Original Article

Intraspecific variations in Cyt b and D-loop sequences of Testudine species, *Lissemys punctata* from south Karnataka

R. Lalitha, V.R. Chandavar *

Molecular Endocrinology Laboratory, Department of Biotechnology, Yuvaraja’s College, A Constituent Autonomous College of University of Mysore, Mysore 570005, India

**G R A P H I C A L A B S T R A C T**

**A B S T R A C T**

The freshwater Testudine species have gained importance in recent years, as most of their population is threatened due to exploitation for delicacy and pet trade. In this regard, *Lissemys punctata*, a freshwater terrapin, predominantly distributed in Asian countries has gained its significance for the study. A pilot study report on mitochondrial markers (Cyt b and D-loop) conducted on *L. punctata* species from southern Karnataka, India was presented in this investigation. A complete region spanning 1.14 kb and ~1 kb was amplified by HotStart PCR and sequenced by Sanger sequencing. The Cyt b sequence revealed 85 substitution sites, no indels and 17 parsimony informative sites, whereas D-loop showed 189 variable sites, 51 parsimony informative sites with 5 functional domains TAS, CSB-F, CSBs (1, 2, 3) preceding tandem repeat at 3’ end. Current data highlights the intraspecific variations in these target regions and variations validated using suitable evolutionary models points out that the overall point mutations observed in the region are transitions leading to no structural and functional alterations. The mitochondrial data generated uncover the genetic diversity within species and conservationist can utilize the data to estimate the effective population size or for forensic identification of animal or its seizures during unlawful trade activities.

© 2017 Production and hosting by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Introduction**

*Lissemys punctata* is a cryptodiran omnivorous freshwater turtle that belongs to order Testudines, family Trionychidae. The Reptile Database [1] provides comprehensive information regarding
species taxonomic position, distribution, number of species, and their conservation status. Trionychidae turtles (soft shelled turtles) are the preferred group of freshwater turtles for delicacy in Asia, due to its low bone to body ratio [2]. According to Bhupathy et al. [3] _L. punctata_ is being exploited in mass for their meat and eggs in recent years at various parts of India. TRAFFIC, a wildlife trade monitoring network (Wildlife Institute) of India and IUCN (2000) conservation action report reveals that _L. punctata_ is threatened due to high trade activity (CITES Appendix II). However, IUCN (Red list category) has categorized it as Low risk/least concern, which needs up gradation. It is possible only through usage of quick and consistent molecular tool for unbiased identification of the species during illegal trade activity.

DNA based forensic identification is gaining significance in current era, where morphological identification of species is no longer perceptible, such as in animal derived products (carapace powder, fragments of shells, meat etc). In this context, mitochondrial DNA and its loci have made remarkable contributions in turtle studies. Pereira [4] has explained the vertebrate mitogenome organization and its loci have made remarkable contributions in turtle studies. Praschag et al. [7] identified three distinct clade of phylogenetic relationship using _Cyt b_ along with other mitochondrial loci in _L. punctata_ subspecies from India and Sri Lanka. _Cyt b_ sequence is used as valuable locus on mtDNA in forensic sciences for wildlife species identification [8,9]. D-loop is the predominant regulatory most variable region of mtDNA, used to identify genetically discrete populations, foraging ground, and nesting behavior of turtles, intraspecific variability, and phylogenetic relationship [10–13]. D-loop mutation has promoted longevity in centenarians [14] and occurrence of tandem repeats (VNTRs) is regarded as valuable molecular marker in turtle species studies [15].

Current mtDNA study in _L. punctata_ was started by specifically extracting _Cyt b_ and D-loop sequence from complete mitogenome sequence of _L. punctata_ available at NCBI with Accession no. NC_012414.1. The mitogenome was composed of 13 protein coding genes, 22 tRNAs, 2 rRNAs with single D-loop/ control region as that of other vertebrates [4]. Partial _Cyt b_ sequences for _L. punctata_ from various states of India (Andra Pradesh, Goa, Gujarat, Karnataka (Mangalore district), Kerala, Maharashtra, Odisha and Tamil Nadu) with respective Accession numbers at NCBI (FR850622, FR850625, FR850626, FR850631, FR850632, FR850635, FR850637, FR850642, and FR850643) have been submitted by different authors [7]. In this regard, current study is the first to report mitochondrial sequence data of _L. punctata_ from Mysore districts of Karnataka, India. This work focused on intraspecific variations in _Cyt b_ and D-loop region, which would be cumulative to the existing knowledge and growing database of mitogenome. This provides an imperative groundwork for future conservational studies on _L. punctata_ species from south Karnataka, India.

### Material and methods

Blood samples from _L. punctata_ (named Lp1 to Lp8) were collected during the field work period from nearby freshwater bodies at Mysore district of Karnataka, India. **Permission for collecting blood**

### Table 1

List of _Cyt b_ gene primers designed using Primer3 online tool.

| Oligo (5’-3’) | Length | Tm (°C) | Ta (°C) | Amplicon (bps) |
|--------------|--------|---------|--------|----------------|
| F-GCAAAACATCACGAAAATCACAC | 24 | 57 | 55 | 226 |
| R-CGTATGATATGTAAGGTTGAGGCTG | 20 | 58 | | |
| F-GCCAGGACGATCACTATT | 20 | 58.3 | 55 | 298 |
| R-ACTGCAAGGTTAGAGACTGCTG | 21 | 59 | | |
| F-CAGCAATGCGAAATATACC | 22 | 58.4 | 55 | 178 |
| R-TCCCGGATGAACTGCTTCT | 20 | 58.4 | | |
| F-CCAATCAACCGCAACCACCCCATTTCGATT | 25 | 57 | 55 | 162 |
| R-AGGGCAGAGGGGCTATCAGAGAG | 20 | 58 | | |
| F-TAACATGTCGACCGAAATACC | 21 | 59.6 | 55 | 180 |
| R-TTGCTTCTGATAGGCTCGGA | 21 | 59.8 | | |
| F-TACAATGAAATTTGCCGAGTCCTC | 23 | 60 | 55 | 200 |
| R-TTGTGACGAGAGTTGGTCCGGA | 23 | 60 | | |
| F-ACCCAAATATTTCCGAGCC | 21 | 57 | 55 | 400 |
| R-TAATGGAATTTCTCCTGCTT | 26 | 57 | | |
| F-CATCTACCAAAAATACCATATACGCA | 27 | 58 | 55 | 155 |
| R-CCCTGAATTTTACTGCGTC | 21 | 59 | | |

### Table 2

List of D-loop primers designed using Primer3 online tool.

| Oligo (5’-3’) | Length | Tm (°C) | Ta (°C) | Amplicon (bps) |
|--------------|--------|---------|--------|----------------|
| F-TCCCTGGATGCTGACATATCACATAT | 20 | 58.5 | 55 | 247 |
| R-CTTCTGAACTTGAATGCCTGGA | 20 | 58.9 | | |
| F-ATCCGGCAATGGGTAATGTTGAC | 20 | 59 | 55 | 201 |
| R-CCGCGGATGAGCTGGAAGAAG | 20 | 59 | | |
| F-CCCATGATGACGGGAGG | 20 | 59 | 55 | 217 |
| R-TCCCGACGACATGATGATGC | 20 | 59 | | |
| F-ATCCTCGGACATGGCCTG | 20 | 59 | | |
| R-CTTGCTGACAGATGTCGG | 20 | 59 | | |
| F-CCCTGAGAGCCAACATGATTTA | 19 | 59 | 55 | 345 |
| R-AGCTATGCAAGGCTGGAAA | 20 | 59 | | |
samples was obtained from Principal Chief Conservator of Forest (Wildlife), Bangalore, Karnataka, India vide letter No. D/WL/CR/149/2010 and PS/WL/CR/21/2013. Without sacrificing or anesthetizing the animal, about 0.3 mL venous blood was drawn from the hind limb femoral vein using lithium/sodium heparin coated BD vacutainers [16]. The animals were left to their habitat after medicating the spot using Betadine solution. Whole blood was immediately stored in 5 mL of Longmire lysis buffer (100 mM

Fig. 1. Multiple sequence alignment of *L. punctata* Cyt b sequences with that of reference (BioEdit v7.2.5). Dots represent consensus nucleotides and Nucleotide (A, T, G, C) denotes variations.
TrisHCl pH 8, 100 mM EDTA, 10 mM NaCl and 0.5% SDS) [17] that is recommended for field and room temperature storage of blood samples for future genomic DNA extraction. Genomic DNA extraction followed conventional phenol–chloroform method with proteinase K digestion [18]. Finally, the pellets were resuspended in 1x TE buffer and stored in −20 °C for future genetic analysis.

Set of primers were designed for amplifying Cyt b and D-loop region using specific regions from reference sequence (NC_012414.1) with the help of Primer 3 online tool [19] and presented in Tables 1 and 2. Hotstart PCR was performed at room temperature using JumpStart Taq Ready mix (Sigma-Aldrich Co. LLC, Bangalore, India) with total reaction volume of 50 μL consisting components according to manufacturer’s protocol. Amplifications were performed in Applied BiosystemsVeriti™ 96 well thermal cycler with initial denaturation at 95 °C for 1 min, denaturation at 94 °C for 30 sec, respective annealing temperatures for Cyt b and D-loop primers (Tables 1 and 2) for 1 min, extension at 72 °C for 2 min for 45 cycles and final extensions at 72 °C for 10 min and hold at 4 °C. Amplified products were inspected in 2% low melting agarose gel and amplicon size was assessed by running prestained 50 bp DNA ladder (MolBio™ HiMedia, Mysore, India).

PCR products were submitted for Sanger Sequencing at Chromous Biotech Pvt. Ltd, Bangalore, India. The products were column purified using Chromous PCR clean-up kit and standard protocol was followed to sequence both strands in Applied Biosystem 3500xl genetic analyzer. Further obtained fasta sequences were aligned using MUSCLE online tool. They were further edited and analyzed using BioEdit v7.2.5 [20] and MEGA v6.06 [21] user friendly software.

Results and discussion

As per Parson et al. [22] the complete Cyt b (1140 bp) and D-loop region (~1000 bp) were identified solely by comparing with database mitogenome sequence of L. punctata (NC_012414.1). All the sequences subjected to preferential BLAST with L. punctata, scored highest BLAST hits (i.e., E-value less than equal to 0) provides first level confirmation to rule out the chances of ‘numts’ (nuclear copies of mitochondrial origin) [23]. The common but, not universal phenomenon of transposition of mtDNA fragments into nuclear genome is referred to ‘numts’ [23]. These are referred...
as common contaminant encountered, while using genomic DNA
from blood sample as a source for mtDNA studies in lower verte-
brates [23,24]. Sorenson and Quinn [23] suggested hints to recog-
nize and avoid ‘numts’ from actual mtDNA sequence. It includes
avoid using universal primers designed for the taxa, instead sug-
gest to use newly designed primers using reference sequence avail-
able [23]. Accordingly, all the primers in the current study were
newly designed choosing particular region of interest from the ref-
ence sequence. Also, the presence of unusual substitutions or
stop codons [23] in the sequenced fragments was verified to con-
firm the absence of ‘numts’. According to Spinks and Shaffer [25]
the presence of multiple peaks in the sequenced chromatogram
indicates ‘numts’. The chromatogram obtained in the current study
showed clear single peaks and no mixed peaks were observed.

Intraspecific Cyt b sequence variations

Multiple sequence alignment of Cyt b sequence with its refer-
ence (Fig. 1) revealed that the open reading frame begins with
‘ATG’ (Methionine) and terminates with ‘TAA’ was similar to other
reported Trionychidae species in database, except Pelodiscus sinen-
sis from Republic of Korea (AY962573.1) reported ‘ATT’ start codon
for Cyt b [26]. The AT rich (60%) nucleotide composition of Cyt b
gene in current study reflects the characteristic vertebrate mito-
genome composition [27]. ‘TGA’ (or UGA) believed to be a universal
termination codon for nuclear genes appeared in the current Cyt b
sequence stretch but, here it codes for Tryptophan as in all verte-
brate mtDNA [28].

Sequence identity for Cyt b calculated in BioEdit v7.2.5 for pair-
wise alignment of all L. punctata samples showed 98.6 to 99.9%
identity between the individuals. The percentage identity is analo-
gous to Shen et al. [29] report for intraspecific variations that is
usually less than 1–2% and do not exceed 5%. The 1140 bp Cyt b
sequence analyzed in MEGA v6.06 [21] (Fig. 1) revealed 85 sub-
stitution sites, which includes high transition (55) when compared to
transversion (28), only 2 sites showed both type of mutations, no
indels and 17 parsimony informative sites. With respect to amino
acid changes observed after translating the Cyt b nucleotide
sequence, there were 45 – 3rd codon transitions and 6 – 1st codon
transitions together 51 synonymous substitutions and only 22
non-synonymous substitutions. Avise et al. [30] reported that pro-
tein coding genes will have synonymous substitutions higher than
non-synonymous substitutions; this appears to be true with our
data.

Further, the Cyt b sequences were individually subjected to CDD
at NCBI [31] for annotation of protein specific domains. The search
parameters included selection of CDD v3.16 database, with no low
complexity filter and E-value threshold of 0.01. To increase the
consistency of domain search rescue borderline hits and suppress
weak overlapping hit parameters were also selected to perform
live search at CDD. Out of 500 maximum hits, the best five hits
are depicted in Fig. 2, which showed E-value < 0.01. All 8 query
sequences showed specific hits with CYT B (MTH00119) a member
of cl27766/QcrB protein superfamily representing proteins
involved in energy production and conversion of taxon Sauropsida
(snakes, beaked reptiles and turtles) (Fig. 2). The conserved func-
tional domains of Cyt b gene like hemebl binding site, hemebH
binding site and Qi binding sites spans between nucleotide resi-
dues 52nd to 6219 in current data (Fig. 2). In spite of nucleotide
variations observed in Cyt b sequence the presence of functional
conserved domains highlights its role in electron transport system,
thereby confirm the protein itself. The Cyt b sequences of L. punct-
tata from the present study can be accessed in future at NCBI using
Accession numbers (KY946735, KY946736, KY946737, KY946738,
KY946739, KY946740, KY946741 and KY946742) respectively.

Intraspecific D-loop/control region sequence variations

The length of D-loop region flanked by tRNA- Pro and Phen,
amplified in the current study range from 999 to 1008 bp and
the alignment is as shown in Fig. 3. The overall nucleotide compo-
sition consist 33–34% (A), 30–32% (T), 11–12% (G) and 21–22% (C),
highlighting AT rich vertebrate mtDNA composition [27]. D-loop
sequences (~1000 bp) analyzed in MEGA v6.06 [21] for intraspeci-
fic variations revealed 189 variable sites, 51 parsimony informative
sites and only 89 conserved sites. Though D-loop is the most vari-
able, fast evolving part of mitochondrial genome [32], they too pos-
sess functional conserved domains preceding VNTR region like TAS,
CD (CSB-F) and CSBs (1, 2 and 3) as found in mammals [33], fresh-
water turtles of order- Geoemydidae [32] and Trionychidae [34].

Fig. 2. Graphical summary of conserved domains found in Cyt b sequences of current study along with list of domain hits from CDD-NCBI.
These functional sites have been identified in the current data and their intraspecific variations are depicted in Table 3.

Xiong et al. [34] reported that TAS domain is characteristic of both pleurodiran and cryptodiran turtles with the sequence 5'-TACAT-3' and its reverse complementary (RC) sequence 5'-ATGTA-3' near 5' region involved in termination of H strand synthesis via stable hairpin loop formation. Among L. punctata samples studied 'TACAT' and 'ATGTA' was observed close to 5' end of D-loop region (Table 3). According to Xiong et al. [34] the conserved domain (CD) with conserved sequence block (CSB-F) is the only found CD unit among soft shelled turtles. It is characterized by 'AGAAATAAGCATC' sequence. In current data also the
region is conserved with the exception of only 1 intraspecific variation – transversion (G to C-2nd residue) (Table 3). CSB-F has been reported in all different classes of vertebrates along with other units (B, C, D, E) signifying its role in H strand replication across the vertebrate lineage [34,35].

Both Xiong et al. [34] and Zhang et al. [35] observed conserved sequence block CSB1, CSB2, CSB3 in CSB domain in turtles of family Trionychidae and Geoemydidae. These blocks were found along with some regulative regions, H-strand replication origin sites, and transcription promoters for H/L-strand genes [35]. Wang et al. [36] reported that among the three, CSB1 6 bp motif ‘GACATA’ is conserved from fishes to mammals, birds and also in turtles [34]. CSB1 in soft shelled turtles is identified by ‘TTAATGCTAGATAGACATA’ sequence [34]. Present data revealed mutations upstream to conserved 6 bp motif i.e., Lp1, Lp6, Lp7 and Lp8 showed transition at position 7 (C to T) and all 8 samples showed 2 transitions (A to G) at position 11, 13, 1 transversion (T to A) at position 12 in CSB1 sequence (Table 3). Xiong et al. [34] reported that a typical CSB2 is characterized by two (C)6 series separated by ‘TA’, with sequence being ‘TT(A)3(C)6TA(C)6’. Our data deviates by 3 transversions in this sequence resulting in ‘CT(A)3(C)3T(C)2TA(C)5T’ (Table 3). CSB3 is represented by ‘TCGTC(A)3(C)4(A)4TCC’ sequence showed no intraspecific variations in our data which coincides with Xiong et al. [34] report on interspecies data for the region (Table 3). Wang et al. [36] also reported absence of CSB2 and CSB3 in pleurodiran turtles, birds and some mammals.

### Table 3

| Intrasp. variation of D-loop conserved regions after alignment from BioEdit v7.2.5. |
|---------------------------------|-----------------|------------------|-----------------|-----------------|-----------------|-----------------|
| **TAS** (36–40) | **RC of TAS** (26–30) | **CSBF** (254–266) | **CSB1** (595–613) | **CSB2** (680–698) | **CSB3** (729–747) |
| NC_012414.1 | TACAT | TATGTA | AGAATAAGCATC | TTAATGCTAGATAGACATA | TTAACCCCCCTTACCCCCC | TCGCTAAAAACCAAATTC |
| Lp1 | **** | *** | *** | *** | *** | *** |
| Lp2 | **** | **** | *** | *** | *** | *** |
| Lp3 | **** | **** | *** | *** | *** | *** |
| Lp4 | **** | **** | *** | *** | *** | *** |
| Lp5 | **** | **** | *** | *** | *** | *** |
| Lp6 | **** | **** | *** | *** | *** | *** |
| Lp7 | **** | **** | *** | *** | *** | *** |
| Lp8 | **** | **** | *** | *** | *** | *** |
Variable number tandem repeats (VNTRs) as the name suggest is the most variable part both in length and type of repeats, a characteristic feature of most vertebrate D-loop region [35]. The distribution of VNTRs is random but, still occurs at certain hypervariable sites or domains at 5’ or 3’ region of D-loop [35,37]. The dynamic VNTRs decide the length heteroplasmy observed in mitogenome across species lineage, originated by strand slippage or replication mispairing [35]. The D-loop sequences of current study were submitted to Tandem Repeat Finder (TRF) tool [38], revealed only one type of tandem repeat ‘AT’ with copy number range 23.5 to 24.66 at 3’ end. The similar observation is reported by Xiong et al. [34] for L. punctata species. The D-loop sequences of L. punctata obtained in the current study can be accessed from NCBI database with following accession numbers (KY946743, KY946744, KY946745, KY946746, KY946747, KY946748, KY946749 and KY946750).

**Nucleotide substitution analysis for Cyt b and D-loop sequences**

The Cyt b and D-loop sequences were subjected to best fit model test in MEGA v6.06 [21]. This resulted in selection of TN93+G [39] (Table 4) and HKY+G+I [40] (Table 5) models respectively, based on lowest AIC corrected criterion values [41]. The nucleotide frequencies used by these models are A = 33.27%, T = 26.91%, C = 28.75%, G = 11.07% for Cyt b and A = 34.31%, T = 31.63%, G = 12.09%, C = 21.97% for D-loop sequence. The maximum Log likelihood values for substitution matrix under these respective model were calculated to be −2042.852 (Cyt b) and −2280.411 (D-loop).

The transition/transversion rate ratio observed between Cyt b and D-loop region of L. punctata in the current study are 2.08 and 1.56; this may be due to difference in mutation accumulation, mutation rate and frequency between the regions [42]. The ratios are not comparable with that of other species as the frequency varies between species, classes, order, and genera [43]. Also Keller et al. [44] pointed that transition/transversion bias is not ubiquitous to all vertebrates and invertebrates, rather it is species specific. Cyt b sequence in the study showed transitions and transversions leading to synonymous substitutions of amino acid in the conserved domain and other regions of protein, which in turn impact no change in protein folding, structure and activity. In contrast to this, complete D-loop sequence showed more transversions but, maintain high transitions in their conserved domains so that their regulatory function is not disturbed. Probably, these transversions are not harmful to the organism as the region is noncoding.

**Conclusions**

The data uncover the intraspecific variations in mitochondrial Cyt b and D-loop regions of the L. punctata. This helps to understand the genetic structure of L. punctata dwelling in the geographical location mentioned in the article. Hence from the analysis, it is clear that mitochondrial protein coding gene Cyt b showed high synonymous substitutions which indicate natural selection favors transitions in protein coding genes so as to maintain its quaternary structure necessary for functional constancy in turtles as well. In case of complete noncoding D-loop region transversions outnumbered transition but, with respect to functional domains transition were more indicating that they are necessary for its regulatory function.

**Conflict of interest**

The authors declare that there is no conflict of interest.

**Acknowledgements**

The corresponding author acknowledges Department of Science and Technology, New Delhi, India for sanctioning project (No. DST- SERB/SB-EMEQ-124/2013). First author is thankful to DST-INSPIRE, New Delhi, India for providing fellowship. Authors also are grateful to PCCF (Wildlife), Bangalore, Karnataka, India for granting permission to collect blood samples from turtles.

**References**

[1] Uetz P, Freed P, Holec J, editors. The Reptile Database; 1995 [accessed 2006 May 23]. Available from: http://www.reptile-database.org.

[2] Krishnakumar K, Raghavan R, Pereira B. Protected on paper, hunted in wildlife forensic science. Investig Genet 2011;2:1–9.

[3] Linacre A, Tohe SS. An overview to the investigative approach to species testing in wildlife forensics. Investig Genet 2011;2:1–9.

[4] Norman JA, Moritz C, Limpus CJ. Mitochondrial DNA control region polymorphisms: genetic markers for ecological studies of marine turtles. Mol Ecol 1994;3(4):363–73.

[5] Lahanas PN, Miyamoto MM, Bjorndal KA, Bolten AB. Molecular evolution and population genetics of greater Caribbean green turtles (Chelonia mydas) as inferred from mitochondrial DNA control region sequences. Genetica 1994;94:57–67.
