Asian Journal of Biological and Life Sciences | Original Research

Molecular Differentiation of mitochondrial NADH dehydrogenase subunit-4 and Phylogeny of the genus *Pangshura*

Chittaranjan Baruah[1]*, Papari Devi[2] and Dhirendra K. Sharma[3]

1 Department of Zoology, Darrang College, Tezpur- 784 001, Assam, India.
2 Department of Zoology, Gauhati University, Guwahati- 781 014, Assam, India.
3 School of Biological Science, USTM, Meghalaya, India.

E-mail : chittaranjan_21@yahoo.co.in Contact No. : +91-99542-94080

Submitted : 03.05.2016 Accepted :22.07.2016 Published : 30.08.2016

Abstract

The turtle family Geoemydidae comprising mostly of freshwater turtles includes several highly endangered Southeast Asian turtle species. The present study has been performed on the mitochondrial NADH dehydrogenase 4 (ND4) gene sequence analysis and phylogeny of the Southeast Asian turtle genus *Pangshura* (Testudines: Geoemydidae: *Pangshura*). A comparative *in silico* study has been carried out at the transcriptional and translational level for mRNA structure along with structure and evolution of NADH dehydrogenase enzyme. Evolutionary analyses were conducted by using the methods of Maximum Parsimony, Maximum Likelihood and Neighbor-Joining. Sequence analyses, restriction mapping and RNA structure prediction were performed in the CLC Genomics Workbench 4.0 (CLC Bio, Cambridge, MA, USA). Comparative modeling of NADH4 was conducted by using Modeller9v2 program. The study reveals that *P. smithii* and *P. tentoria* are the sister species followed by *P. tecta* and *P. sylhetensis* their successive sister-taxa. Distinctness within the sub-species of *P. tentoria* is not clear in the evolutionary data of ND4 gene and encoded NADH protein. Although, sequence variation has been observed in ND4 gene, minor structural changes have been observed in some parts of mRNA secondary structure with variation in the minimum folding energy. The computational models of NADH dehydrogenase 4 could be of use for further evaluation of molecular mechanism of function. The present study also provides an indispensable groundwork for future conservation and genetic analysis of other turtle genera.

Key words : Chelonian, Conservation, NADH, Geoemydidae, *in-silico*

INTRODUCTION

The turtle family Geoemydidae includes highly endangered South-east Asian turtles group, mostly are of freshwater turtles[1,2]. Most geoemydids are freshwater turtles. The group also occupies a wide range of habitats, from highly aquatic (*Batagur* and *Pangshura*) to highly terrestrial (*Geoemyda*). The genus *Pangshura* having maximum shell length of 2026.5 cm[3,4], comparatively smaller than *Batagur* (maximum shell lengths 4858 cm; Ernst *et al*. 2000) is characterized by sexual dimorphism, now placed them into two distinct genera.

The genus *Pangshura* includes four small-sized species viz. *P. tecta, P. sylhetensis, P. tentoria* and *P. smithii*[5]. Further these species are fragmented to sub species like *P. tentoria tentoria, P. tentoria circumdata, P. tentoria flaviventer, P. smithii* and *P. smithii pallidipes*. A fifth species, *Pangshura tatrotia*, has been described in 2010[6]. However, all these classification were based on the morphological characterization, after stands as puzzle. Therefore, it is quite relevant that molecular understanding at species and sub-species level of the *Pangshura* group in particular and the turtle group in general to formulate conservation strategies for many of the endangered turtle group. However, few data on genetic variation are available until now and no previous study focused on the mitochondrial ND4 gene of the *Pangshura*. Furthermore, there is no transcriptional and translational level study has been focused on South-east Asian turtles group regarding the RNA structure and molecular structure of various proteins.

Based on the patchy taxon sampling, a first molecular hypothesis for *Batagur, Callagur, Hardella, Kachuga* and *Pangshura* was established by, providing evidence that these genera represent a monophyletic group. Praschag et al.[7] used the mitochondrial Cyt b gene for the same. Based on a complete taxon sampling of all species and subspecies of *Pangshura* and the Gen Bank sequence for the *Batagur*, the present study used sequence variation of a highly informative mitochondrial marker, the ND4 gene and its transcribed RNA and encoded NADH4 protein : (i) to see the concordance of currently available morphological data with genetic data along with spatial pattern of distribution, (ii) to test whether the currently recognized sub-species of *P. tenotoria* correspond well with nd4 clades, (iii) to see the sequence polymorphism among the genera *Pangshura* and *Batagur,* (iv) to characterize the structural changes at mRNA level of the transcribed ND4 gene and (v) to see the Structural and evolutionary relationship of mitochondrial NADH dehydrogenase among the genera *Pangshura* and *Batagur.*

MATERIALS AND METHODS

Taxon sampling

The samples were collected from northeastern and northern India including the bordering areas of Assam (India) and Bangladesh. *A complete taxon sampling of the genus Pangshura was carried out* opportunistically *and one representative of each species/sub-species were* used for the present analysis. Taxonomy and nomenclature of the species were followed after Das (2002)[9] and Prachag *et al*.[7]. Tissue samples from tail clippings were stored in 95% ethanol. Samples collected for the study and their sources are listed in Table 1. Further, two available sequences of genus Batagur and a sequence of out group species (*Lissemys punctata*) were downloaded from GenBank.

DNA extraction, PCR amplification and sequencing

The tissue samples collected were used simultaneously for DNA extraction, sequencing of targeted regions of genome to

Table 1: Currently recognized species and subspecies of *Pangshura* and *Batagur* analyzed in the present study [8].

TAXON	Sample/ sequence code	Origin	GPS Location
P. tecta (Gray, 1831)	P7	Hajo, Kamrup district, Assam	26°14'41.1''; 91°31'37.2''
P. sylhetensis Jerdon, 1870	P12	Brahmaputra river, Biswanath Ghat, Assam	26° 39′ 31.46″ N, 93° 10′ 18.91″ E
P. tentoria tentoria (Gray, 1834)	P30	Hajo,Kamrup, Assam	26 ° 14'41.1''N, 91 ° 31'37.2'' E
P. tentoria circumdata (Mertens, 1969)	P38	Kalindri river	27° 13'23.3" N , 79° 43' 22.9" E
P. tentoria flaviventer Günther, 1864	P43	Burhachapari WLS	26°30'41.3"N, 92°41'14.7"E
P. smithii smithii (Gray, 1863)	P50	Tengatoli char, Morigaon,Assam	26° 29' 016" N, 92° 20' 41.5" E
P. smithii pallidipes (Moll, 1987)	P53	Ganga river	N 27° 12' 50.8", E 79° 41' 34.6"
Batagur kachuga (Gray, 1831)	297185716	Unknown	Unknown
Batagur dhongoka (Gray, 1835)	297185713	Unknown	Unknown
Lissemys punctata (Outgroup)	37963315	Unknown	Unknown

resolve the taxonomy. Total genomic DNA was extracted from samples by standard phenol chloroform technique [10]. The informative regions of mitochondrial ND4 gene were PCR amplified from multiple specimens per species using the taxa specific primer pairs ND4(F) 5'-TGACTACCAAAAG CTCATGTACAAGC-3' (Spinks and Shaffer 2005) [11] and Hist-ND4(R) 5'-CCTATTTTAGAGCCACAGTCTAATG-3' [12]. All PCR amplifications were carried out in 25 µ*L* reaction volume, with 1.5 units of Taq DNA Polymerase (Bangalore Genei, Bangalore, India), 0.25 mM of dNTP's (Bangalore Genei), 2.0 mM of MgCl2, 1 µ*L* l of 0.5 mg/ml of BSA, 0.1 lM (Sigma) of each primer and 40 ng of genomic DNA. The condition for amplification was an initial denaturation temperature 94 °C for five min, followed by 35 cycles of 45 sec at 94 °C, then by 45 sec at 50 °C annealing temperature followed by 2 min at 72 °C, and then by a final extension step for 7 min at 72 °C. PCR products were purified using QIAquick PCR Purification kit (Qiagen) and sequences were obtained commercially from MWG Biotech Pvt. Ltd. (Bangalore).

Molecular phylogenetic analysis

The sequences for the ND4 gene were aligned using ClustalW 1.6 integrated in software MEGA [13], using default parameters. ND4 sequences were translated into amino acids of NADH protein prior to analysis. The mitochondrial dataset were subjected to phylogenetic analyses. Evolutionary analyses were conducted in MEGA5 [13]. The evolutionary history was inferred by using three different methods namely the Maximum Parsimony, Maximum Likelihood and Neighbor-Joining methods**.** Maximum parsimony (MP) tree was estimated using the Close-Neighbor-Interchange algorithm.The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates). The branch lengths calculated using the average pathway method [14].There were a total of 623 positions in the final dataset.

Nucleotide substitution model that best fits each dataset and the model parameters were estimated using Akaike information criterion implemented in the program MODELTEST version 3.7[15]. The Maximum Likelihood analysis was carried out on the Hasegawa-Kishino-Yano model [16]. The tree with the highest log likelihood was considered. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was < 100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The Neighbor-Joining tree was computed using the Tamura-Nei method [17] and are in the units of the number of base substitutions per site. The analysis involved 10 nucleotide sequences. There were a total of 623 positions in the final dataset.

Sequence analysis, in silico RFLP and RNA structure prediction

Sequence analyses *in silico* RFLP and RNA structure prediction were performed in the CLC Genomics Workbench 4.0 (CLC Bio, Cambridge, MA, USA). The nucleotide sequences of the genera *Pangshura* and *Batagur* were *in silico* restriction-digested with 10 restriction enzymes namely AluI, BamHI, FokI, HaeIII, HindIII, KpnI, MspI, NcoI, SalI and TaqII.

Three-dimensional structure prediction

Comparative modeling based on the 3D coordinates of pdb ID 1WB1 Chain A (Crystal structure of translation elongation factor SELB from *Methanococcus maripaludis* in complex with gdp)were conducted by using Modeller9v2 program[18]. The final 3D structures for NADH dehydrogenase 4 *were* evaluated by ERRAT and ProCheck [19].

RESULTS

Evolution of Pangshura based on the ND4 gene sequence

In this study the total sequence length for the mitochondrial ND4 gene was 749 bp. Estimates of Base Composition Bias Difference between Sequences of nd4 gene is shown in Table 1. The three tree building methods yielded very similar tree topologies with minor differences. The differences were at the

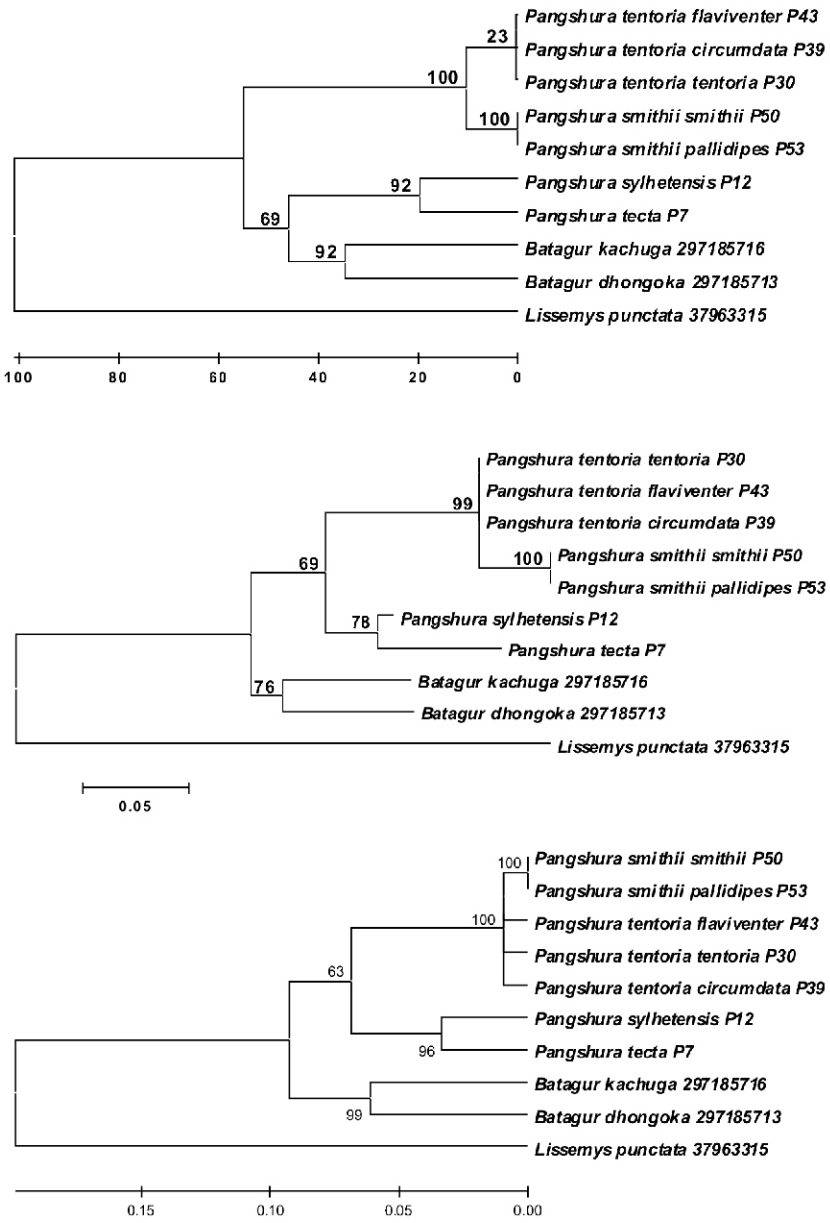


Fig 1 (A-C): Phylogenetic tree based on the ND4 gene sequence analysis. A.Maximum Parsimony tree, B. Maximum Likelihood tree based on the Hasegawa-Kishino-Yano model and C. Neighbor-Joining tree.

base of the tree where the bootstrap support is very low indicating that the relationships are not very robust at deeper nodes. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. All positions containing gaps and missing data were eliminated. There were a total of 623 positions in the final dataset.

All tree-building methods revealed NADH dehydrogenase-4 of *Pangshura* as perfectly supported monophylum with bootstrap or posterior probability values of more than 70% (Fig. 1 A-C). The results are highly consistent with the molecular genetics study involving 12S rRNA genes [20] that within *Batagur and Pangshura,* all species correspond with well-supported clades. However, the evolutionary tree of ND4 gene NADH dehydrogenase 4 supports the fact that *P.tecta* is the closest relative of *P. sylhetensis*, while *P. smithii* and *P. tentoria* are found as sister groups. Distinctness of the subspecies within *P. tentoria* is badly supported by the evolutionary data of NADH dehydrogenase 4 protein. The analysis involved 10 nucleotide sequences (Fig 1 A-C).

A. The evolutionary history was inferred by 12 most parsimonious trees (length = 372) using ND4 gene.

B. The evolutionary history was inferred by using the Maximum Likelihood tree based on the Hasegawa-Kishino-Yano model. The tree with the highest log likelihood (-2348.3680) is shown.

C. The Neighbor-Joining optimal tree with the sum of branch length = 0.67257698 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Tamura-Nei method.

Table 1: Estimates of Base Composition Bias Difference between Sequences of nd4 gene

P. sylhetensis (P12)								
P. tecta (P7)	0.024							
P. tentoria tentoria (P30)	0.136	0.079						
P. tentoria flaviventer (P43)	0.136	0.079	0.000					
P. tentoria circumdata (P39)	0.178	0.108	0.006	0.006				
Pangshura smithii smithii (P50)	0.230	0.175	0.029	0.029	0.016			
P. smithii pallidipes (P53)	0.230	0.175	0.029	0.029	0.016	0.000		
Batagur kachuga (297185716)	0.185	0.246	0.392	0.392	0.392	0.401	0.401	
Batagur dhongoka (297185713)	0.551	0.596	0.663	0.663	0.612	0.573	0.573	0.120
Lissemys punctata (37963315) (outgroup)	0.316	0.502	0.637	0.637	0.727	0.695	0.695	0.474	0.918

Table 2: Estimates of Evolutionary Divergence between nd4 gene sequences

Pangshura_sylhetensis_P12		0.009	0.012	0.012	0.012	0.013	0.013	0.012	0.013	0.018
Pangshura_tecta_P7	0.063		0.013	0.013	0.013	0.014	0.014	0.014	0.014	0.019
Pangshura_tentoria_tentoria_P30	0.093	0.125		0.000	0.002	0.007	0.007	0.014	0.014	0.018
Pangshura_tentoria_flaviventer_P43	0.093	0.125	0.000		0.002	0.007	0.007	0.014	0.014	0.018
Pangshura_tentoria_circumdata_P39	0.095	0.125	0.003	0.003		0.007	0.007	0.014	0.014	0.018
Pangshura_smithii_smithii_P50	0.124	0.156	0.032	0.032	0.035		0.000	0.015	0.015	0.018
Pangshura_smithii_pallidipes_P53	0.124	0.156	0.032	0.032	0.035	0.000		0.015	0.015	0.018
Batagur_kachuga_297185716	0.120	0.149	0.156	0.156	0.157	0.186	0.186		0.012	0.017
Batagur_dhongoka_297185713	0.124	0.144	0.154	0.154	0.152	0.183	0.183	0.111		0.018
Lissemys_punctata_37963315	0.302	0.302	0.294	0.294	0.294	0.311	0.311	0.300	0.307	

The estimated Transition/Transversion bias (*R*) is 1.88. Substitution pattern and rates were estimated under the Kimura (1980) 2-parameter model (+G) (Kimura, 1980). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*], parameter = 0.8134). The nucleotide frequencies are A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. For estimating ML values, a user-specified toplogy was used. The maximum Log likelihood for this computation was -2432.267.

The estimated value of the shape parameter for the discrete Gamma Distribution is 0.6075. Substitution pattern and rates were estimated under the Hasegawa-Kishino-Yano (1985) model (+G) (Hasegawa et al., 1985). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*]). Mean evolutionary rates in these categories were 0.04, 0.21, 0.54, 1.16, 3.06 substitutions per site. The nucleotide frequencies are A = 35.10%, T/U = 25.17%, C = 28.83%, and G = 10.90%. For estimating ML values, a user-specified toplogy was used. The maximum Log likelihood for this computation was -2337.347.

The numbers of base differences per site from and between sequences are shown. Standard error estimate(s) are shown above the diagonal (Table 2). The analysis involved 10 nucleotide sequences. Overall mean distance is 0.145.

Evolution of NADH dehydrogenase- 4 in Pangshura and Batagur

The alignment of translated amino acid sequences revealed

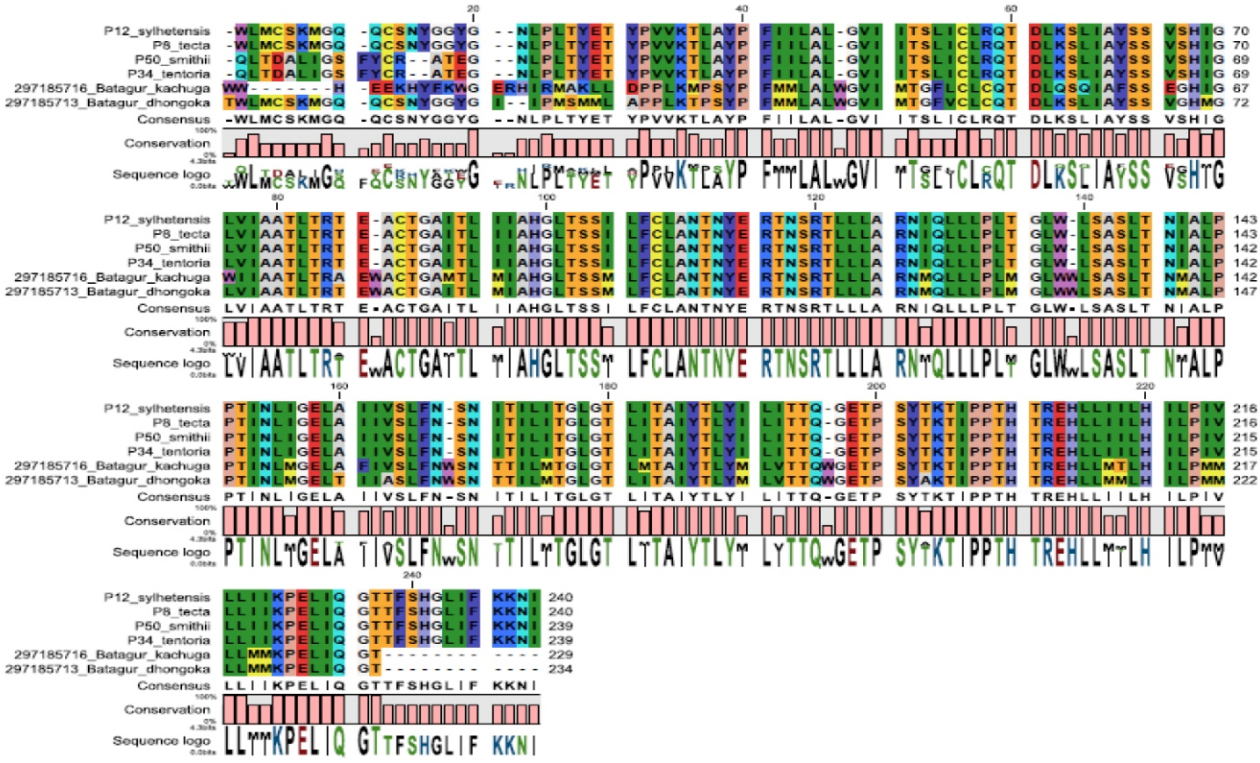


Fig 2: Alignment of NADH protein sequences for *Pangshura* and *Batagur*

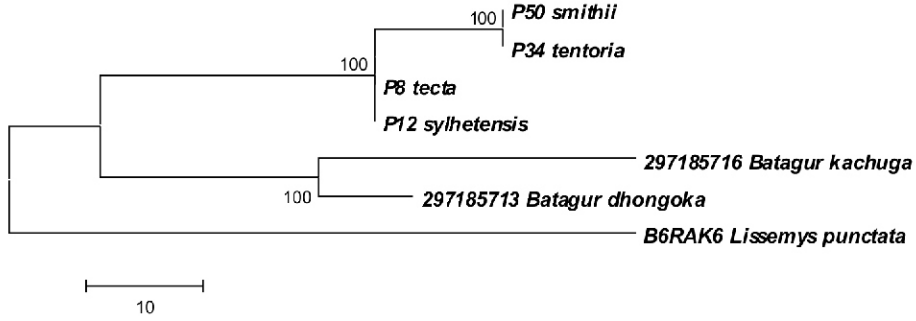
that although there were minor variation in the ND4 gene sequence at the sub-species level of *P. tentoria* and *P. smithii*, the amino acid sequence of NADH dehydrogenase subunit 4 were 100 % identical (Figure 2). Therefore, species level analysis was considered for the structure and evolution of NADH dehydrogenase- 4. All the three tree building method revealed similar topology, except minor variation in the boot strap value. The analysis involved 7 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 223 positions in the final dataset.

Maximum Likelihood Estimate of Gamma Parameter for Site Rates revealed that the estimated value of the shape parameter for the discrete Gamma Distribution is 0.5206. Substitution pattern and rates were estimated under the Adachi-Hasegawa (1996) mitochondrial DNA model (+G) (Adachi and Hasegawa, 1996). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories, [+*G*]). Mean evolutionary rates in these categories were 0.02, 0.17, 0.48, 1.12, 3.21 substitutions per site. The amino acid frequencies are 7.20% (A), 1.90% (R), 3.90% (N), 1.90% (D), 0.60% (C), 2.50% (Q), 2.40% (E), 5.60% (G), 2.80% (H), 8.70% (I), 16.80% (L), 2.30% (K), 5.30% (M), 6.00% (F), 5.50% (P), 7.20% (S), 8.80% (T), 2.90% (W), 3.30% (Y), and 4.40% (V). The maximum Log likelihood for this computation was -1403.854.

The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Nei and Kumar, 2000) with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The tree is drawn to scale; with branch lengths calculated using the average pathway method. (Figure 3 A-C)

A. The evolutionary history was inferred using the Maximum Parsimony method. The consensus tree inferred from 3 most parsimonious trees is shown. Branches corresponding to partitions reproduced in less than 50% trees are collapsed. The consistency index is (0.938272), the retention index is (0.915254), and the composite index is 0.884746 (0.858757) for all sites and parsimony-informative sites (in parentheses).

B. The evolutionary history was inferred by using the Maximum Likelihood method based on the General Reversible



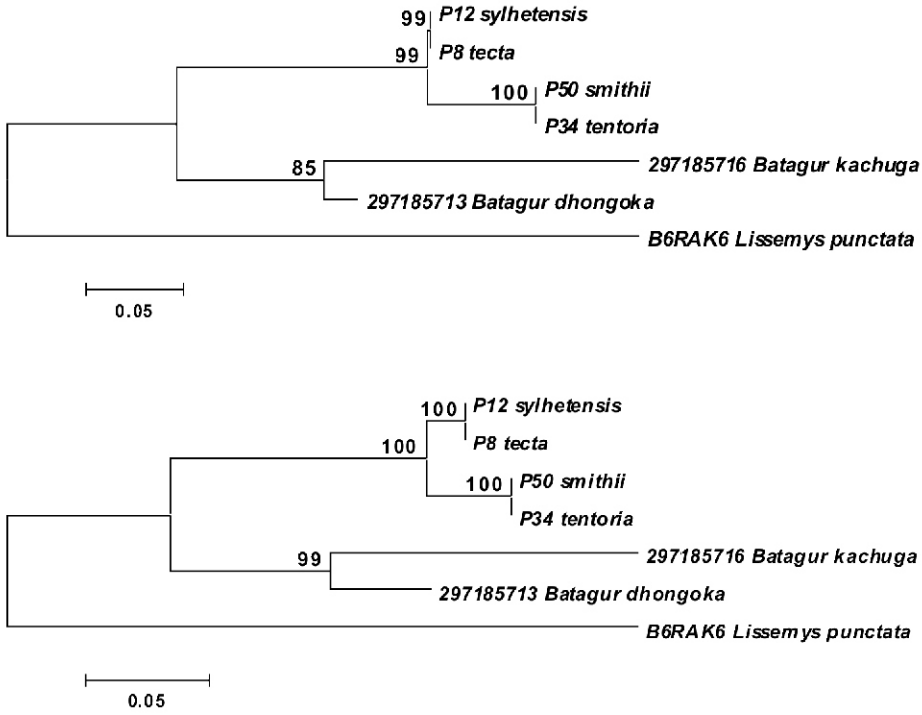


Fig 3 : Evolutionary relationship of NADH dehydrogenase subunit-4 in Pangshura and Batagur.A.Maximum Parsimony tree; B. Maximum Likelihood phylogeny; C. Neighbor-Joining tree.

Mitochondrial + Freq. model. The tree with the highest log likelihood (-1417.7929) is shown. The percentage of trees in which the associated taxa clustered together is shown above the branches.

C. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 0.70710277 is shown.

Sequence analysis, in silico RFLP and RNA structure prediction

Analysis of ND4 sequence alignment position for *P. tecta* and *P. sylhetensis* showed transition (C→T) at 144th, 328th and 346th have shown their genetic identity. Sequence analysis of ND4 gene might be considered to be effective in establishing the closeness of *P. tecta* and *P. sylhetensis*. Multiple sequence alignment of ND4 gene revealed following major transition, transversion, substitution and deletion, helpful for identifying *Pangshura*.

(i) The genus *Batagur* differed from *Pangshura* by transition (C→T) at 77th , (A→G) at 139th , transition (T→C) at 448th, (A→G) at 521st , 662nd, (C/T→A) at 592nd position in the alignment.

(ii) *P. tecta* and *P. sylhetensis* along with *Batagur* sequences have transition (T→C) at 23rd , 228th , 238th , 467th , 499th , 575th , 566th , 612th position, transversion (T→A) at 73rd ,199th , transition (G→A) at 76th , transversion (C→A) at 178th, transition (GA) at 193rd , 535th , transversion (T→G) at 497th position. *P. tecta* along with *Batagur* have transition (T→C) at 131st positions in the alignment.

(iii) *P. tecta* and *P.sylhetensis* have transition (C→T) at 144th , 328th, 346th , transition (A→G) at 397th , 620th , (T→C) at 591st , 649th (G→A) at 631st , transversion (G→T) at 673rd positions in the alignment.

(iv) *P. sylhetensis* sequence have transversion (T→A) at 23rd, (C→A) at 45th , transition (C→T) at 54th ,166th , transition (GA) at 158th position. *P. sylhetensis*16 sequence has a transversion (C→A) at 367th , transition (C→T) at 607th position.

Molecular weight of 666- 681 bp ND4 gene sequence ranged from 212.758 kDa to 217.935 kDa. The nd4 sequences of of the study are found to be AT rich (Table 3).

The haplotypes generated with various enzymes for both the mitochondrial genes were identified on the basis of RFLP patterns are shown in Figure 4 and Tables 4. Identical RFLP patterns generated with the same enzyme in different turtle species were designated as same haplotypes, while dissimilar patterns, as different types. Six, out of the 10 enzymes used, produced species-specific patterns in the 4 *Pangshura* species.

The sequence analysis indicated that within the genus *Pangshura*, there are minor sequence variations. Analysis of 16S rDNA sequence showed that there is minor variation in melting temperature and being ribosomal RNA gene sequence, there is no coding regions defined in sequence. Minor structural changes have been observed in some parts of 16S rRNA secondary structure with variation in the minimum folding energy.

mRNA structure: The present study find very clear signs that the mRNAs had an evident evolutionary relationship, although it is not totally clear whether this was a relationship of homology or complementarity between the 5prime and 3prime halves of mRNAs. The minimum enrgy nd4 mRNA structure have been shown in Table 5 and Figure 5. Overall, these data favour models for the evolution of the mRNA molecule postulating that a duplication event involving hairpin structures.

Table 3: Results of comparative nucleotide sequence statistics of the conserved region of mitochondrial ND4 gene among the 4 species of *Pangshura*

Information	Taxon						
	P. tecta	*P. sylhetensis*	*P. smithii smithii*	*P. smithii pallidepes*	*P. t. tentoria*	*P. tentoria circumdata*	*P. tentoria flaviventer*
Length	680	681	666	674	681	681	681
Molecular weight (kDa)	217.625	217.92	212.758	215.359	217.919	217.935	217.919
Melting temperature [salt] = 0.1M	80.94	81.16	80.97	81.14	81.10	81.16	81.10
C + G	266	270	261	267	269	270	269
A + T	414	411	405	407	412	411	412

Table 4: Summary of restriction digestion of *nd4* gene with 10 restriction enzymes namely AluI, BamHI, FokI, HaeIII, HindIII, KpnI, MspI, NcoI, SalI and TaqII.

Taxon	Total number of restriction sites
P. sylhetensis (P12)	8
P. tecta (P7)	8
P. tentoria tentoria (P30)	4
P. tentoria flaviventer (P43)	4
P. tentoria circumdata (P39)	4
Pangshura smithii smithii (P50)	7
P. smithii pallidipes (P53)	6
Batagur kachuga (297185716)	7
Batagur dhongoka (297185713)	8

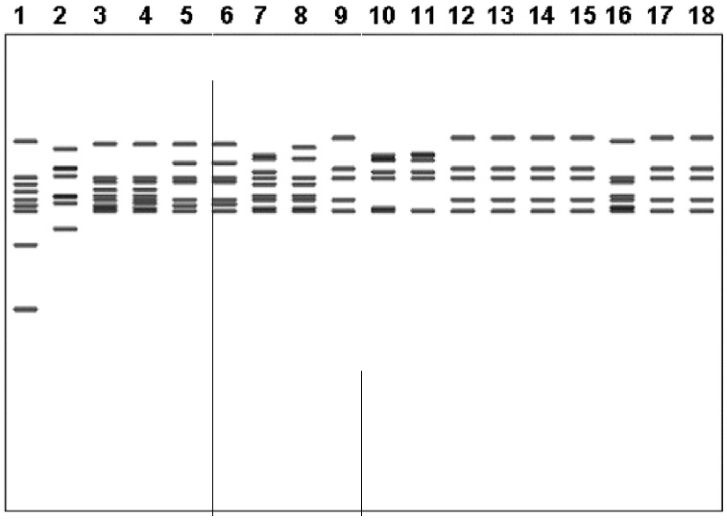
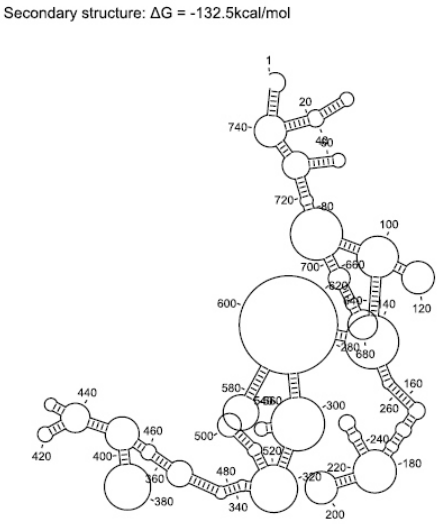


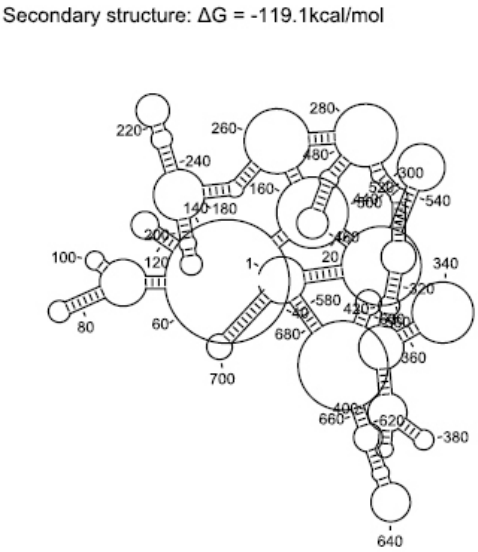
Table 4: *In silico* Restriction digestion profile for ND4 gene using 10 Restriction enzymes. [1. *Batagur dhongoka* , 2. *Batagur kachuga* 1, 3. *P. smithii pallidipes* 1, 4. *P. smithii pallidipes* 2, 5. *P. smithii smithii* 1, 6. *P. smithii smithii* 3, 7. *P. sylhetensis* 1, 8. *P. sylhetensis* 16, 9. *P. t. circumdataX flaviventer,* 10. *P.tecta* 1, 11. *P. tecta* 9, 12. *P. tentoria circumdata* 1, 13. *P. tentoria circumdata* 3, 14. *P. tentoria flaviventer* 1, 15. *P. tentoria flaviventer* 3, 16. *P. t tentoriaXflaviventer* , 17. *P. tentoria tentoria* 1, 18. *P. tentoria tentoria* 6]

Table 5: nd4 gene mRNA structure

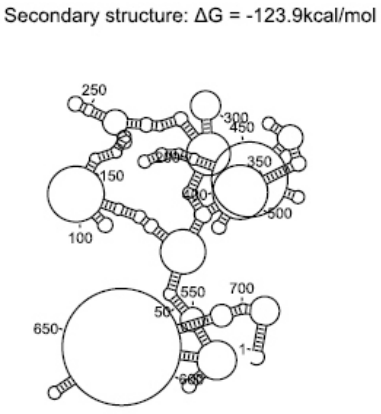
Taxon	Minimum folding energy, ΔG (kcal/mol)
P. sylhetensis (P12)	-119.1
P. tecta (P7)	-132.5
P. tentoria tentoria (P30)	-123.6
P. tentoria flaviventer (P43)	-114.7
P. tentoria circumdata (P39)	-117.9
Pangshura smithii smithii (P50)	-123.9
P. smithii pallidipes (P53)	-128.3
Batagur kachuga (297185716)	-138.3
Batagur dhongoka (297185713)	-142.3



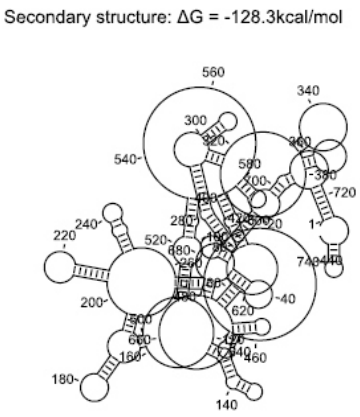
P. tecta (P7)



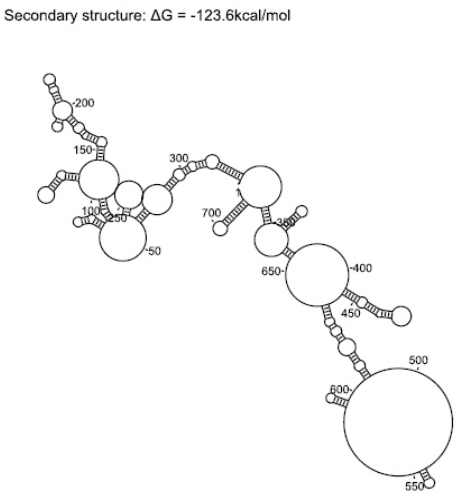
P. sylhetensis (P12)



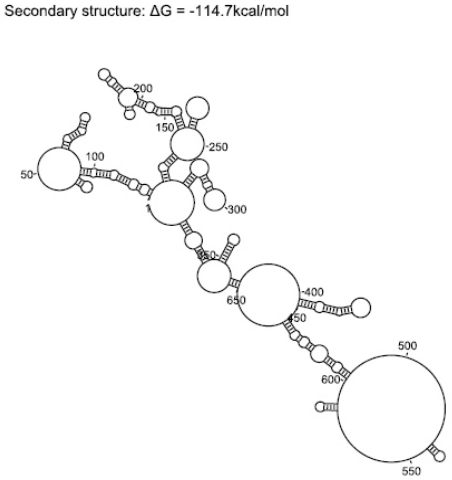
Pangshura smithii smithii (P50)



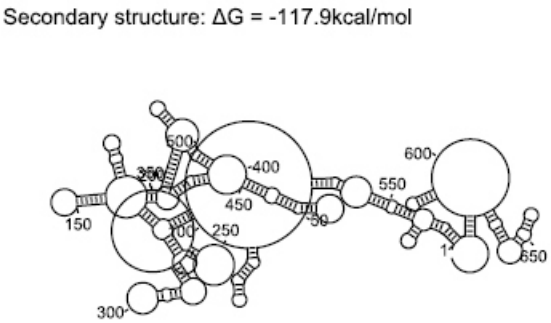
P. smithii pallidipes (P53)



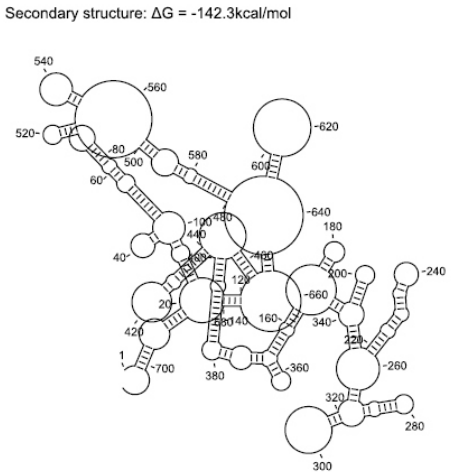
P. tentoria tentoria (P30)



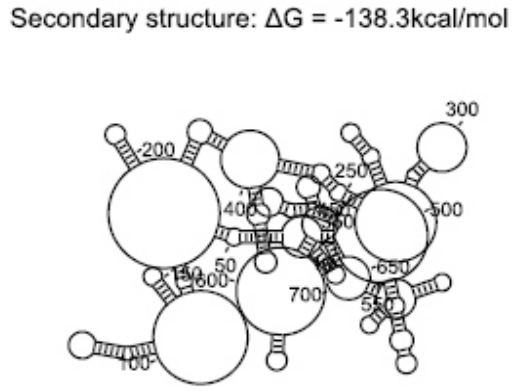
P. tentoria flaviventer (P43)



P. tentoria circumdata (P39)



Batagur dhongoka (297185713)



Batagur kachuga (297185716)

Fig 5: Secondary structure of nd4 mRNA (2D view)

The NADH4 protein has isoelectric point 9.15 in *P. tecta* and *P. sylhetensis* and 8.92 *in P. smithii* and *P. tentoria.* The aliphatic index of NADH4 is 141.042 in *P. tecta* and *P. sylhetensis* and 145.732 in *P. smithii* and *P. tentoria* (Table 6).

The computational model of NADH dehydrogenase-4 of *P tecta* has 2 sheets, 3 beta alpha beta units, 3 beta hairpins, 1 beta bulge, 10 strands, 5 helices, 3 helix- helix interacs, 20 beta turns and 1 gamma turns. The model of *P. sylhetensis* has 4 sheets, 2 beta alpha units, 3 beta hairpins, 3 beta bulges, 12 strands, 5 helices, 2 helix- helix interacs, 20 beta turns and 4 gamma turns. *P. tentoria* has 2 sheets, 3 beta alpha beta units, 2 beta hairpins, 1 beta bulge, 9 strands, 5 helices, 3 hellix-helix interacs, 27 beta turns, 3 gamma turns.*P. smithii* has 3 sheets, 2 beta alpha beta units, 3 beta

Table 6: Amino acid sequence statistics of NADH dehydrogenase subunit-4

Protein statistics	*P. tecta*	*P. sylhetensis*	*P. smithii*	*P. tentoria*
Number of amino acids	240	240	239	239
Molecular weight	26.203 kDa	26.203 kDa	26.035 kDa	26.035 kDa
Isoelectric point	9.15	9.15	8.92	8.92
Total number of negatively charged residues (Asp + Glu	8	8	10	10
Total number of positively charged residues (Arg + Lys)	13	13	13	13
Aliphatic index	141.042	141.042	145.732	145.732
Total number of Hydrophobic residues (A,F,G,I,L,M,P,V,W)	134	134	134	134
Total number of Hydrophilic residues (C,N,Q,S,T,Y)	79	79	76	76

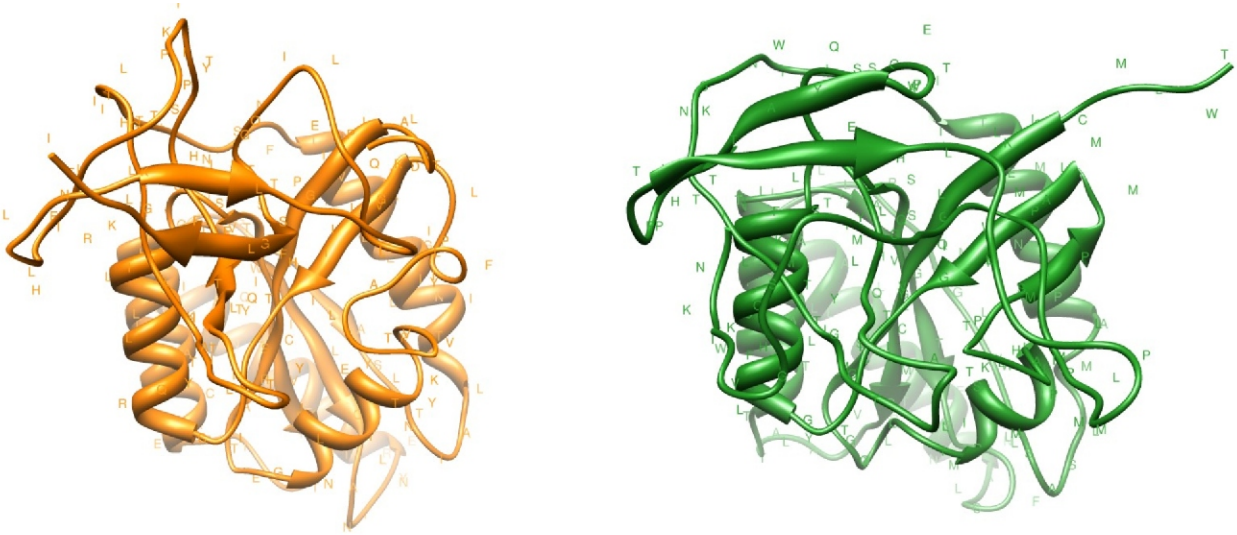
hairpins, 1 beta bulge, 11 strands, 5 helices, 3 helix-helix interacs, 18 beta turns, 3 gamma turns. *Batagur kachuga* has 3 sheets, 2 beta alpha beta units, 2 beta hairpins, 7 strands, 7 helices, 2 helix-helix interacs, 23 beta turns 4 gamma turns. *Batagur dhongoka* has 3 sheets, 2 beta alpha beta units, 2 beta hairpins, 1 psi loop, 2 beta bulge, 10 strands, 5 helices, 3 helix-helix interacs, 30 beta turns, 7 gamma turns(Fig. 6).

Procheck verification proved that the models are of good quality as judged by Ramachandran Plot (Fig. 7). The overall Quality factors predicted by ERRAT verification programme for the predicted 3D structures of NADH dehydrogenase 4are more than 93% (Fig. 8). The structures were successfully deposited to PMDB Protein Model Database [21] of University of Rome and now available for download. Each 3D structure of NADH dehydrogenase 4 has been assigned a unique PMDB ID for the coordinate entry.

ProFunc analysis revealed that the the predicted protein structures belong to nadh-ubiquinone oxidoreductase (4.89) dehydrogenase (1.07) gropup. The biological process associated is ATP synthesis coupled electron transport (0.80) oxidation reduction (0.80) cellular process (0.80) cellular metabolic process (0.80). The Biochemical function: NADH dehydrogenase (ubiquinone) activity (0.80) catalytic activity (0.80) oxidoreductase activity (0.80) oxidoreductase activity\, acting on NADH or NADPH (0.80).

DISCUSSION

Use of ND4 sequence alignment in the monophyletic analysis of *Pangshura species* is not known. Yet, Praschag *et al.*[7] suggested using Cyt *b* that within *P. smithii* there is no phylogenetic differentiation at all paralleling the two recognized subspecies of *P. s. smithii* and *P.s. pallidipes.*Introduction of sequence analysis in the genetic identification of freshwater turtle species is not known and this would be the first report of its kind. Vargas-Ramirez *et al.*[22] aligned sequences for protein coding regions of ND4, cyt *b*, CO1 and Rag2 and obtained no insertion or deletion in Malagasy river turtles *Podocnemis.* However, they



A **B**

Fig 6: IThe predicted 3D structures of NADH dehydrogenase 4displayed by UCSF Chimera (A. *Pangshura*, B. *Batagur*)

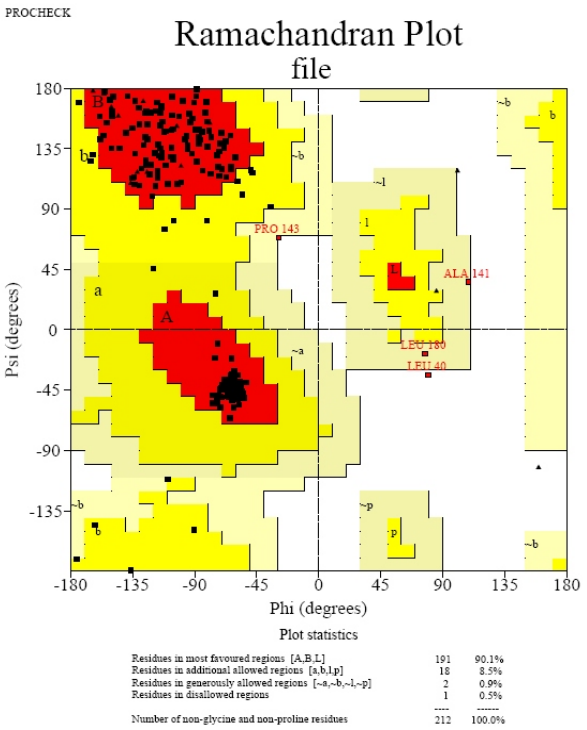


Fig 7: Ramachandran analysis of the backbone dihedral angles PSI (y) and PHI (f) for the final structure of NADH dehydrogenase 4 protein (from *P. tecta*) Red region represents the most favored region, yellow = allowed region, light yellow = generously allowed region, white = disallowed region [ProCheck].

obtained a highly variable region with several insertions and deletions in 'D loop' in *P. expansa*.In their study, the base frequencies for ND4 gene were A=27.99%, C=12.28%, G=24.01% and T=35.71%. Both these protein coding gene regions were A and T rich, exhibiting a pattern typical of protein-coding mtDNA loci[23].

The present study suggests that *P. smithii* and *P. tentoria* as sister species followed by *P. tecta* and *P. sylhetensis* as their successive sister-taxa. Distinctness within the three currently recognized sub-species of *P. tentoria* is not clearly visible by the evolutionary analysis. The predicted 3D structures can serve as a guide for the allocation of amino acid residues involved in each fold, which is important for further investigations on molecular mechanism of functions. The molecular evolutionary analysis underline that detailed population genetics study on *P. tentoria* is essential for developing effective conservation strategies. A detailed phylogeographic study by the authors is in progress. The molecular method developed in the present study is simple, rapid, reliable and reproducible; hence could be routinely applied for species identification; essential for conservation and management of endangered chelonian species.

The structure NADH dehydrogenase 4 can be helpful in structural biology for further investigations on allocation of amino acid residues in each fold, prediction of active sites, molecular mechanism of function. The present analysis corroborate that the genus *Pangshura* is monophyletic. The modeling of NADH dehydrogenase 4 of *genus Pangshura* gains importance for the structural biology and even to the conservation genetic research from several angles.

The present evolutionary study on NADH dehydrogenase 4 (gene and protein) provides a stable phylogenetic hypothesis for all *Pangshura* species, with the suggestion that *P.smithii* and *P.tentoria* as sister species followed by *P.tecta* and *P.sylhetensis* as their successive sister-taxa. The present study provides an indispensable groundwork for future molecular analyses at the protein level. The choice of molecular data is crucial for phylogenetic analyses and molecular studies can now be tailored specifically for particular phylogenetic groups and/ or questions.

CONCLUSION

The present study can be used as an additional method for identification of species as well as for identification of unknown samples with unusual appearances and could be made available for the identification of confiscated specimens. The predicted 3D structures presented here can serve as a guide for the allocation of amino acid residues involved in each fold, which is important for further investigations on molecular mechanism of functions. The molecular evolutionary analysis underline that further sampling is in dire need for developing effective conservation strategies.

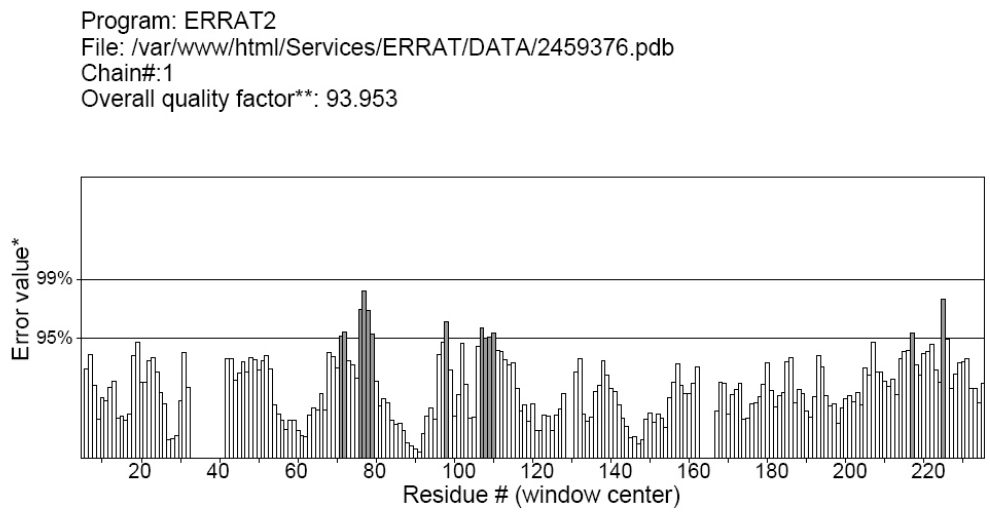


Fig 8: ERRAT verification for the overall Quality factors of the predicted 3D structures of NADH dehydrogenase 4 (Pangshura smithii)

Pangshura represent distinct genera with four well supported species. Thus, the present analysis could be deployed in resolving the position and lineage of each species and subspecies of the genus *Pangshura.* The AT rich gene in the study was suggestive of the monophyletic origin of *Pangshura*, though such data are not readily available to support this proposition. Secondary structure of *nd4* mRNA structure has presented distinct variation in different taxon of this investigation. As in the present study, detection of transition or transversion or deletion of nucleotide base(s) at specific locus in the genome could readily help in species identification. Larger number of sample groups and complete genome information are needed towards getting a definitive understanding of the grand evolution problem.

ACKNOWLEDGEMENTS

The authors are thankful to the Ministry of Environment and Forests, Govt. of India for necessary approval letter and the state forest departments of Assam, West Bengal and Uttar Pradesh for necessary permission to carry out the research. The authors are also thankful to Rufford Small Grants Foundation, UK for financial support during field work.

REFERENCES

1. Van Dijk PP, Stuart BL, Rhodin AGJ. Asian Turtle Trade. Proceedings of a Workshop on Conservation and Trade of Freshwater Turtles and Tortoises in Asia, Phnom Penh, Cambodia (14 December 1999). *Chelonian Research Monographs.* 2000: 2: 1164.

2. IUCN [International Union for the Conservation of Nature and Natural Resources]. IUCN Red List of Threatened Species, 2014:*www.iucnredlist.org*.

3. Ernst CH, Altenburg RGM, Barbour RW. Turtles of the World- World Biodiversity Database, CD-ROM Series, Windows, Version 1.2. Amsterdam: Biodiversity Center of ETI, 2000.

4. Sarma PK. Habitat Ecology, population status and distribution of *Kachuga sylhetensis* (Jerdon) in certain districts of Assam. PhD. Thesis (Unpublished), Gauhati University, 2007. P. 137.

5. Das I. Die Schildkröten des Indischen Subkontinents. Frankfurtam Main: Chimaira.2001.

6. Walter GJ, Tyler RL. *Pangshura tatrotia*, a new species of pond turtle (Testudinoidea) from the Pliocene Siwaliks of Pakistan. *Journal of Systematic Palaeontology*, 2010:8(3): 449458.

7. Praschag P, Hundsdorfer AK, Fritz U. Phylogeny and taxonomy of endangered South and South-east Asian freshwater turtles elucidated by mtDNA sequence variation (Testudines: Geoemydidae: Batagur, Callagur, Hardella, Kachuga, Pangshura). *Zoologica Scripta*, 2007: 36: 429442.

8. Fritz U, Havas P. Checklist of Chelonians of the World. Bonn and Dresden: German Federal Ministry of Environment, Nature Conservation and Nuclear Safety and Museum of Zoology Dresden, 2007.

9. Das I. A photographic guide to snakes and other reptiles of India. *New Holland Publishers (UK) Ltd, Garfield House, 86-88 Edgware Road, London w2 2EA, UK*, 2002. p.144.

10. Sambrook J, Russell DW. Molecular cloning: a laboratory manual, 3rd ed. New York: Cold Spring Harbor Laboratory Press, New York, 2001.

11. Spinks PQ, Shaffer HB. Conservation phylogenetics of the Asian box turtles (Geoemydidae, Cuora): mitochondrial introgression, numts, and inferences from multiple nuclear loci. *Conservation Genetics*, 2007: 8: 641657.

12. Engstrom TN, Shaffer, HB, McCord WP. Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trionychidae). Syst Biol., 2004: 53: 693710.

13. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution,* 2013: 30: 2725-2729.

14. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. Oxford University Press, New York, 2000.

15. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 1998: 14: 817818.

16. Hasegawa M, Kishino H, Yano T. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol*., 1985: 22: 160-174.

17. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution,* 1993:10:512-526.

18. Marti-Renom MA, Stuart A, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Bimolecular. Structure,* 2000: 29, 291-325.

19. Laskowski RA, Watson JD, Thornton JM. From protein structure to biochemical function. *Journal of Structural and. Functional Genomics,* 2003: 4:167177.

20. Shaffer HB, Meylan PA, McKnight ML. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. *Systematic Biology,* 1997: 46: 235-268.

21. Tiziana C, Paolo DM, Domenico C, Ivano GT, Anna T. The PMDB Protein Model Database CASPUR, *Nucleic Acids Research*, 2006 34 (Database Issue): D306-D309.

22. Vargas-Ramirez M, Castano-Mora OV, Fritz U. Molecular phylogeny and divergence times of ancient South American and Malagasy river turtles (Testudines: Pleurodira: Podocnemididae). *Org. Divers. Evol*. 2008: 8: 388398.

23. Daniels SR, Hofmeyr MD, Henen BT, Crandall KA. Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise Chersina angulata. *Mol. Phylogenet. Evol*., 2007: 45: 915926.