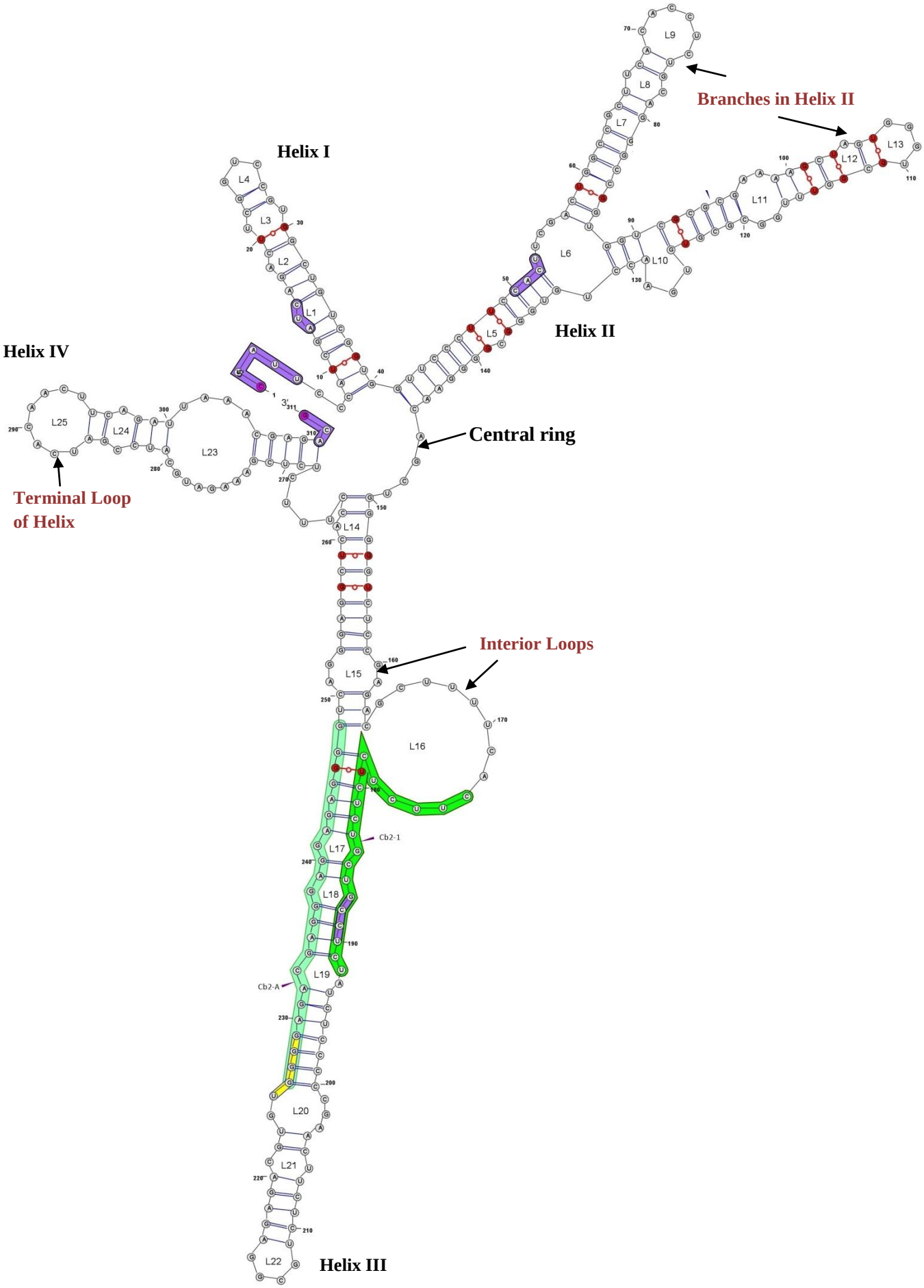


Figure 15A. ITS2 secondary structure of *Clarias batrachus*. The 4-helix secondary structure showing the central ring with its 4 helices from Helix I to Helix IV arranged in clockwise direction. The helices are interrupted with loops (L) of variable size labeled as L1 to L25. The annotation was done with software VARNA.

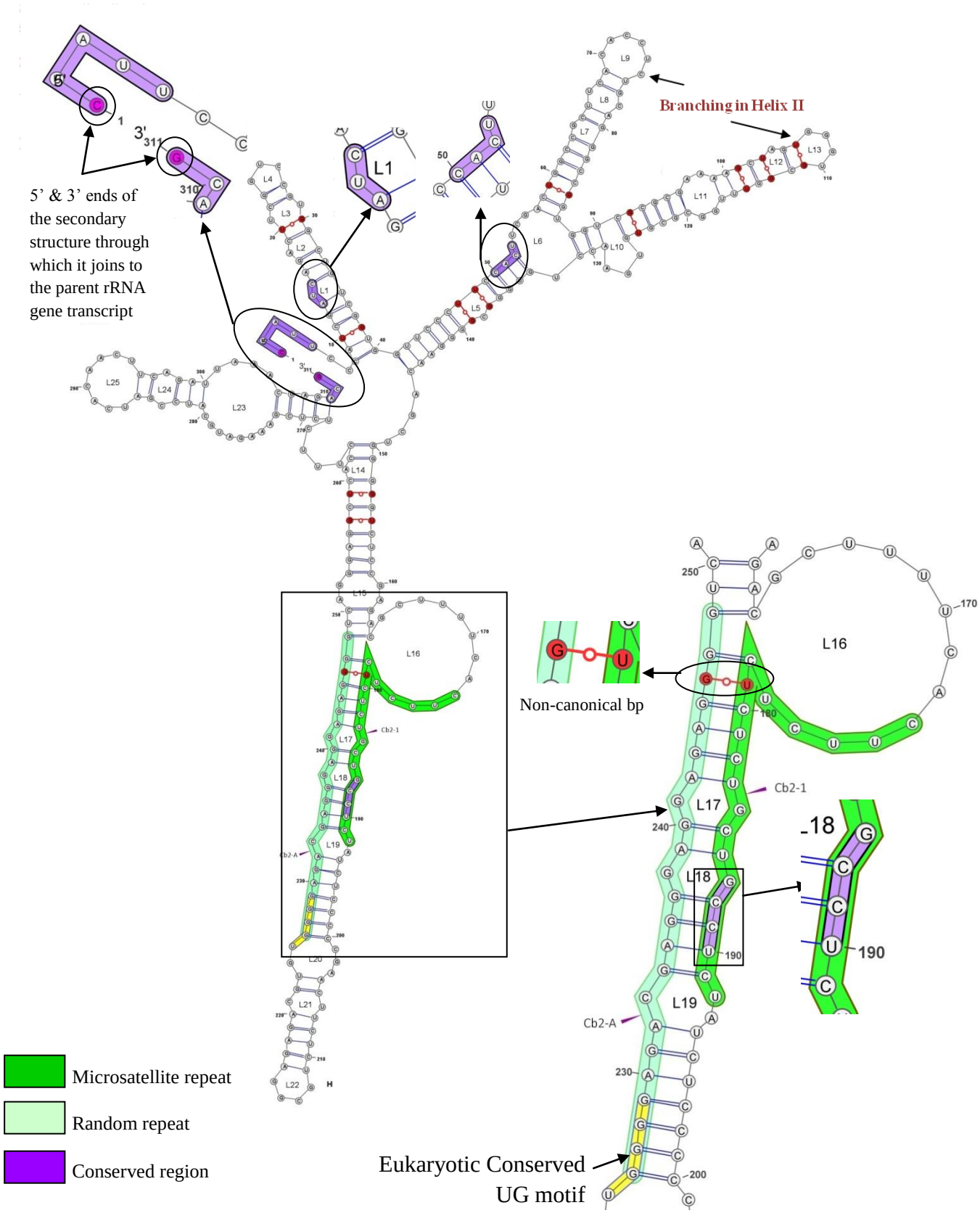


Figure 15B. ITS2 secondary structure of *Clarias batrachus.* Microsatellite repeats, Random repeats, eukaryotic conserved motifs, catfish conserved motifs, non-canonical bp are shown in magnified view round the secondary structure.

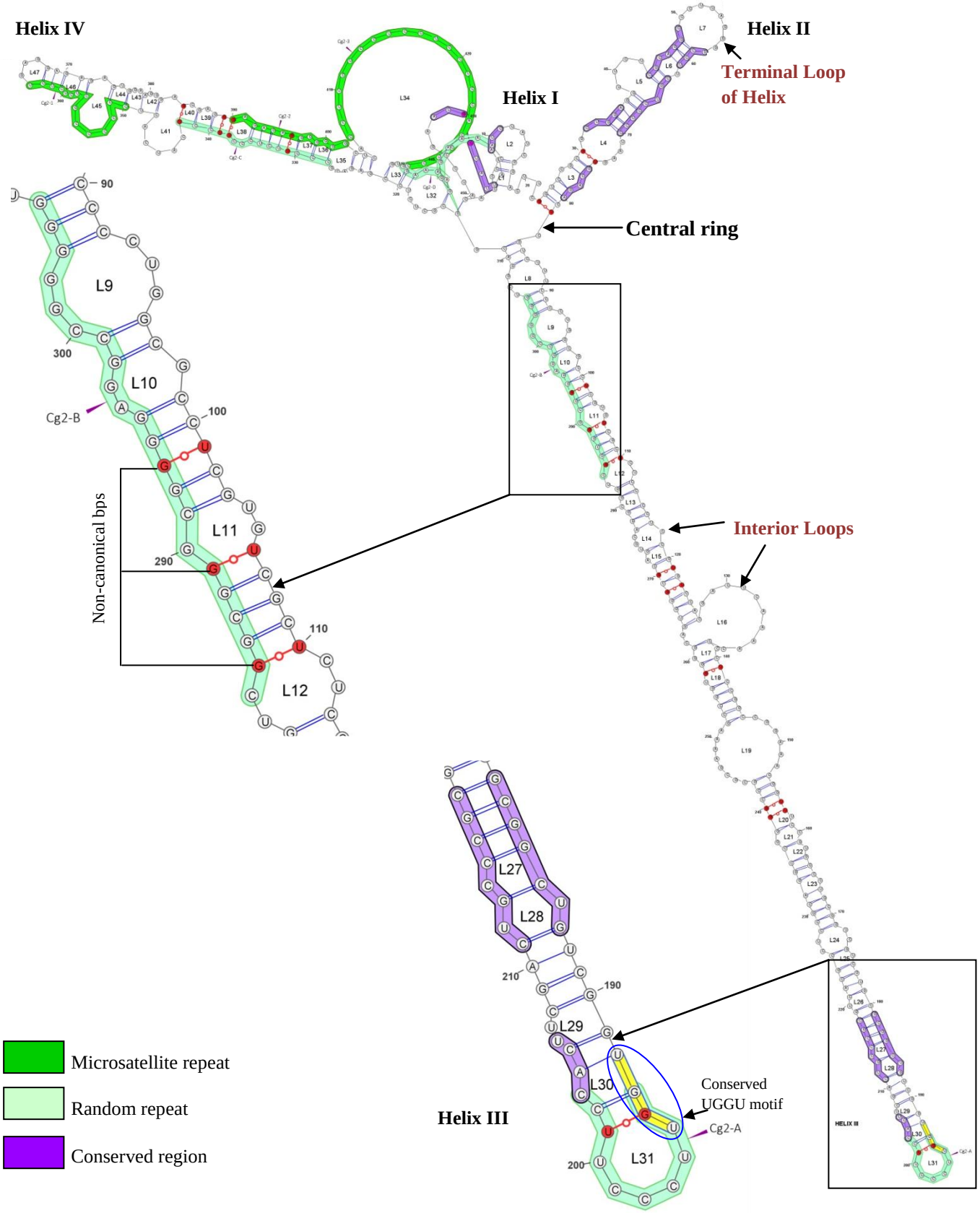


Figure 16A. ITS2 secondary structure of *Clarias gariepinus*. The 4-helix secondary structure is showing the central ring with its 4 helices from Helix I to Helix IV arranged in clockwise direction. The helices are interrupted with loops (L) of variable size labeled as L1 to L34. Microsatellite repeats, Random repeats, eukaryotic conserved motifs, catfish conserved motifs, non-canonical bp are shown in magnified view round the secondary structure.

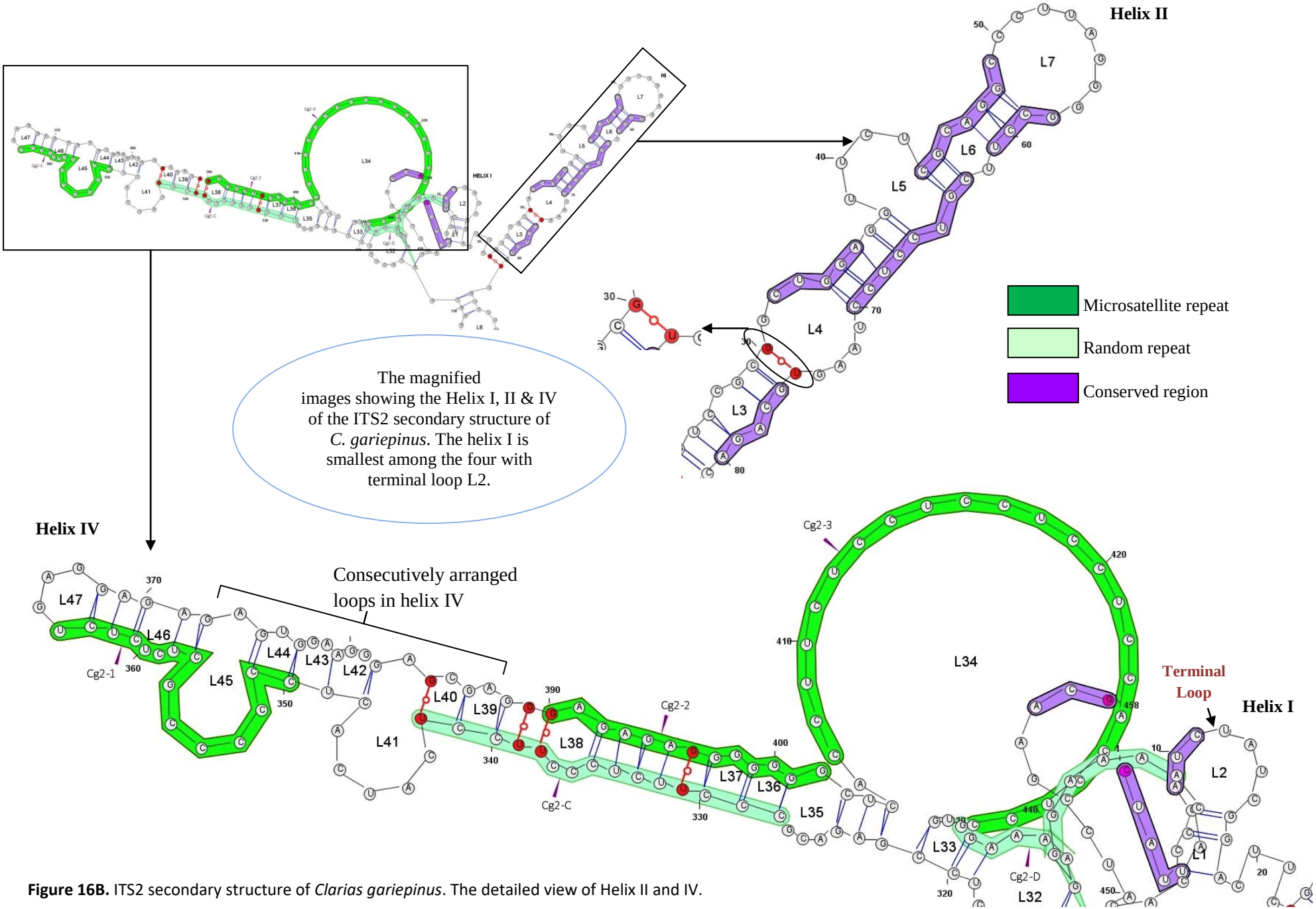


Figure 16B. ITS2 secondary structure of *Clarias gariepinus*. The detailed view of Helix II and IV.

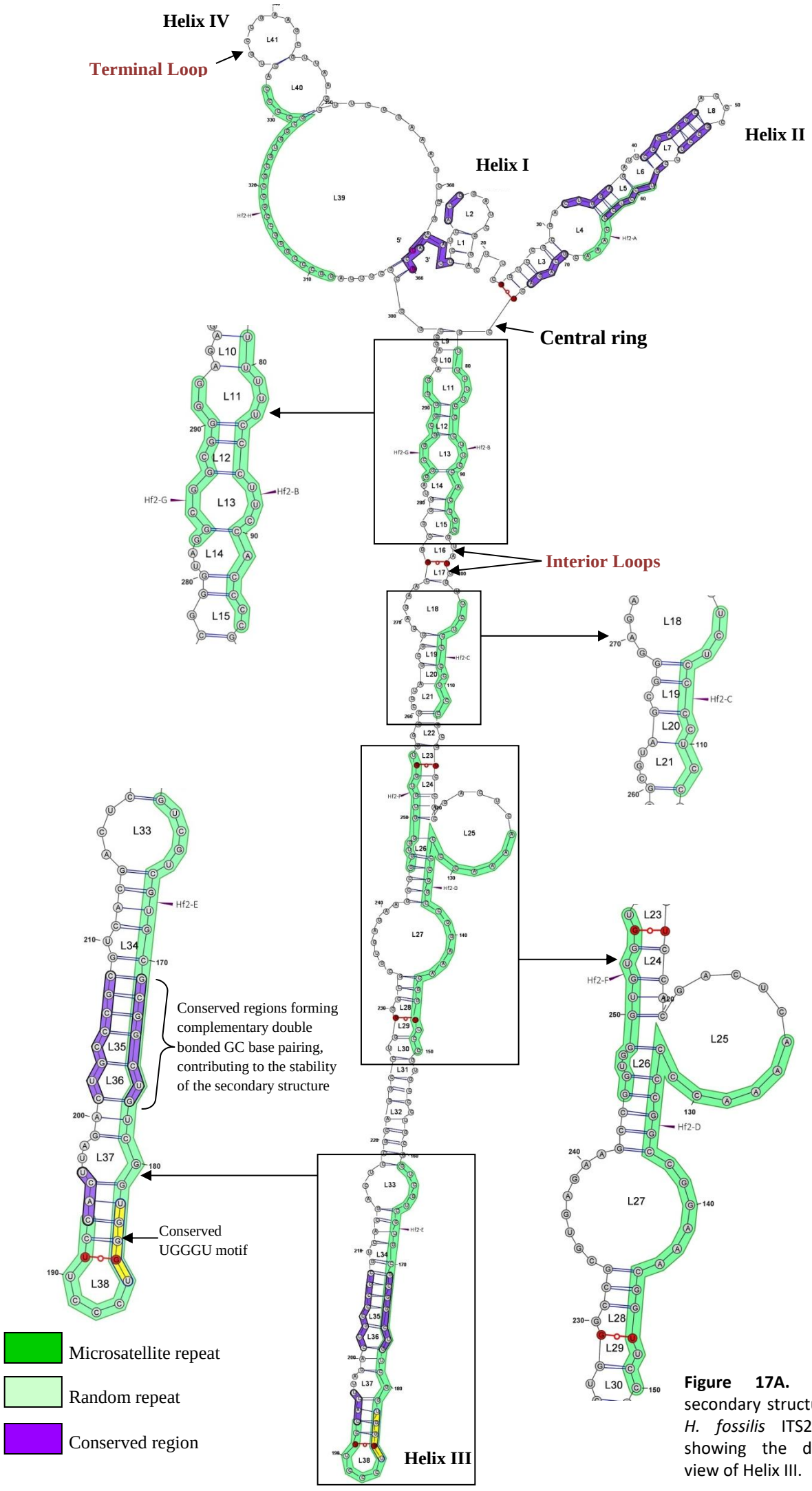


Figure 17A. 4-helix secondary structure of *H. fossilis* ITS2. Inset showing the detailed view of Helix III.

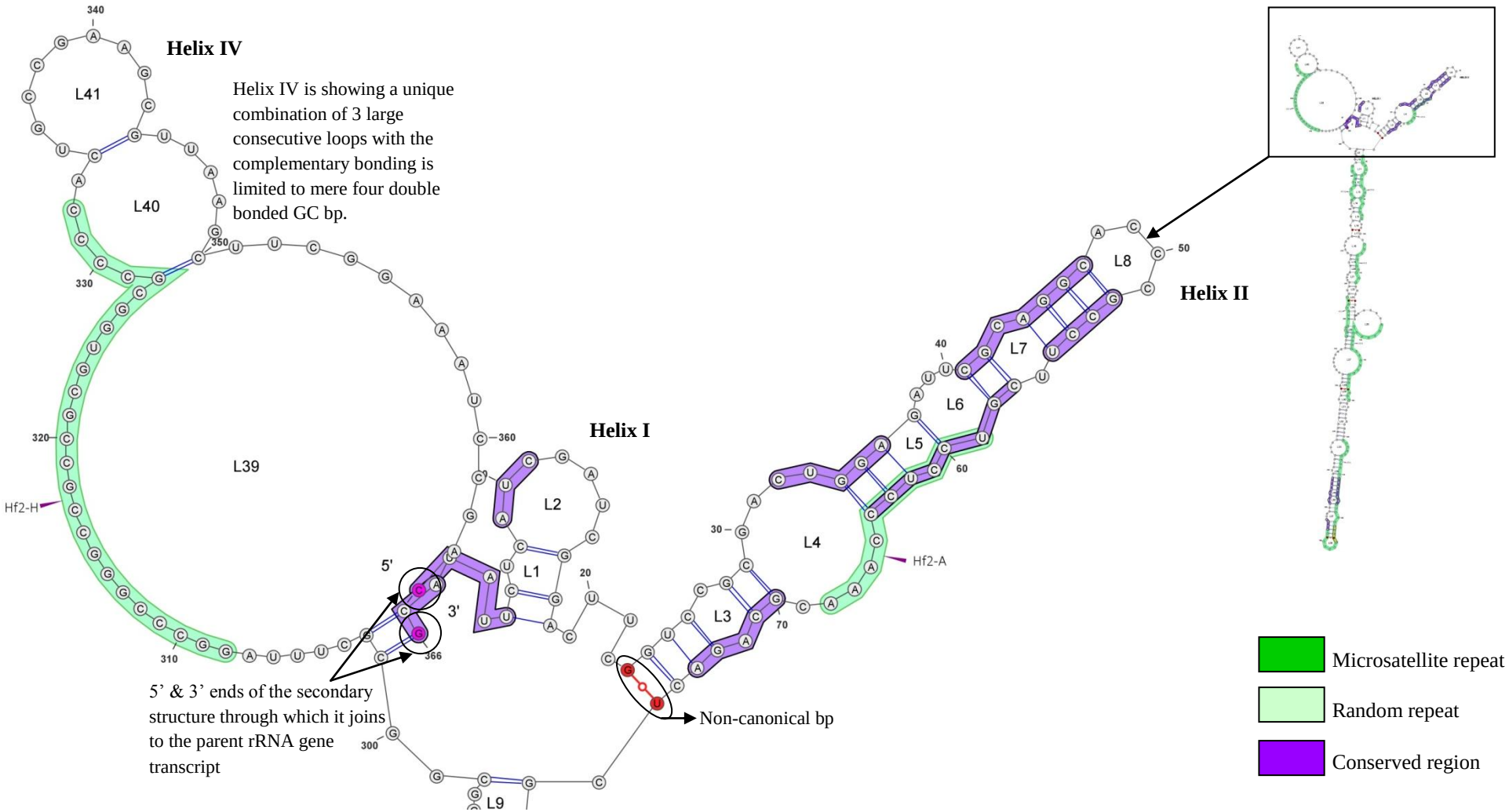


Figure 17B. ITS2 secondary structure of *H. fossilis.* Inset showing the detailed view of central ring and the helix I, II and IV.

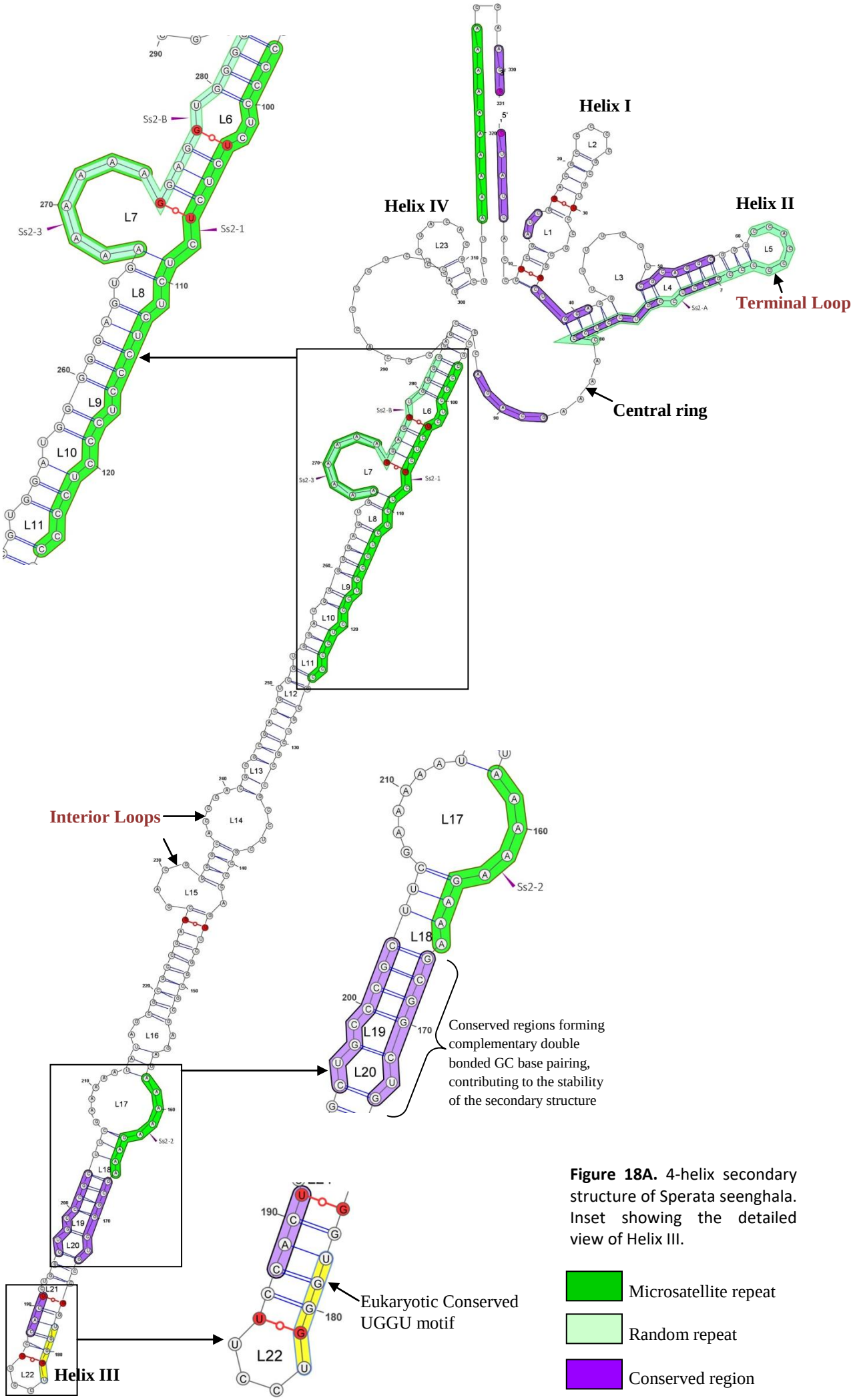


Figure 18A. 4-helix secondary structure of Sperata seenghala. Inset showing the detailed view of Helix III.

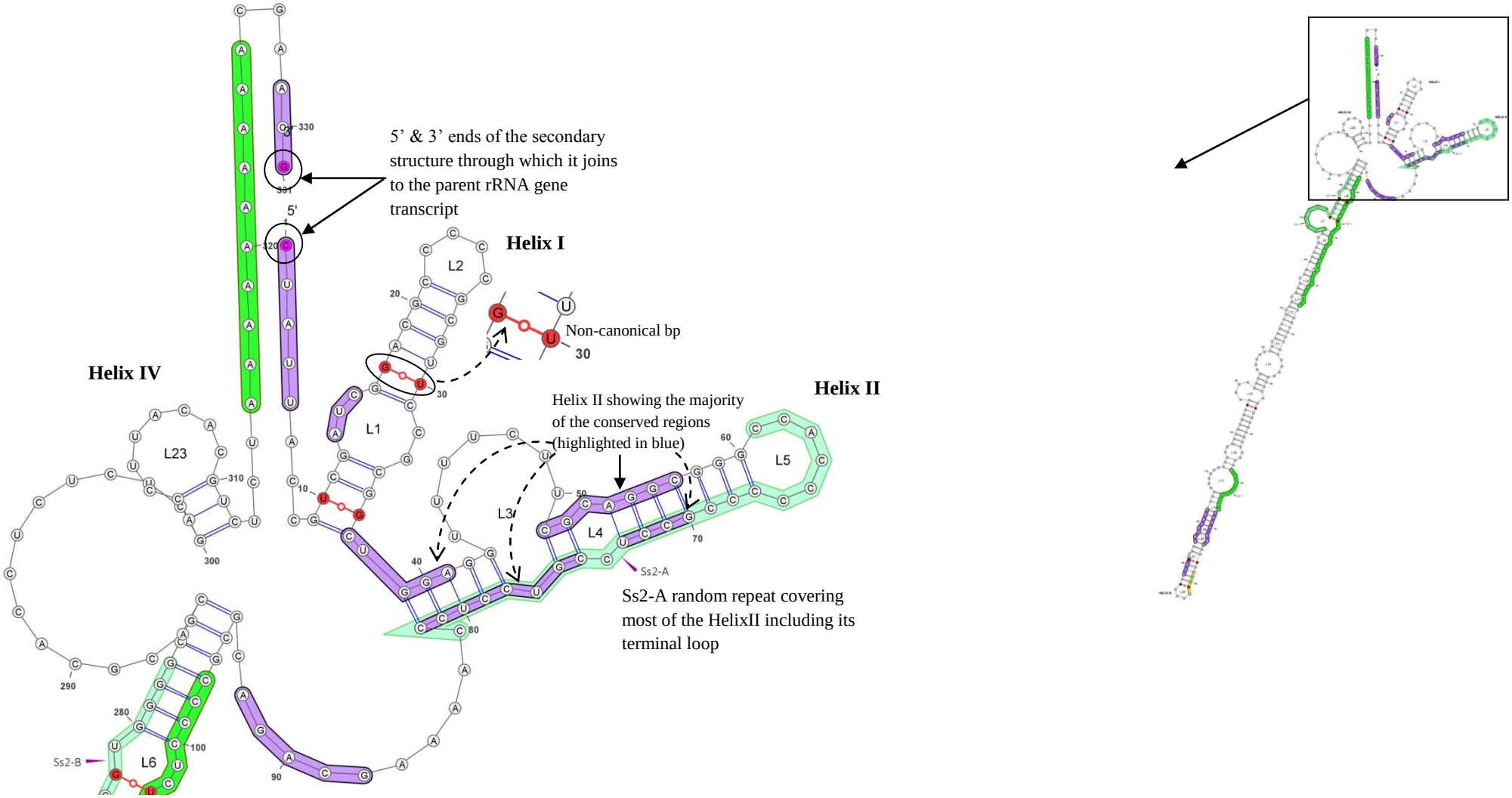


Figure 18B. 4-helix secondary structure of *Sperata seenghala* ITS2. Inset showing the detailed view of central ring and the helix I, II and IV.

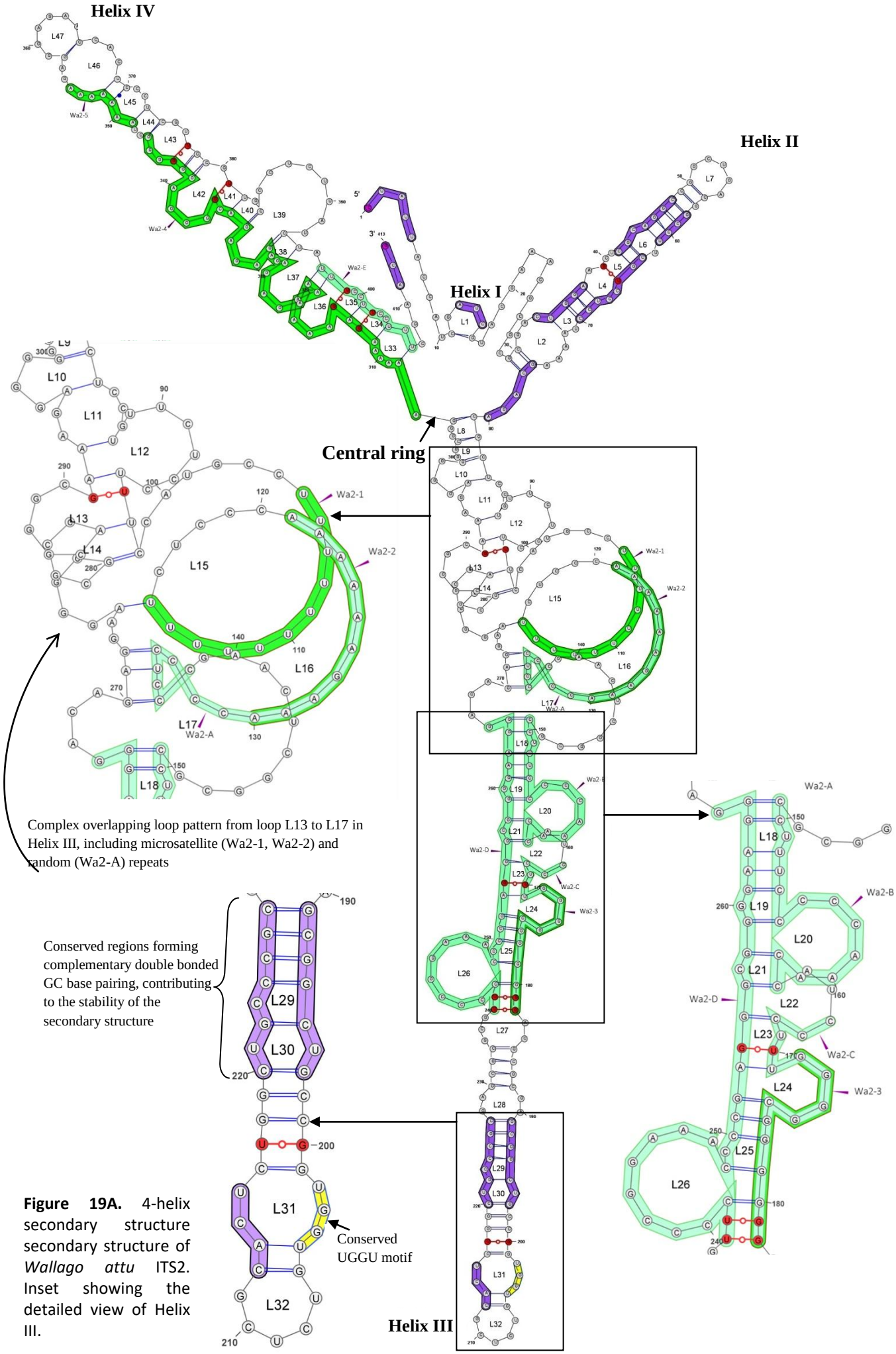


Figure 19A. 4-helix secondary structure secondary structure of *Wallago attu* ITS2. Inset showing the detailed view of Helix III.

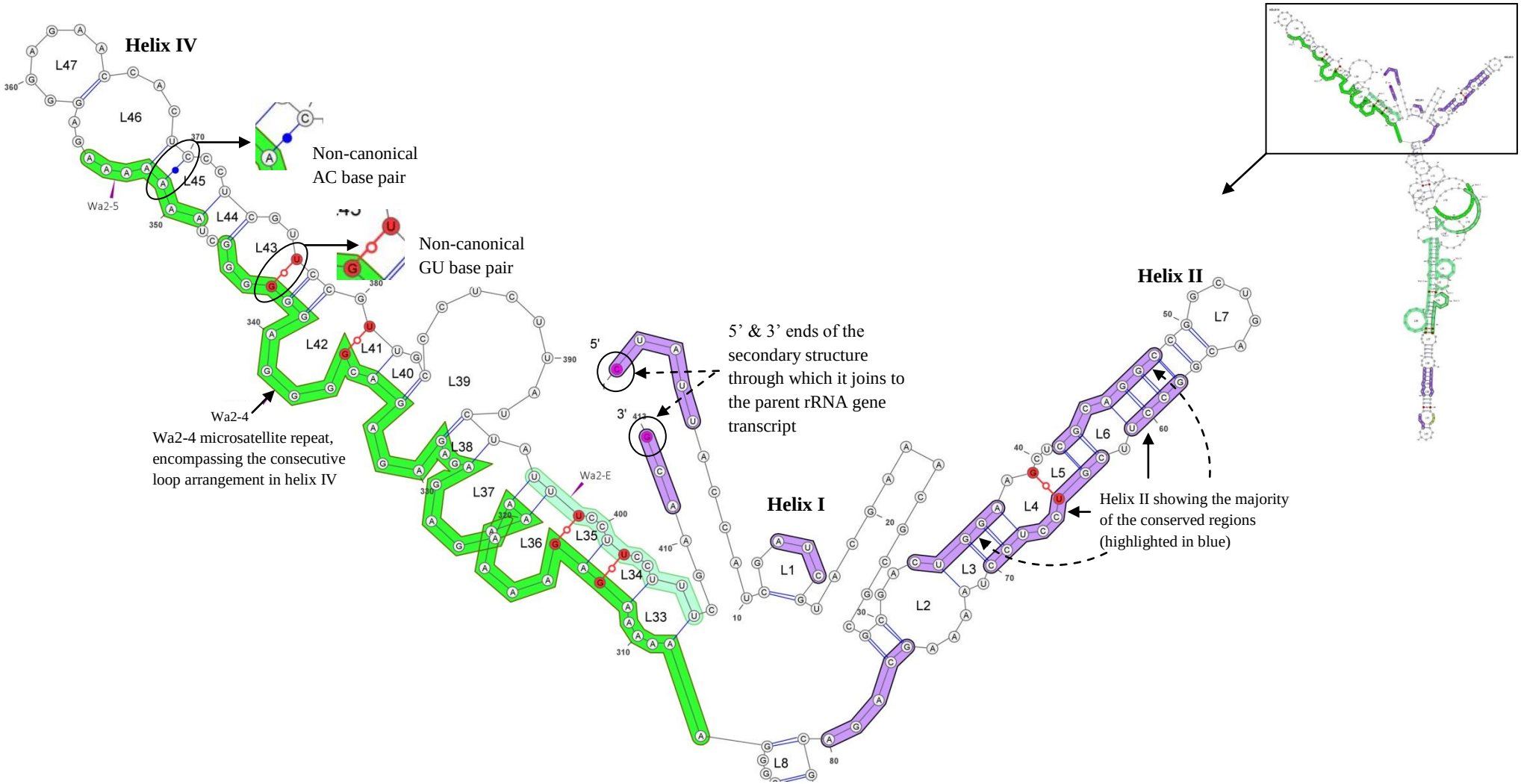


Figure 19B. 4-helix secondary structure of *Wallago attu* ITS2. Inset showing the detailed view of central ring and the helix I, II and IV.

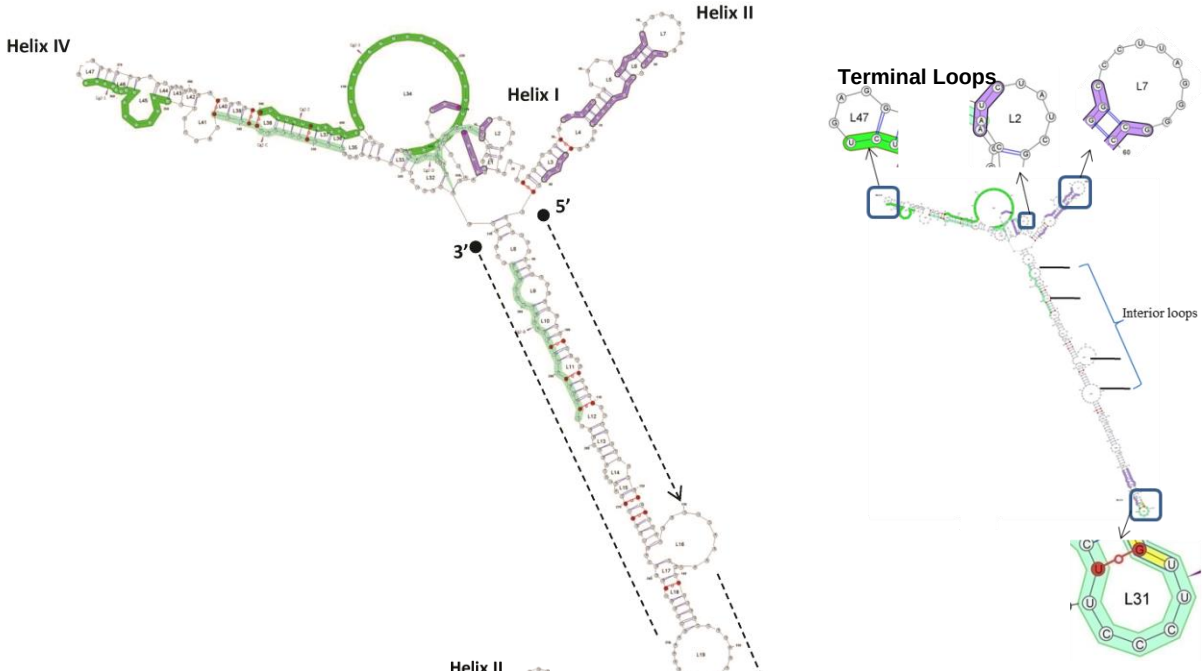


Figure 20. Loop distribution in ITS2 secondary structure.

Figure 22. Helix Length. Dotted arrow indicates the 5'-3' length direction.

Figure 21. Non-canonical bp in secondary structure.

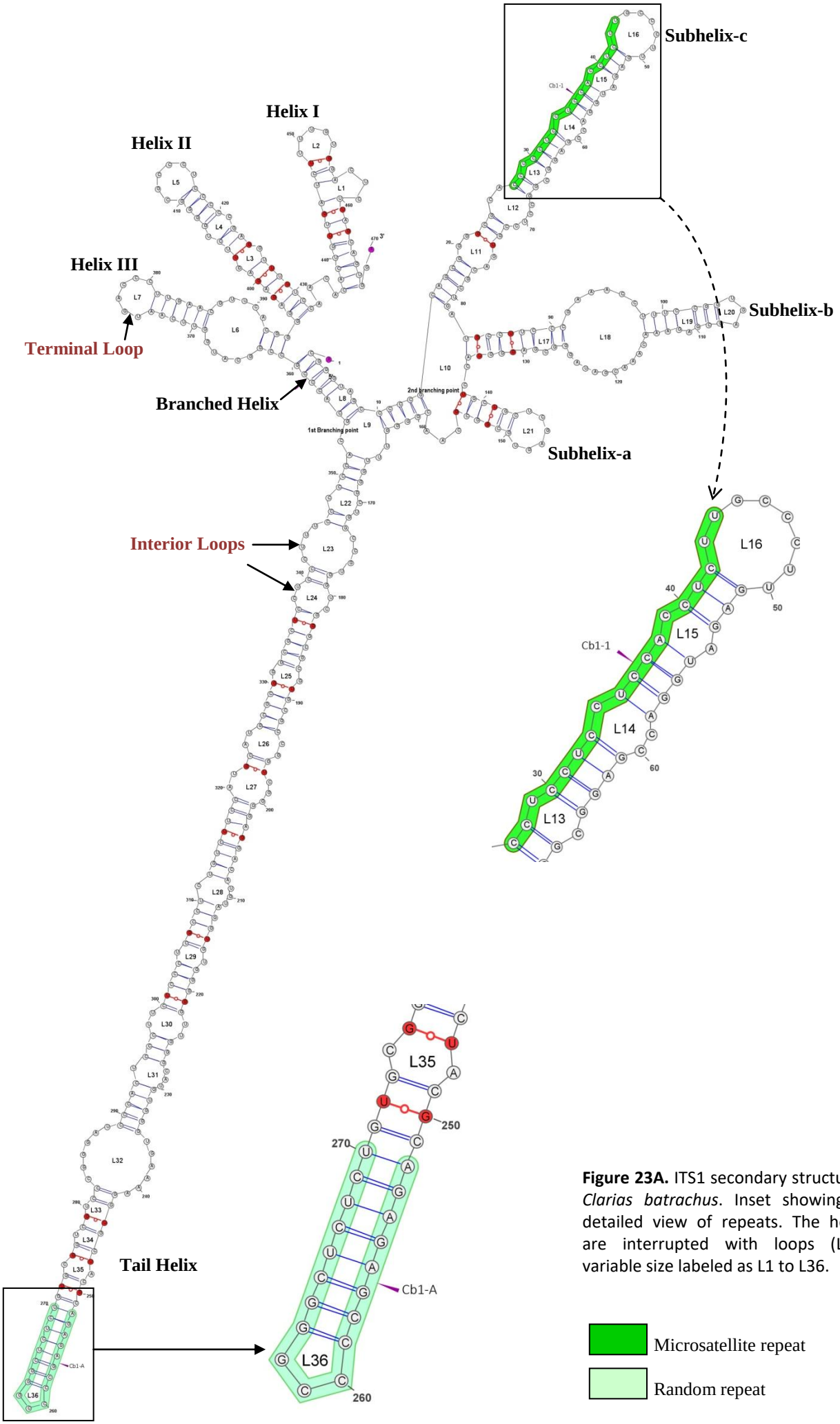


Figure 23A. ITS1 secondary structure of *Clarias batrachus*. Inset showing the detailed view of repeats. The helices are interrupted with loops (L) of variable size labeled as L1 to L36.

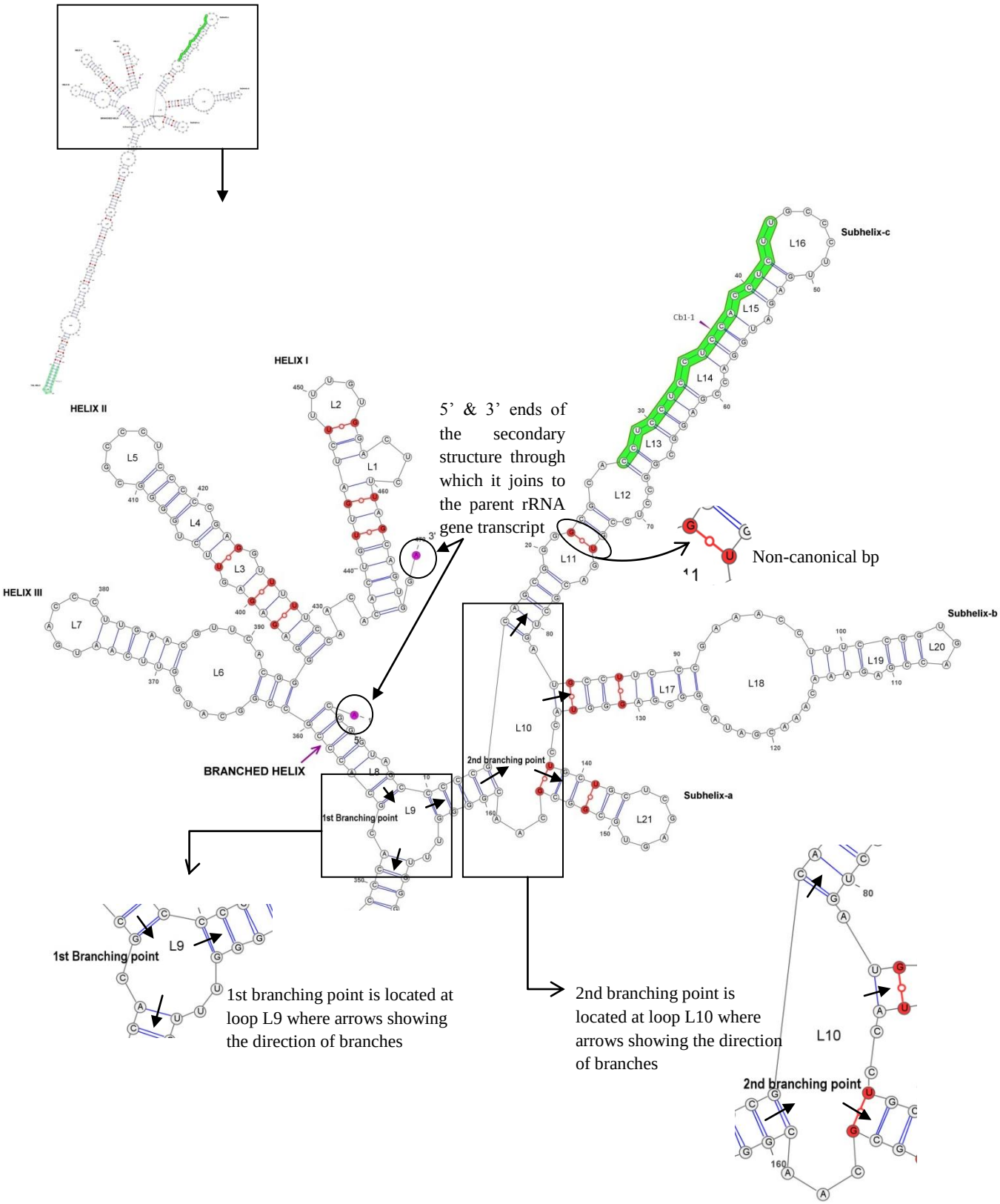


Figure 23B. ITS1 secondary structure of *Clarias batrachus*. Inset showing the detailed of branched helix.

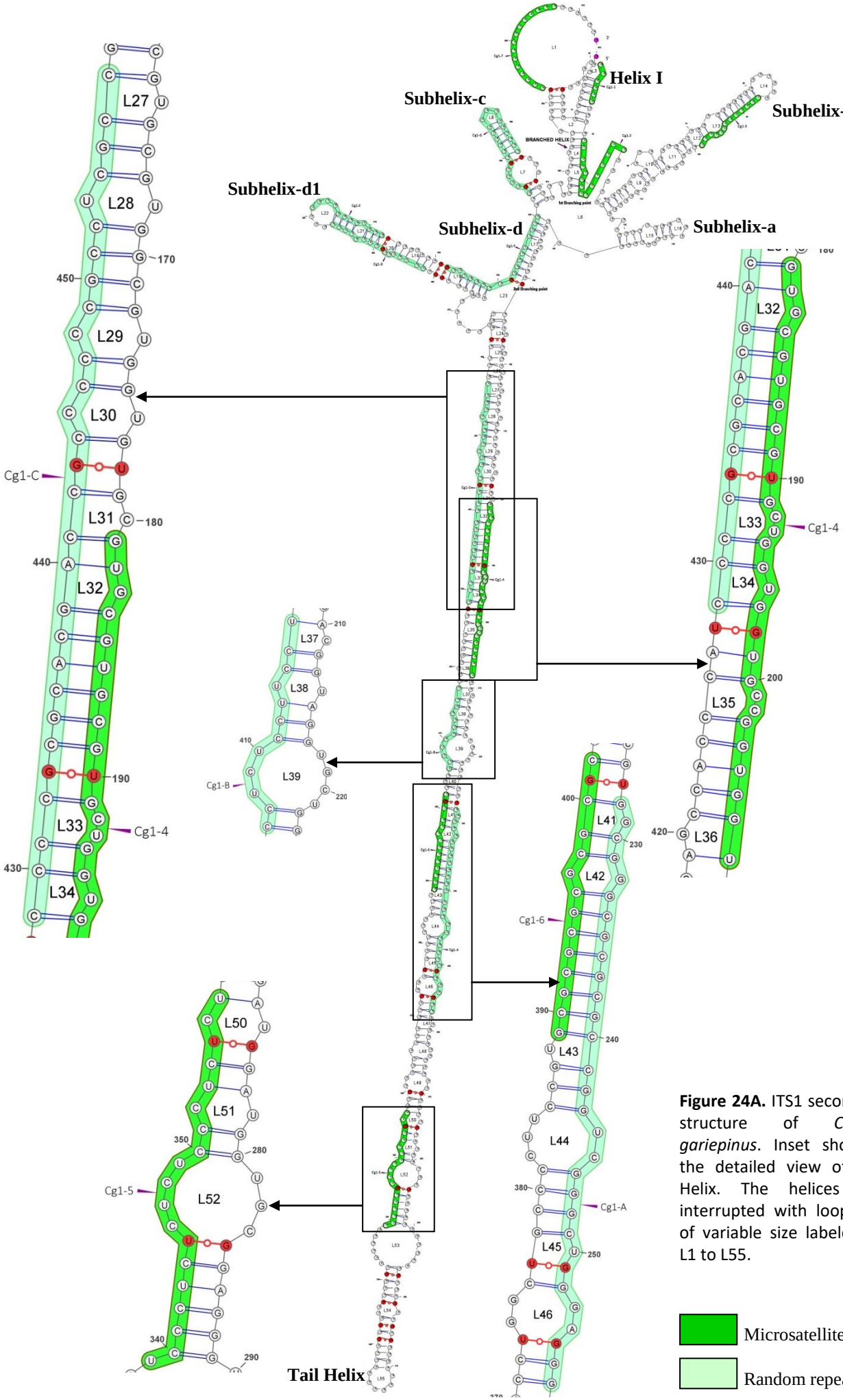


Figure 24A. ITS1 secondary structure of *Clarias gariepinus*. Inset showing the detailed view of Tail Helix. The helices are interrupted with loops (L) of variable size labeled as L1 to L55.

Microsatellite repeat

Random repeat

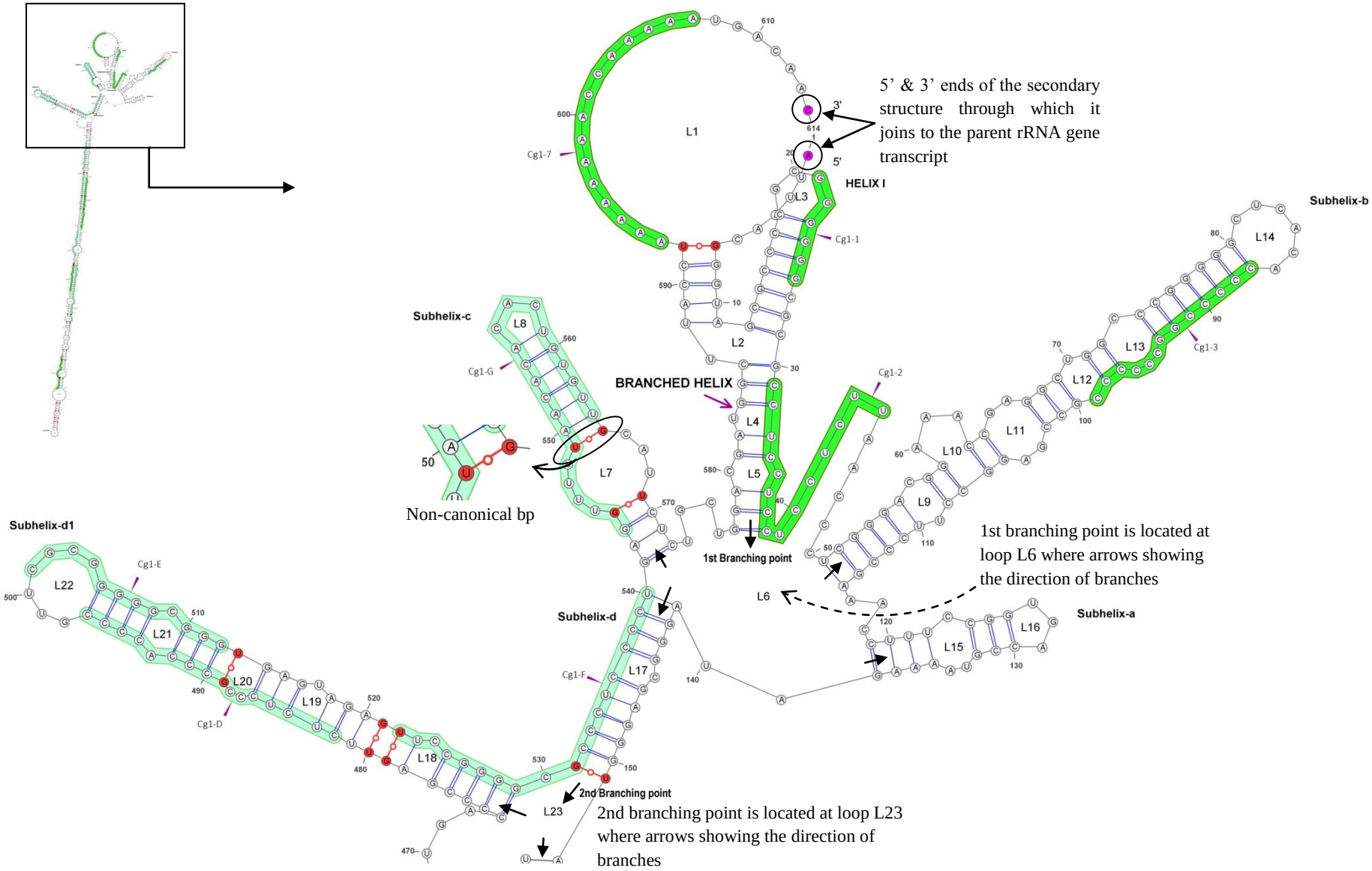


Figure 24B. ITS1 secondary structure of *Clarias gariepinus*. Inset showing the detailed view of upper branched region.

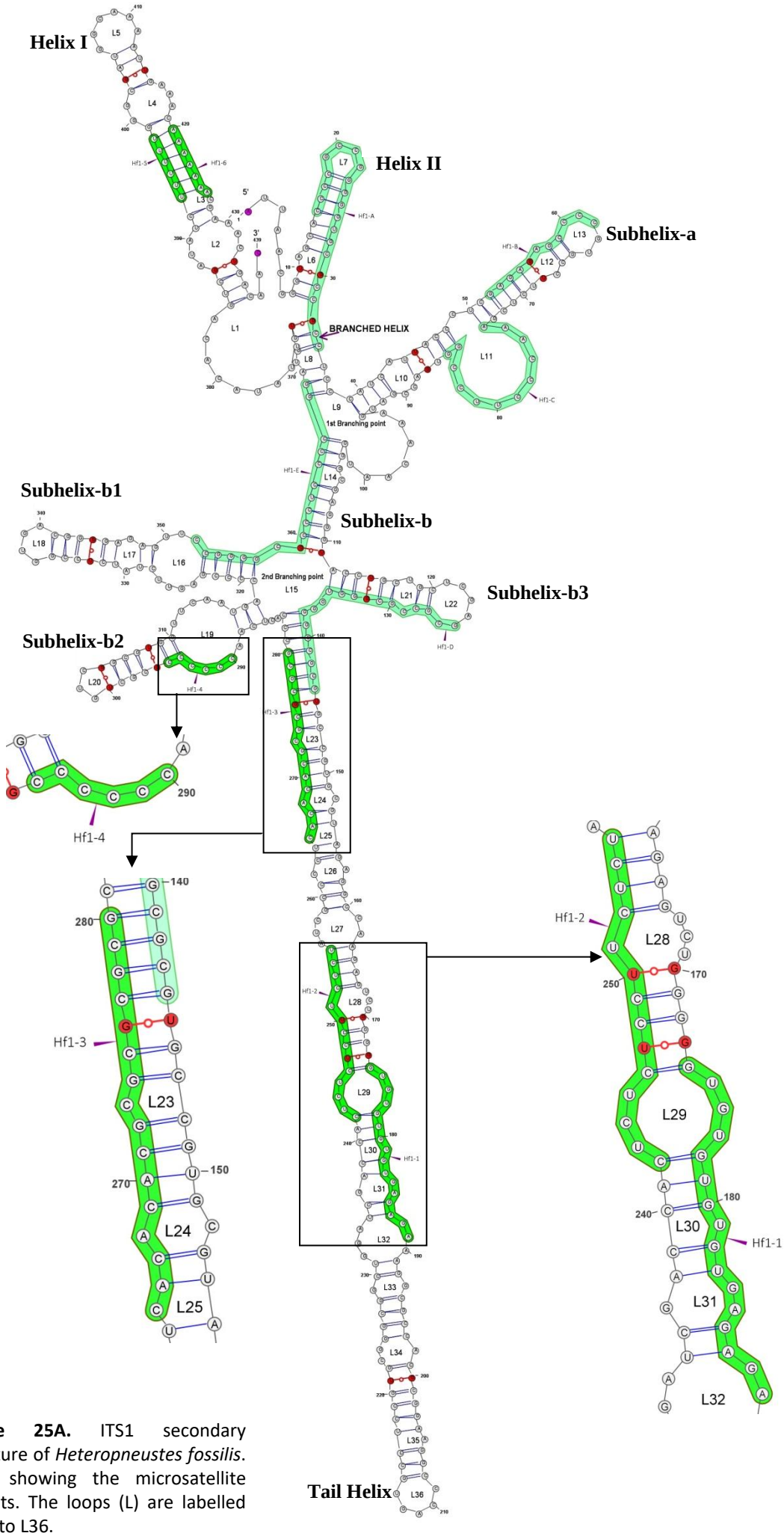


Figure 25A. ITS1 secondary structure of *Heteropneustes fossilis*. Inset showing the microsatellite repeats. The loops (L) are labelled as L1 to L36.

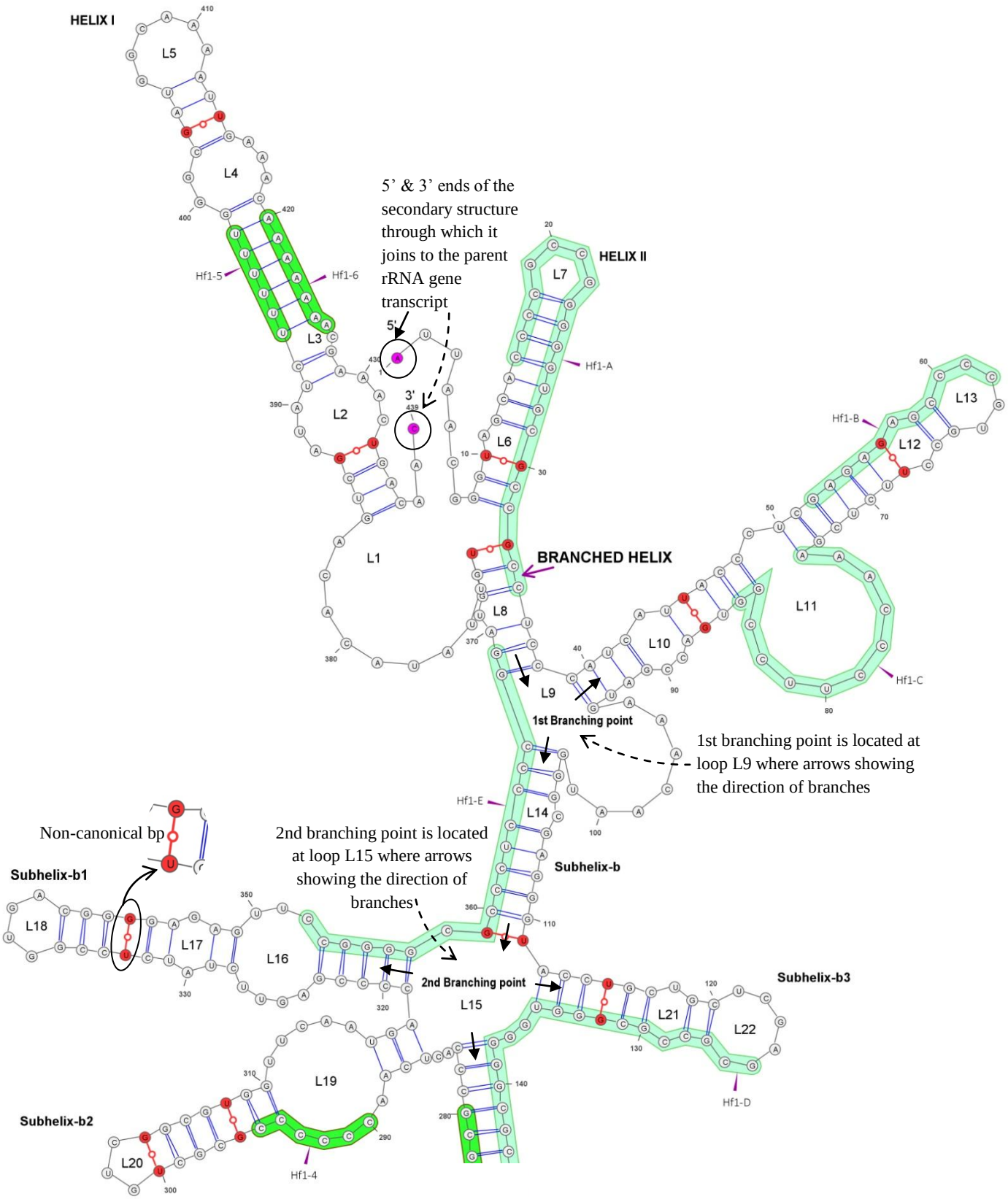


Figure 25B. ITS1 secondary structure of *Heteropneustes fossilis*. Detailed view of upper branched region showing the branching pattern of helix and subhelices.

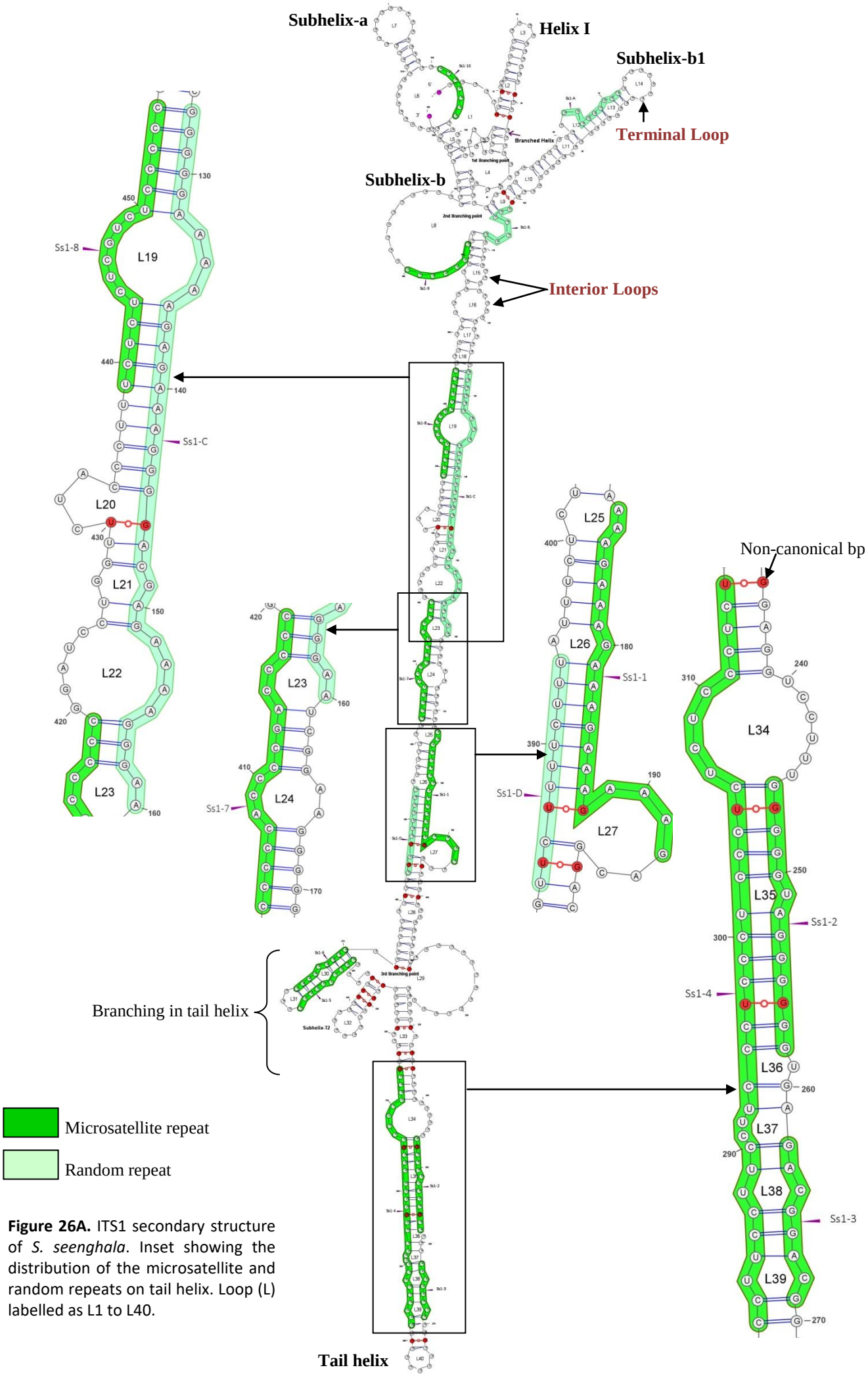


Figure 26A. ITS1 secondary structure of *S. seenghala*. Inset showing the distribution of the microsatellite and random repeats on tail helix. Loop (L) labelled as L1 to L40.

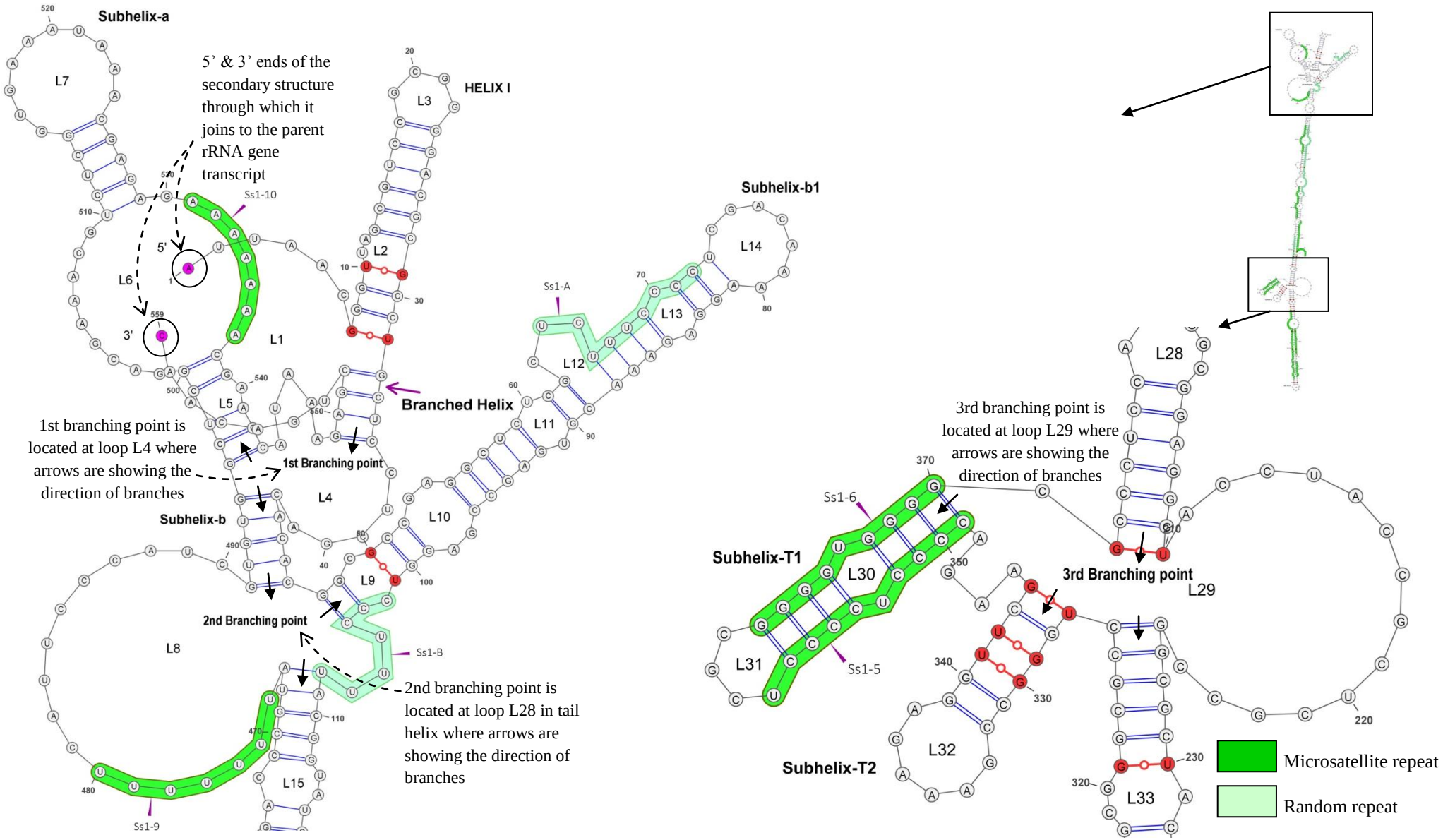


Figure 26B. ITS1 secondary structure of *S. seenghala*. Inset showing the detailed view of upper branched region and branching on tail helix.

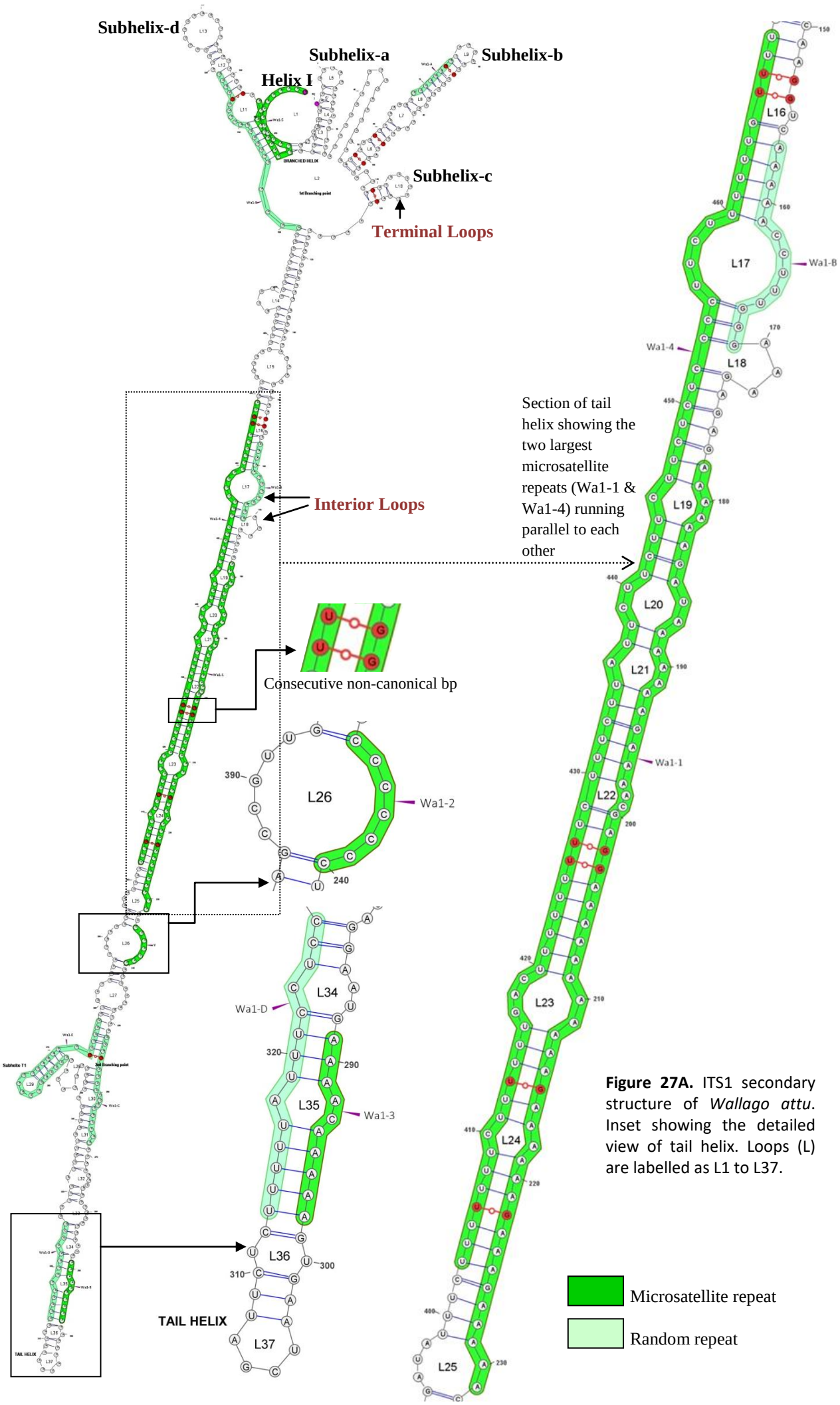


Figure 27A. ITS1 secondary structure of *Wallago attu*. Inset showing the detailed view of tail helix. Loops (L) are labelled as L1 to L37.

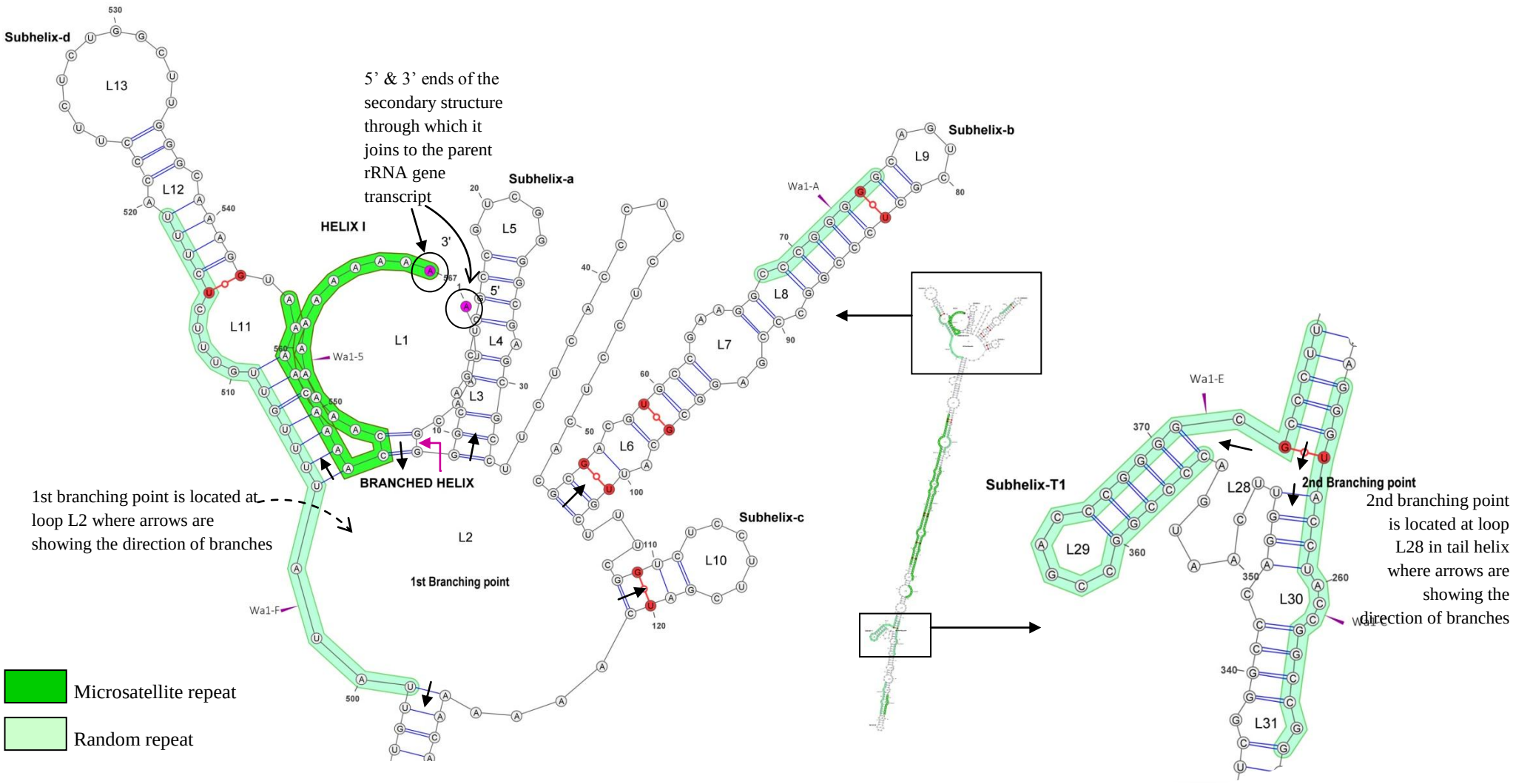


Figure 27B. ITS1 secondary structure of *Wallago attu*. Inset showing the detailed view of upper branched region and branching on tail helix.