

## A secondary structural common core in the ribosomal ITS2 (internal transcribed spacer) of *Culex* species from diverse geographical locations

Rayavarapu Bhargavi<sup>1</sup>, Siddharth Vishwakarma<sup>2</sup> and Upadhyayula Suryanarayana Murty<sup>1\*</sup>

<sup>1</sup>Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Hyderabad, India; <sup>2</sup>Department of P.G. Studies and Research in Biological Sciences, RDVV, Jabalpur, India;

Upadhyayula Suryanarayana Murty\* - Email: murty\_usn@yahoo.com; Fax: +91-40-27193227

\* Corresponding author

received August 15, 2005; revised September 19, 2005; accepted October 3, 2005; published online October 5, 2005

### Abstract:

In the present study, sequence and structural analysis of ITS2 region (the spacer segment between 5.8S and 28S rRNA of mature rRNA sequences) of 7 *Culex* species belonging to 5 different geographical locations was carried out. Alignment of the ITS2 sequence from the 7 species revealed 8 homologous domains. Four species namely *C. vishnui*, *C. annulus*, *C. pipiens*, *C. quinquefasciatus* showed high sequence (98-100%) and RNA secondary structure similarity. The ITS2 similarity among different species is high despite their varying geographical locations. Several common features of secondary structure are shared among these species, with some of them supported by compensatory changes, suggesting the significant role by ITS2 as an RNA domain during ribosome biogenesis.

**Key words:** *Culex*; ITS2; secondary structure; geographical locations; ribosome biogenesis

### Background:

*Culex* mosquito species have been described from a wide range of environments and are involved in pathogen transmission from human to reservoirs and vice-versa. [1] Sequence similarity have been reported for ITS2 (the spacer segments between 5.8S and 28S rRNA of mature rRNA sequences) from 5 *Culex* species of diverse geographic locations. [3,4,5,9,20] The ITS spacers are versatile as genetic markers and have been used for the determination of taxonomic classification. [17] Recent functional analyses performed on yeast ribosomal RNA genes clearly show that the structural integrity of the transcribed spacer regions is an essential prerequisite for the correct processing of mature rRNA and for the biogenesis of active ribosomal subunits. [2,3] The derivation of reliable secondary structure models for each transcribed spacer region would represent a major step towards a detailed understanding of their biological roles.

Comparative sequence analysis provides a powerful tool for identifying biologically relevant folding patterns in RNA molecules. [4] This involves collection of sequences exhibiting substantial nucleotide differences while retaining unequivocal sequence similarity. However, a high degree of sequence variation in the transcribed spacers is known even among closely related species. [5,6,7] Here, we analyze ITS2 from 7 *Culex* species characteristic of 5 different geographical locations to assemble common RNA structural features.

### Methodology:

### Dataset:

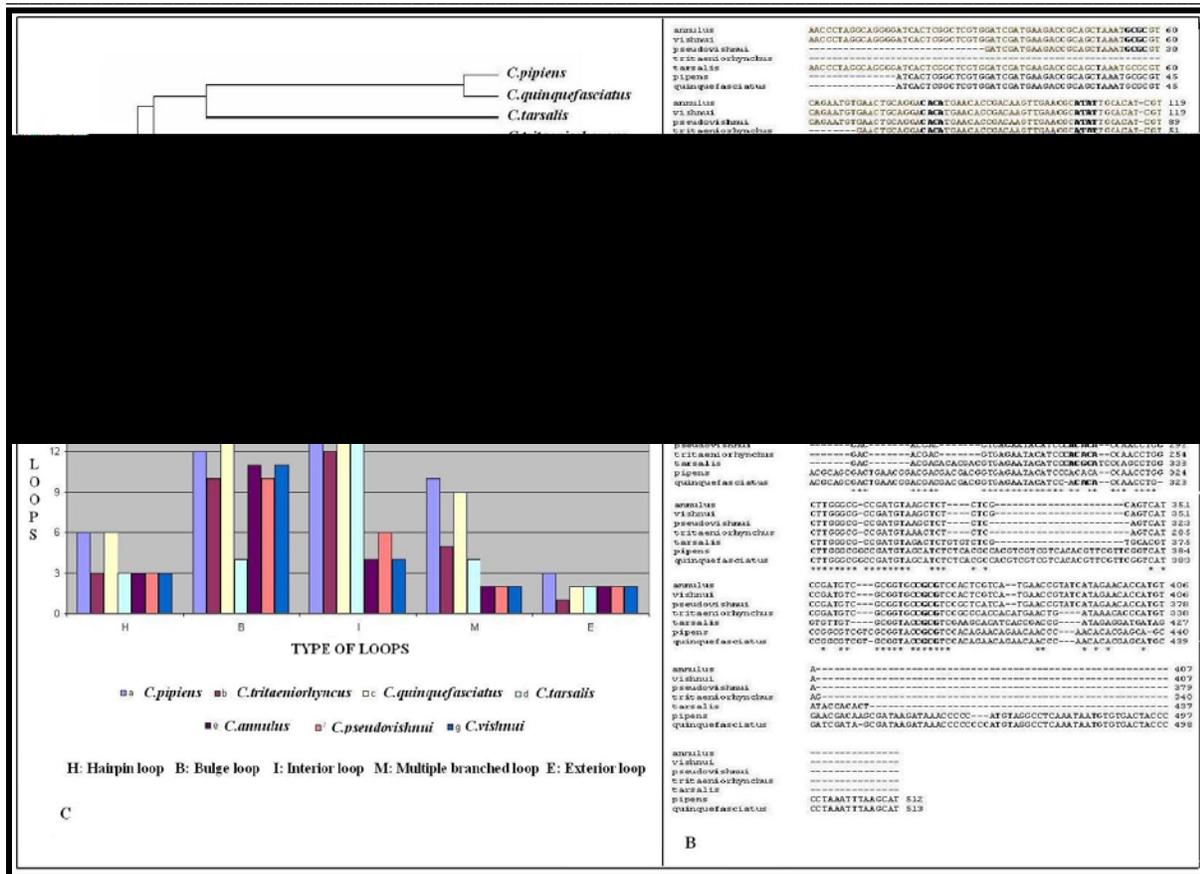
ITS2 nucleotide sequences from 7 *Culex* mosquitoes characteristic of 5 geographical locations (Italy, China, Africa, America, and Japan) were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank). The GenBank accession numbers for ITS2 sequences from *C. pipiens* (North America), *C. tritaeniorhynchus* (China), *C. quinquefasciatus* (Africa), *C. tarsalis* (Italy), *C. annulus* (China), *C. pseudovishnui* (China), and *C. vishnui* (Japan) were X75817, AF305558, Z48468, U33031, AF453488, AF453498 and AF165900, respectively.

### Sequence alignments:

Multiple sequence alignments were performed using CLUSTALW with a gap opening penalty of 15 and gap extension penalty of 6.66. [8]

### Secondary structure prediction:

The RNA secondary structures for ITS2 were predicted using RNADRAW. [21] RNADRAW predicts RNA structures by identifying suboptimal structures using the free energy optimization methodology at a default temperature of 37°C. In the current study, ITS2 and 5.8S regions (the first 170 nucleotides) were used for RNA structure prediction. The minimum energy structure prediction algorithm in RNADRAW was ported from the RNAFOLD program included in the Vienna RNA package. [24] The dynamic programming algorithm employed in RNADRAW was based on the work of Zuker and Stiegler [25] and uses energy parameters taken from Turner [26], Freier [27] and Jaeger. [28]



**Figure 1:** (A) Phylogenetic tree construct of ITS2 sequences from 7 *Culex* species. (B) Multiple sequence alignment of 5.8S rDNA and ITS2 from 7 *Culex* species. Colored segments = 5.8S rDNA; **Bold segments** = repeat regions. (C) Distribution of different types of loops (hairpin, bulge, multi branched, interior and exterior) among different isolates of *Culex*

Species	ITS2		5.8S + ITS2			No of Stems	Energy
	Length	G+C	GC	AU	GU		
<i>C. tritaeniorhynchus</i>	260	55	47	37	10	18	-82.57
<i>C. pipiens</i>	512	53	85	68	12	32	-149.38
<i>C. quinquefasciatus</i>	513	53	85	69	12	33	-148.23
<i>C. tarsalis</i>	282	57	73	46	20	28	-129.2
<i>C. annulus</i>	253	54	64	48	16	25	-115.3
<i>C. pseudovishnui</i>	255	54	57	49	12	22	-105.66
<i>C. vishnui</i>	253	54	64	48	16	25	-115.3

**Table 1:** Lengths (nucleotides); G+C content (%); GC, AU, GU base pairs (numbers); energy (Kcal) for the secondary structures of 5.8S and ITS2 rDNA sequences of *Culex* species

### Phylogenetic analysis:

The phylogenetic Genebee service was used for phylogenetic tree construction. [22]

### RNA fold:

The *Stribo* program in *Sfold* (Statistical Folding and Rational Design of Nucleic Acids) was used to predict the probable target accessibility sites (loops) for trans-cleaving ribozymes in ITS2. [23] The prediction of accessibility is based on a statistical sample of the Boltzmann ensemble for secondary structures. Here, we assessed the likelihood of unpaired sites for potential ribozyme target. Each mRNA exists as a population of different structures. Hence, stochastic approach to the evaluation of accessible sites was found appropriate. [29] The probability profiling approach by Ding and Lawrence [30] reveals target sites that are *commonly accessible* for a large number of statistically representative structures in the target RNA. This novel approach bypasses the long-standing difficulty in accessibility evaluation due to limited representation of probable structures due to high statistical confidence in predictions. The probability profile for individual bases ( $W = 1$ ) is produced for the region that includes a triplet and two flanking sequences of 15 bases each in every site of the selected cleavage triplet (e.g., GUC).

### Results and Discussion:

ITS2 sequences showed a taxonomic trend similar to that in phylogeny construction (Figure 1A). A multiple sequence alignment of 5.8S rDNA and ITS2 showed in Figure 1B indicated that more distantly related species have lower sequence similarity in the ITS2 region. The predicted features of ITS2 using RNADRAW are given in Table 1. The stems (double stranded paired regions) stabilize RNA secondary structures and the number of stems present in each ITS2 is given (Table 1). ITS2 RNA structures from *C. pipiens* and *C. quinquefasciatus* have the highest negative free energy (-149.38 Kcal and -148.23 Kcal) followed by *C. tarsalis* (-129.2), *C. annulus* and then by *C. vishnui* (-115.3), *C. pseudovishnui* (-105.66) and *C. tritaeniorhynchus* (-82.57). Visual comparison shows that this is related to the trend in the cladogram given in Figure 1A. A high degree of sequence similarity is observed at the 5' end compared to the 3' end (Figure 1A). This is due to factors such as genetic drift, the relative number and size of repeats, rates of unequal crossover, gene conversion, immigration and the number of the loci influencing the length. [10] Furthermore, a high level of sequence similarity is found between *C. annulus* and *C. vishnui* as well as *C. quinquefasciatus* and *C. pipiens*. The simple tandem repeats shown in Figure 1B as bolded regions are found to be similar. This similarity is seen in corresponding RNA structures and computed energies.

The *Culex* species considered in this study were then grouped into three classes based on the similarity of RNA stems and loops. Despite their different geographic locations with varying eco-climatic conditions, class I (*C. annulus*, *C. pseudovishnui* and *C. vishnui*) isolates (China and Japan) showed high similarity in secondary structural features. Similar observations were seen in class II (*C. pipiens*, *C. quinquefasciatus* and *C. tarsalis*) isolates from North America and Africa. [19] A high degree of similarity in the 5.8S region unlike the ITS2 is shown for different isolates due to relative evolutionary selection. [9] The loops 11 and 12 having sequences UGUCG and CUUCGUG, respectively are highly conserved in all classes.

Figure 1C shows the distribution of different types of loops (hairpin, bulge, multi branched, interior and exterior) among different isolates. The segments of the ITS2 having score  $\geq 50$  are further probed carefully for target site to assess the likelihood of unpaired segments. Interestingly, the observed phylogenetic trend was identified with respect to the target accessibility sites for the seven *Culex* isolates. The order of preference is interior loop, bulge loop, multiple branched loop, hairpin loop and exterior loops in all the isolates.

Several homologous domains were observed in the ITS regions of *Aedes* mosquito species by Wesson *et al.*, [11] and was shown that these domains base pair to form a core region central to several stem features. This suggested that conserved core region of rDNA is more important for a functional rRNA folding pattern. Barker found that ITS2 is unique in the 16 populations of *Boophilus microplus*, *Boophilus decoloratus*, *Rhipicephalus appendiculatus*, *Rhipicephalus zambeziensis*, *Rhipicephalus evertsi*, *Rhipicephalus sanguineus*, *Rhipicephalus turanicus*, *Rhipicephalus pumilio* and *Rhipicephalus camicasi* from different geographical locations. [12] These results suggest that the differences and similarity observed in the ITS2 of different species are not simply accumulated due to random mutation and have evolved for functional selection in ribosome biogenesis. However, it was shown in three related mosquito genera (*Aedes*, *Psorophora* and *Haemogogus*) that the intra-spacer variable regions appear to co-evolve and that ITS2 variation is constrained by its secondary structure. [11] Further studies have demonstrated that the ITS2 is essential for the correct and efficient processing and maturation of 26S rRNA ribosomal units. [13] Furthermore, the information required for the efficient removal of ITS2 from its RNA precursor is not localized but dispersed throughout the ITS2 region. Thus, insertions and deletions (indels) that affect secondary structures alter rRNA processing. Critical changes in the rRNA folding pattern due to evolutionary sequence variation in the ITS spacer regions may thus have an important role on the kinetics of precursor rRNA formation for the efficient functioning of rDNA clusters.

**Conclusion:**

Comparison of ITS2 sequences from different isolates of *Culex* show similarity and variations. Surprisingly, species displaying sequence similarity belong to different geographical locations with diverse climatic and ecological conditions. This implies that the ITS2 regions have less selective pressure than the ribosomal regions. Several common structural folds were shared among the selected mosquitoes for maintaining functional equivalents. Construction of an evolutionary tree using more isolates of *Culex* will provide an understanding for their functional selection.

**Acknowledgement:**

Authors are thankful to the Director Dr. J. S. Yadav, IICT, Hyderabad for his encouragement and support. RB thanks ICMR (Indian Council of Medical Research) for Junior Research Fellowship.

**References:**

- [1] K. Nusha, *et al.*, *Molecular Ecology Notes*, 4:20 (2004)
- [2] W. Musters, *et al.*, *EMBO J.*, 9:3989 (1990) [PMID: 2249660]
- [3] C. A. Van der Sande, *et al.*, *J. Mol. Biol.*, 223:899 (1992) [PMID: 1538404]
- [4]