International Journal of Molecular Biology
ISSN: 0976-0482 & E-ISSN: 0976-0490, Volume 6, Issue 1, 2015, pp.-102-111.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000235

SEQUENCE ANALYSIS, STRUCTURE PREDICTION AND PHYLOGENY OF GLUTAMINE SYNTHETASE OF TWO EEL SPECIES, *Monopterus cuchia* (HAMILTON) AND *Monopterus albus* (ZUIEW)

DEVI P.[1], BARUAH C.[2] AND SHARMA D.K.[1]*

[1]Department of Zoology, Gauhati University, Guwahati- 781 014, Assam, India.
[2]Department of Zoology, Darrang College, Tezpur- 784 001, Assam, India.
*Corresponding Author: Email- dksguniv@gmail.com

Received: December 20, 2014; Revised: January 06, 2015; Accepted: January 08, 2015

Abstract- The present investigation includes *in silico* sequence analysis, structure prediction and phylogeny of Glutamine Synthetase from freshwater bony fishes. The analyses were performed for the sequence data of three glutamine synthetase genes namely *gs01, gso2* and *gso3* and their translated Gutamine Synthetase enzymes (GS-I, GS-II and GS-III). The evolutionary analyses were performed using Maximum Likelihood and Maximum Parsimony methods. The structures of Gutamine Synthetase were predicted using the template of Crystal Structure of Human Glutamine Synthetase (PDB ID: 2OJW). The computed instability index (40.60 to 46.96) classifies GS of *M. albus* as unstable. After verification, the structures of GS have been deposited to Protein Model Database (PMDB). The predicted structures of GS could be of use for further evaluation of molecular mechanism of function.

Keywords- bony fish, Gutamine Synthetase, *in-silico*, multifunctional enzyme

Citation: Devi P., Baruah C. and Sharma D.K. (2015) Sequence Analysis, Structure Prediction and Phylogeny of Glutamine Synthetase of Two Eel Species, *Monopterus cuchia* (Hamilton) and *Monopterus albus* (Zuiew). International Journal of Molecular Biology, ISSN: 0976-0482 & E-ISSN: 0976-0490, Volume 6, Issue 1, pp.-102-111.

Introduction

Glutamine Synthetase (GS; L-glutamate-ammonia ligase, EC 6.3.1.2) is a multifunctional enzyme, catalyzes the ATP-dependent conversion of glutamate and ammonium to glutamine. The formation of Glutamine plays an important role in nucleotide biosynthesis, neurotransmitter metabolism, amino acid balance and ammonia detoxification. Gene sequences for glutamine synthetase have been reported in several fish species [1-3]. GS is critical in the detoxification process of the highly mobile and toxic ammonia [4,5]. Recently, four glutamine synthetase isoforms were isolated from adult bony fish (GS-I - GS-IV) [3,6]. All four genes were expressed during early development, but only GS-I and GS-II were expressed at appreciable levels in liver of adults and expression was very low in muscle tissue. The high level of expression of GS-I and GS-III prior to hatching corresponded to a linear increase in glutamine synthetase activity.

Although, there has been availability of sequence information for GS from different fish groups, yet species-specific structural information are lacking. Therefore, the biochemistry and molecular mechanism of their functions in fishes are still not very well understood due to lack of their structural information. Thus, an attempt has been made for sequence analysis, 3D structure prediction [7] and phylogeny of of GS from *M. cuchia* and *M. albus*.

Materials and Methods

Acquisition and Alignment of Sequences

The study was extended to data mining and sequence analyses of *gs* gene (*gs01, gso2* and *gso3*) and GS protein from the sequence information extracted from GenBank (NCBI) and Protein Knowledgebase (UniProtKB), respectively [Table-1], [Table-2], [8,9]. The sequences were simultaneously aligned using CLUSTAL-W [20] and Modeller 9.12 [21] programs.

Comparative Sequence Analysis

The nucleotide (*gs* genes) and protein (GS) sequence analyses and were performed in the CLC Genomics Workbench 7.0.3 (CLC Bio, Hyderabad). The physico-chemical parameters of GS were computed using CLC Genomics Workbench and ProtParam [10]. The important calculations for the amino acid composition, atomic composition, molecular weight, Formula, theoretical pI, extinction coefficients, instability index, half-life, hydrophobicity, aliphatic index, charge *vs.* pH were carried out under sequence analysis.

Molecular Phylogenetic Analysis

The sequences for the *gs* gene were separately aligned using ClustalW 1.6 [11] integrated in software MEGA6 [12], using default parameters. *gs* sequences were translated into amino acids of GS protein prior to analysis. Both *gs* and GS datasets were subjected

to phylogenetic analyses. Evolutionary analyses were conducted in MEGA6 [12]. The evolutionary history was inferred by using the Maximum Parsimony [13] and Maximum Likelihood methods [14].

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 [Table-3] [15].

Tertiary Structure Prediction

BlastP [16] and FASTA [17] searches were performed independently with PDB [18] for obtaining an approriate template. The significance of the BLAST results was assessed by expect values (e-value) generated by BLAST family of search algorithm [17]. The target-template alignment [19] was carried out using ClustalW 2.1 [20] and Modeller 9.12 [21] programmes. Comparative (Homology) modelling was conducted by the Modeller 9.12 [22]. The loop regions were modeled using MODLOOP server [23]. The final 3D structures with all the coordinates for GS were obtained by optimization of a molecular probability density function (pdf) of Modeller [24]. The molecular pdf for homology modelling was optimized with the variable target function procedure in Cartesian space that employed the method of conjugate gradients and molecular dynamics with simulated annealing [25].

The 3D structures for GS were evaluated [26] by ERRAT [27] and ProCheck [28] programmes. After fruitful verification, the coordinate files were successfully deposited to PMDB [29] and RasMol programs [30].

Preliminary investigations of function of the modelled proteins were performed from the 3D structure using ProFunc [28, 31]. A number of databases like PFam, PROSITE, PRINTS, ProDom, InterProScan [32] were used for functional characterization.

Results

Comparative Sequence Analysis

The three *gs* genes of the present study ranged from 541(gs01 of *M. cuchia*) to 1951(gs03 of *M. albus*) nucleotide long and with molecular weights 167.581 kDa to 603.143 kDa. The melting temperature range was from 83.72 (gs02 of *M. albus*) to 86.49 (gs02 of *M. cuchia*) at 0.1M salt concentration [Table-1]. The frequency of GC ranged from 0.443 (in gs03 of *M. albus*) to 0.526 (in gs02 of *M. albus*). On the other hand frequency of AT in *gs* mRNA (cDNA) sequence in different fishes of the present study ranged from 0.477 (in gs01 of *M. cuchia*) to 0.558 (in gs03 of *M. albus*) [Table-1]. The sequences of *M. albus* were found to be A:T rich for all the three gs gene sequences. On the other hand, the gs genes of *M. cuchia* were very rich in G:C frequency [Table-1], [Fig-1].

Table 1- Nucleotide sequence statistics of the *gs* cDNA sequence

Statistical parameter	Gs01		Gs02		Gs03	
	M. cuchia	*M. albus*	*M. cuchia*	*M. albus*	*M. cuchia*	*M. albus*
GenBank Accession numbers	HQ667788	JF694448	KF672794	JF694447	JX112754	JF694446
Length (bp)	541bp	1,471bp	868bp	1,704bp	938bp	1,951bp
MW in single stranded condition (kDa)	167.581 kDa	455.436 kDa	268.773 kDa	527.099 kDa	290.571 kDa	603.143 kDa
Melting temperature (^{0}C) [salt] = 0.1M	86.35	84.33	86.49	83.72	86.41	83.01
Frequency of A + T	0.477	0.526	0.474	0.541	0.475	0.558
Frequency of C + G	0.523	0.474	0.526	0.459	0.525	0.442

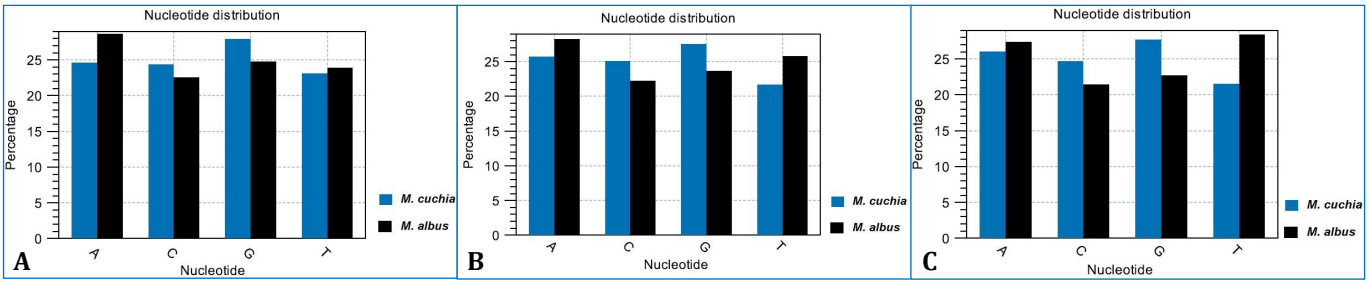

Fig. 1- Comparative nucleotide composition (% in average) in the *gs* cDNA sequence of *M. cuchia* and *M. albus*; **(A)** *gs01* gene, **(B)** *gs02* gene, **(C)** *gs03* gene.

The sizes of protein sequences of GS enzyme in the present study ranged from 180 (GS-I of *M.* cuchia) to 377 (GS-II of *M. albus*) amino acids. The amino acid Glycine (frequency=0.111) and Proline (frequency=0.083) has been found predominantly rich in the GS-I of *M. albus*. However, GS-I of *M. albus* was found to be rich in Glycine (frequency= 0.094) and Glutamic Acid (frequency= 0.075). GS-II of *M. cuchia* is rich in Glycine (frequency=0.114) and Arginine (frequency=0.069), whereas *M. albus* is rich in Glycine (frequency= 0.095) and Alananine (frequency= 0.069). But the GS-III of *M. cuchia* is found to be rich in Glycine (frequency=0.109) and Arginine (frequency=0.071) where *M. albus* is rich in Glycine (frequency= 0.108) and Glutamic Acid (frequency= 0.075) [Fig-2]. Sequence analysis of GS revealed -ve hydropathy on average (-0.366 to -0.661) [Table-2]; [Fig-3]. The molecular weight of GS in *Monopterus* ranged from 20.186 kDa (in GS-I of *Monopterus cuchia*) 41.572 kDa (in GS-III of *M. albus*). The isoelectric point of the GS ranged from 5.59 (in GS-III of *M. albus*) to 7.73 (in GS-II of *M. cuchia*) [Table-2]; [Fig-4]. The Instability index of GS of the present study ranged from 25.19 to 46.96 [Table-2]. Multiple amino acid sequence alignment of Glutamine Synthetase protein of M. cuchia and M. albus showed higher degree of conservation of respective amino acid in each alignment position for GS-I, GS-II, and GS-III, respectively [Fig-5].

Molecular Evolution of gs Genes

The Maximum-likelihood model parameters for data sets as estimated in Modeltest [15] are listed in [Table-3]. Pairwise distances of *gs01*, *gs02* and *gs03* genes have been depicted in the [Table-4] [Table-5], [Table-6] respectively. The bootstrap consensus tree

inferred from 1000 replicates was taken to represent the evolutionary history of the taxa analyzed [33]. There were a total of 1107, 1118 and 545 positions in gs01, gs02 and gs03 gene final dataset.

gs01 Gene Phylogeny

The Pairwise distance of *gs01* gene sequences among the different fish species of the present study revealed shortest genetic distance (0.022) *Haplochromis burtoni* and *Oreochromis niloticus. Monopterus albus* and *Tetraodon nigroviridis* have the longest genetic distance (0.369). The gs01 gene of *Monopterus cuchia* and *Monopterus albus* showed a genetic distance of 0.133, which is the shortest genetic distance for *M. cuchia* among the gs01 sequences of nine species. The longest genetic distance was showed by *Monopterus cuchia* (0.302) with *Lepisosteus oculatus* [Table-4].

Table 2- Glutamine Synthetase protein statistics

Statistical parameter	GS-I		GS-II		GS-III	
	M. cuchia	*M. albus*	*M. cuchia*	*M. albus*	*M. cuchia*	*M. albus*
UniProtKB Accession number	E7EDT7	H2BL60	V5UW28	H2BL59	I7BEK1	H2BL58
No. of amino acids	180	371	289	377	312	371
MW (kDa)	20.186	41.433	32.437	42.229	34.944	41.572
pI	6.06	5.87	7.73	6.61	6.35	5.59
-ve charged residues	20	46	34	48	42	52
+ve charged residues	17	38	35	46	39	41
Formula	$C_{907}H_{1353}N_{245}O_{253}S_{14}$	$C_{1815}H_{2793}N_{513}O_{554}S_{24}$	$C_{1433}H_{2179}N_{415}O_{419}S_{16}$	$C_{1855}H_{2862}N_{526}O_{560}S_{23}$	$C_{1532}H_{2351}N_{443}O_{462}S_{18}$	$C_{1830}H_{2790}N_{516}O_{553}S_{22}$
II	25.19	46.96	39.63	40.88	42.09	40.6
AI	62.83	67.55	60.07	63.93	60.03	63.13
GRAVY	-0.366	-0.506	-0.615	-0.582	-0.661	-0.593

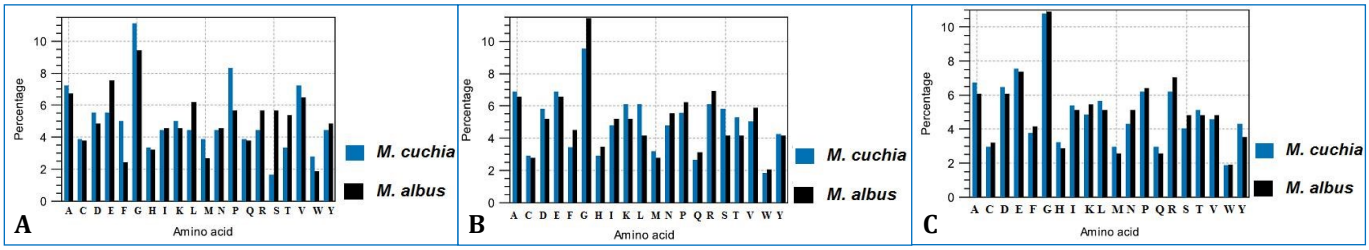

Fig. 2- Distribution of amino acids for Glutamine Synthetase protein in the ornamental fishes; **(A)** GS-I, **(B)** GS-II, **(C)** GS-III.

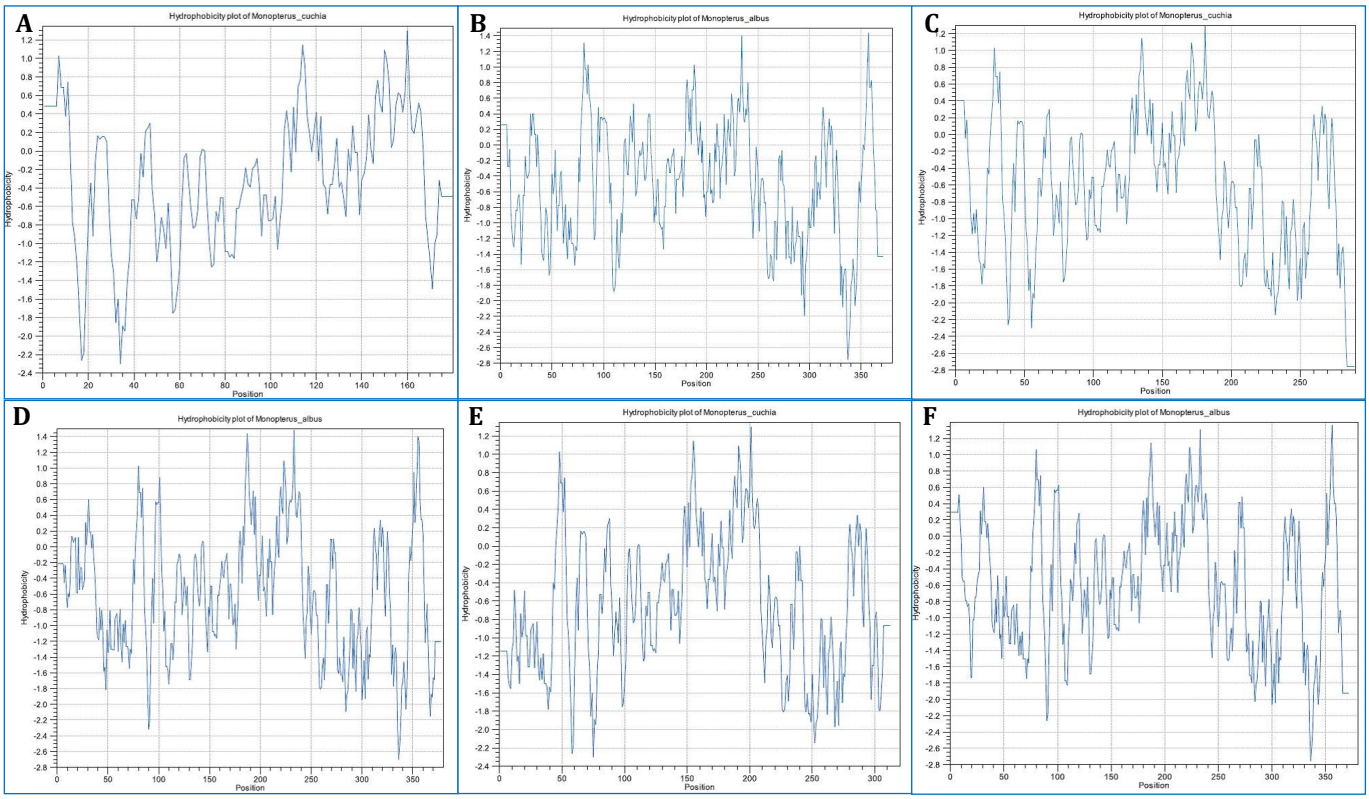

Fig. 3- Plot of local Hydropathy for GS (Kyte-Doolittle scale [38]). **(A)** GS-I of *M. cuchia*, **(B)** GS-I of *M. albus,* **(C)** GS-II of *M. cuchia*, **(D)** GS-II of *M. albus*, **(E)** GS-III of *M. cuchia*, **(F)** GS-III of *M. albus*.

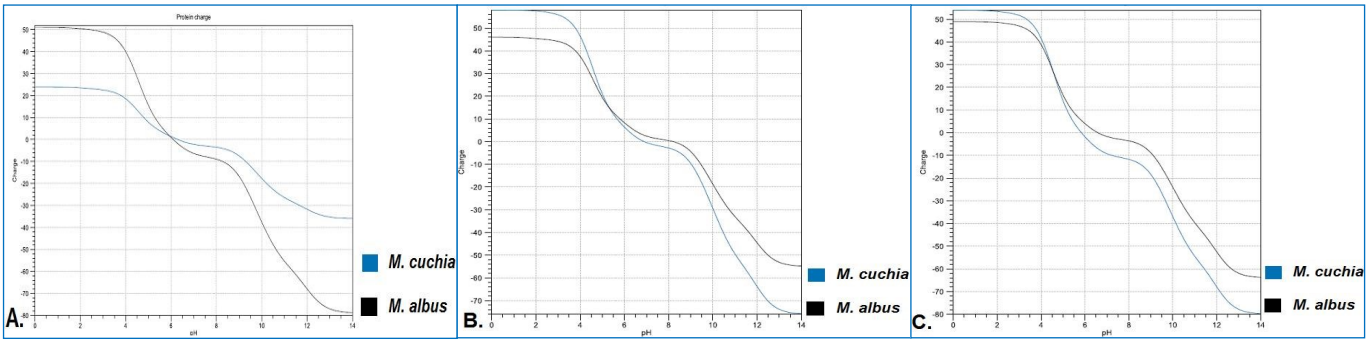

Fig. 4- Electrical charge as a function of pH for GS in the ornamental fishes; **(A)** GS-I, **(B)** GS-II, **(C)** GS-III.

Table 3- Maximum-likelihood model parameters for data sets as estimated in Modeltest [15]

Parameter	Gs01 gene	Gs02 gene	Gs03 gene
Model	TN93+G	T92+G+I	K2+G
Bayesian Information Criterion (BIC) scores	11277.2	13401.4	6932.3
Akaike Information Criterion, corrected (AICc) value	11125.9	13213.9	6729.27
Maximum Likelihood value (lnL)	-5541.9	-6581.94	-3335.5
Gamma distribution (G)	0.3919	1.2105	0.2556
invariable (I)	n/a	0.48	n/a
Transition/Transversion bias (R)	1.748	2.27388	2.1084
Total positions in the final dataset	1107	1118	545

Table 4- Pairwise distance *gs01* gene

		1	2	3	4	5	6	7	8	9
1	*Monopterus cuchia*	-								
2	*Monopterus albus*	0.133	-							
3	*Oryzias latipes*	0.26	0.362	-						
4	*Haplochromis burtoni*	0.261	0.362	0.155	-					
5	*Oreochromis niloticus*	0.256	0.356	0.155	0.022	-				
6	*Gasterosteus aculeatus*	0.278	0.368	0.154	0.146	0.143	-			
7	*Tetraodon nigroviridis*	0.266	0.369	0.186	0.184	0.182	0.164	-		
8	*Heteropneustes fossilis*	0.296	0.366	0.254	0.241	0.243	0.233	0.244	-	
9	*Lepisosteus oculatus*	0.302	0.361	0.233	0.216	0.219	0.214	0.209	0.259	-

The evolutionary history was inferred using the Maximum Parsimony method. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [34]. The analysis involved 9 numbers of nucleotide sequences [Fig-6](A).

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura and Nei model [35]. The final tree shown had the highest log likelihood (-5542.0024). The discrete Gamma distribution was applied to model evolutionary rate differences among sites [5 categories (+*G*, parameter = 0.4419)]. The analysis involved 9 nucleotide sequences [Fig-6](B).

gs02 Gene Phylogeny

The Pairwise distance of *gs02* gene sequences among the different fish species of the present study revealed shortest genetic distance (0.005) *Haplochromis burtoni* and *Pundamilia nyererei.* The longest genetic distance (0.280) exists between *Lepisosteus oculatus* and *Misgurnus anguillicaudatus.* The gs02gene of *Monopterus cuchia* and *Monopterus albus* showed a genetic distance of 0.177, which is the shortest genetic distance for *M. cuchia* among the gs01 sequences of the twelve species. *Monopterus cuchia* showed longest genetic distance (0.255) with *Lepisosteus oculatus* and *Misgurnus anguillicaudatus*, while *Monopterus albus* showed longest genetic distance (0.278) with *Opsanus beta* [Table-5].

The evolutionary history was inferred using the Maximum Parsimony method. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [34] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The analysis involved 12 nucleotide sequences [Fig-7](A).

The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura 3-parameter model [36]. The highest log likelihood in the final tree was -6576.6604. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+*G*, parameter = 0.2657)]. The branch lengths in the tree was measured in the number of substitutions per site for 12 number of nucleotide sequences [Fig-7](B).

gs03 Gene Phylogeny

The Pairwise distance of *gs03* gene sequences among the different fish species of the present study revealed shortest genetic distance (0.002) *Haplochromis burtoni* and *Pundamilia nyererei.* The longest genetic distance (0.313) exists between *Lepisosteus oculatus* and *Misgurnus nguillicaudatus.* The *gs03* gene of *Monopterus cuchia* and *Monopterus albus* showed a genetic distance of 0.081, which is the shortest genetic distance for *M. cuchia* among the gs01 sequences of the twelve species. Both *Monopterus cuchia* and *Monopterus albus* showed longest genetic distance with *Lepisosteus oculatus* (0.271 and 0.269 respectively) [Table-6].

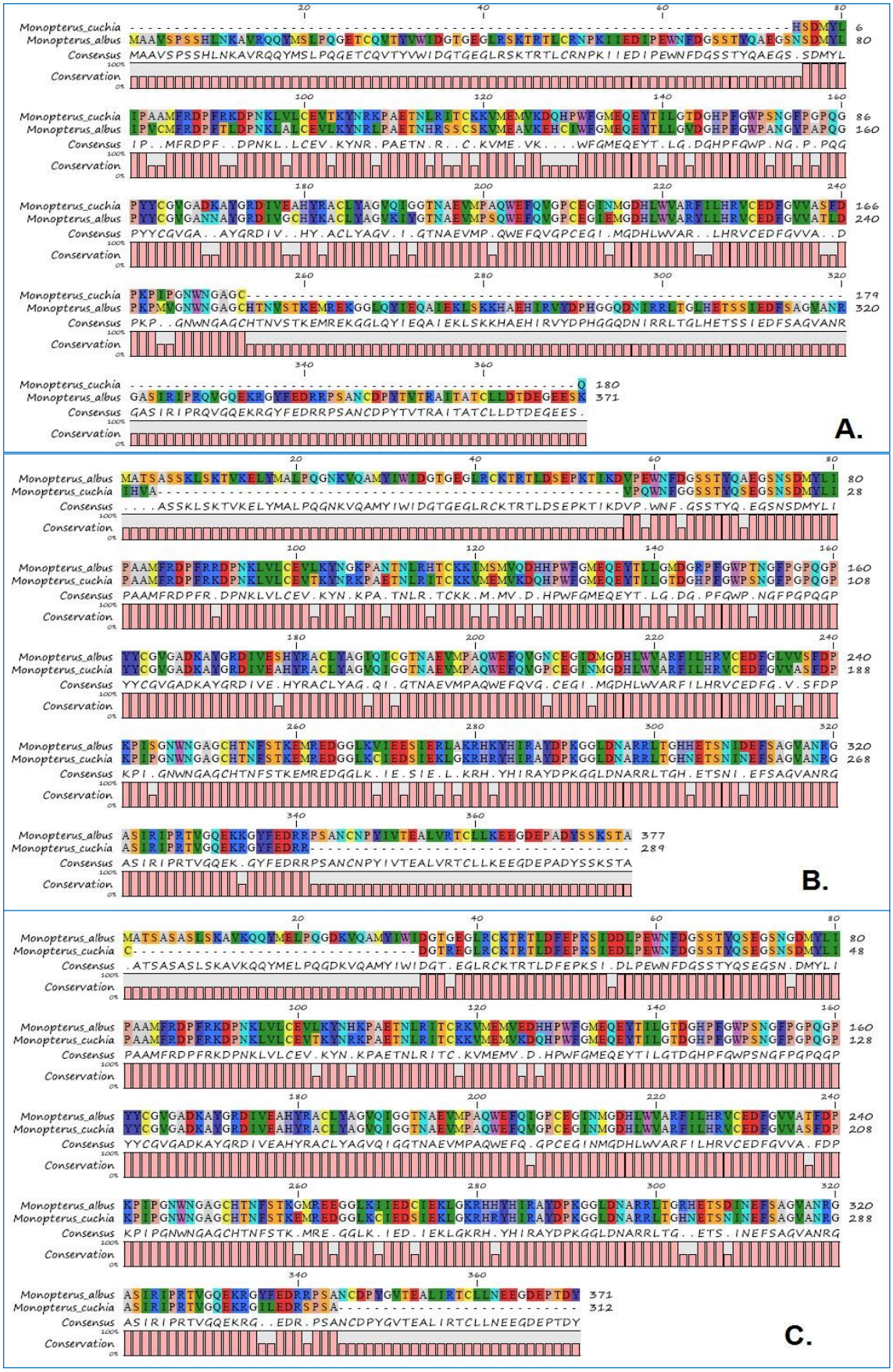

Fig. 5- Multiple amino acid sequence alignment of Glutamine Synthetase protein in *M. cuchia* and *M. albus.* '-' represent sequence not conserved. The sizes of the letter in the sequence logo represent the degree of conservation of respective amino acid in each alignment position. **(A)** GS-I, **(B)** GS-II, **(C)** GS-III.

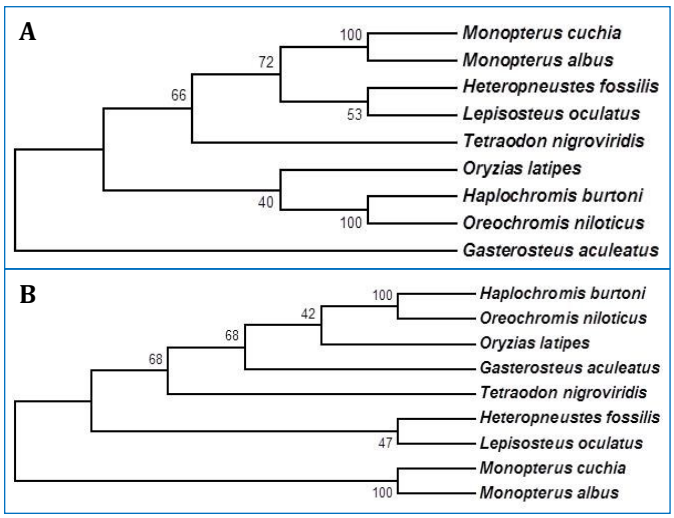

Fig. 6- Molecular phylogenetic anaylsis of *gs01*gene. **(A)** Maximum Parsimony tree, **(B)** Maximum Likelihood tree based on the Tamura -Nei model [35]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches [33]. The scale bars represent the branch lengths measured in the number of changes (substitutions per site) over the whole sequence.

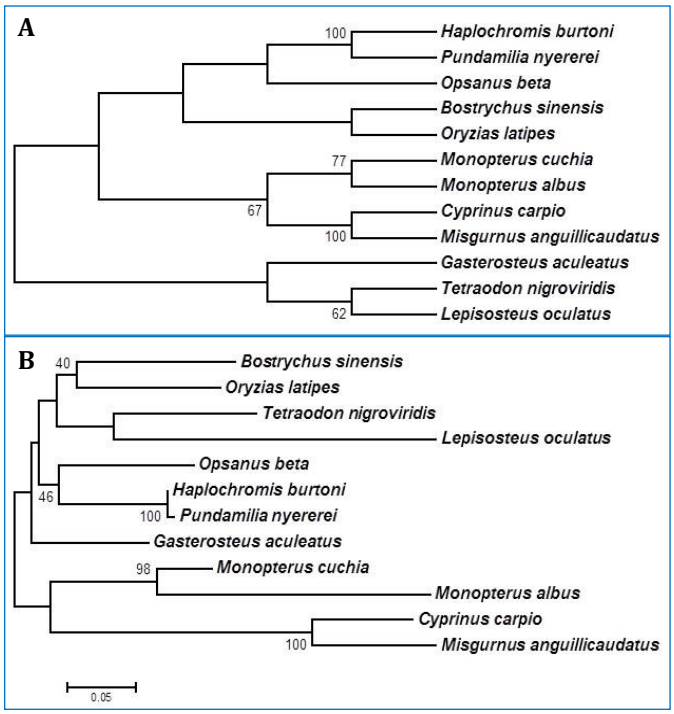

Fig. 7- Molecular phylogenetic anaylsis of bony fish *gs02*gene. **(A)** Maximum Parsimony tree, **(B)** Maximum Likelihood tree based on the Tamura 3-parameter model [36]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches [33]. The scale bars represent the branch lengths measured in the number of changes (substitutions per site) over the whole sequence.

The evolutionary history was inferred using the Maximum Parsimony method. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [34]. The tree is drawn to scale, with branch lengths calculated using the average pathway method [34] and are in the units of the number of changes over the whole sequence. The analysis involved 15 nucleotide sequences [Fig-8](A).

The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model [37]. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+*G*, parameter = 0.3493)]. 15 nucleotide sequences were considered for the analysis [Fig-8](B).

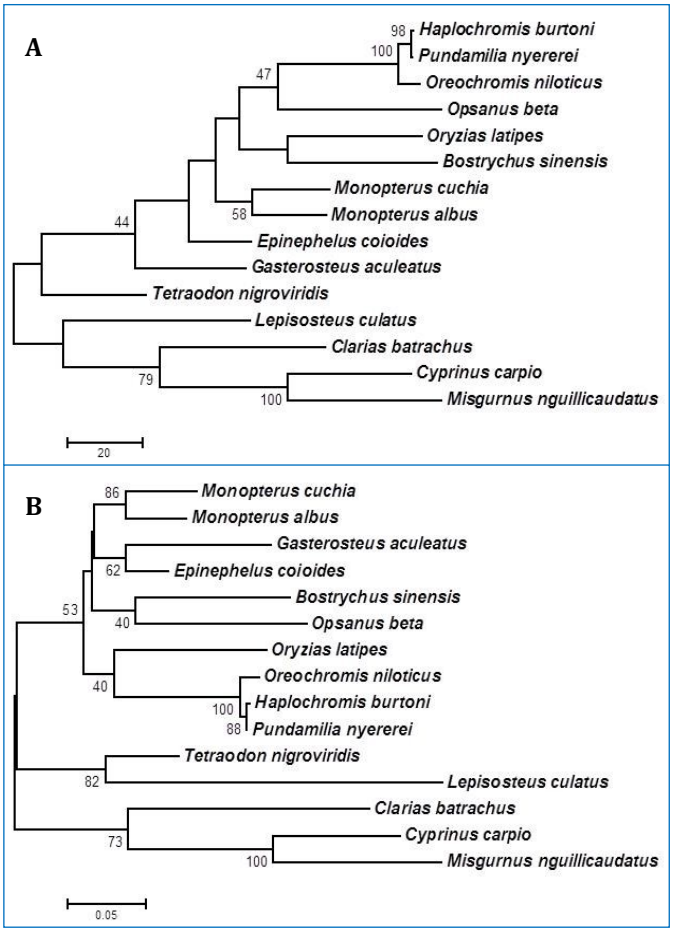

Fig. 8- Molecular phylogenetic anaylsis of ornamental fish *gs03* gene. **(A)** Maximum Parsimony tree, **(B)** Maximum Likelihood tree based on the Kimura 2-parameter model [37]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches [33]. The scale bars represent the branch lengths measured in the number of changes (substitutions per site) over the whole sequence.

Protein Tertiary Structures

The tertiary structure of GS-I for *M. cuchia* has 6 helices, 1 sheet, 3 gamma turns, 23 beta turns, 2 helix-helix interacs, 1 beta bulge, 3 beta hairpins and 5 strands on the other hand GS-I for *M. albus* has 11 helix-helix interacs, 11 helices, 2 sheets, 5 gamma turns, 39 beta turns, 5 beta hairpins, 5 beta bulges and 13 strands [Fig-9](A-B); [Table-7].

The computational model of GS-II for *M. cuchia* has 9 helices, 6 helix-helix interacs, 27beta turns, 6 gamma turns, 3sheets, 4 beta hairpins,3 beta bulge and 10 strands on the other hand GS-II for *M. albus* has 11 helices, 10 helix-helix interacs, 36 beta turns,7 gamma turns, 3 sheets, 5 beta hairpins, 4 beta bulges and 14 strands [Fig-9](C-D); [Table-7].

Table 5- Pairwise distance *gs02* gene

		1	2	3	4	5	6	7	8	9	10	11	12
1	*Monopterus cuchia*	-											
2	*Monopterus albus*	0.177	-										
3	*Haplochromis burtoni*	0.177	0.262	-									
4	*Pundamilia nyererei*	0.179	0.267	0.005	-								
5	*Bostrychus sinensis*	0.203	0.272	0.162	0.163	-							
6	*Gasterosteus aculeatus*	0.19	0.262	0.147	0.15	0.181	-						
7	*Oryzias latipes*	0.2	0.266	0.157	0.155	0.172	0.158	-					
8	*Opsanus beta*	0.195	0.278	0.149	0.15	0.18	0.147	0.186	-				
9	*Tetraodon nigroviridis*	0.206	0.251	0.186	0.186	0.189	0.166	0.19	0.196	-			
10	*Cyprinus carpio*	0.254	0.273	0.236	0.233	0.221	0.245	0.273	0.262	0.237	-		
11	*Lepisosteus oculatus*	0.255	0.251	0.219	0.221	0.227	0.217	0.239	0.264	0.212	0.261	-	
12	*Misgurnus anguillicaudatus*	0.255	0.269	0.251	0.249	0.238	0.255	0.269	0.259	0.252	0.135	0.28	-

Table 6- Pairwise distance *gs03* gene

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*Monopterus cuchia*	-														
2	*Monopterus albus*	0.081	-													
3	*Oreochromis niloticus*	0.15	0.16	-												
4	*Haplochromis burtoni*	0.144	0.15	0.019	-											
5	*Pundamilia nyererei*	0.142	0.152	0.017	0.002	-										
6	*Gasterosteus aculeatus*	0.155	0.149	0.19	0.198	0.195	-									
7	*Oryzias latipes*	0.145	0.158	0.165	0.158	0.156	0.176	-								
8	*Bostrychus sinensis*	0.152	0.139	0.201	0.179	0.181	0.207	0.172	-							
9	*Tetraodon nigroviridis*	0.174	0.183	0.2	0.196	0.194	0.176	0.191	0.205	-						
10	*Opsanus beta*	0.173	0.173	0.179	0.176	0.178	0.193	0.204	0.178	0.222	-					
11	*Clarias batrachus*	0.218	0.234	0.238	0.247	0.244	0.206	0.248	0.266	0.219	0.286	-				
12	*Cyprinus carpio*	0.247	0.244	0.236	0.241	0.244	0.261	0.301	0.23	0.253	0.291	0.248	-			
13	*Epinephelus coioides*	0.108	0.096	0.149	0.138	0.141	0.111	0.137	0.151	0.161	0.158	0.225	0.22	-		
14	*Misgurnus nguillicaudatus*	0.258	0.26	0.264	0.261	0.264	0.276	0.276	0.243	0.271	0.291	0.263	0.158	0.24	-	
15	*Lepisosteus oculatus*	0.271	0.269	0.251	0.25	0.247	0.264	0.263	0.269	0.201	0.317	0.288	0.297	0.255	0.313	-

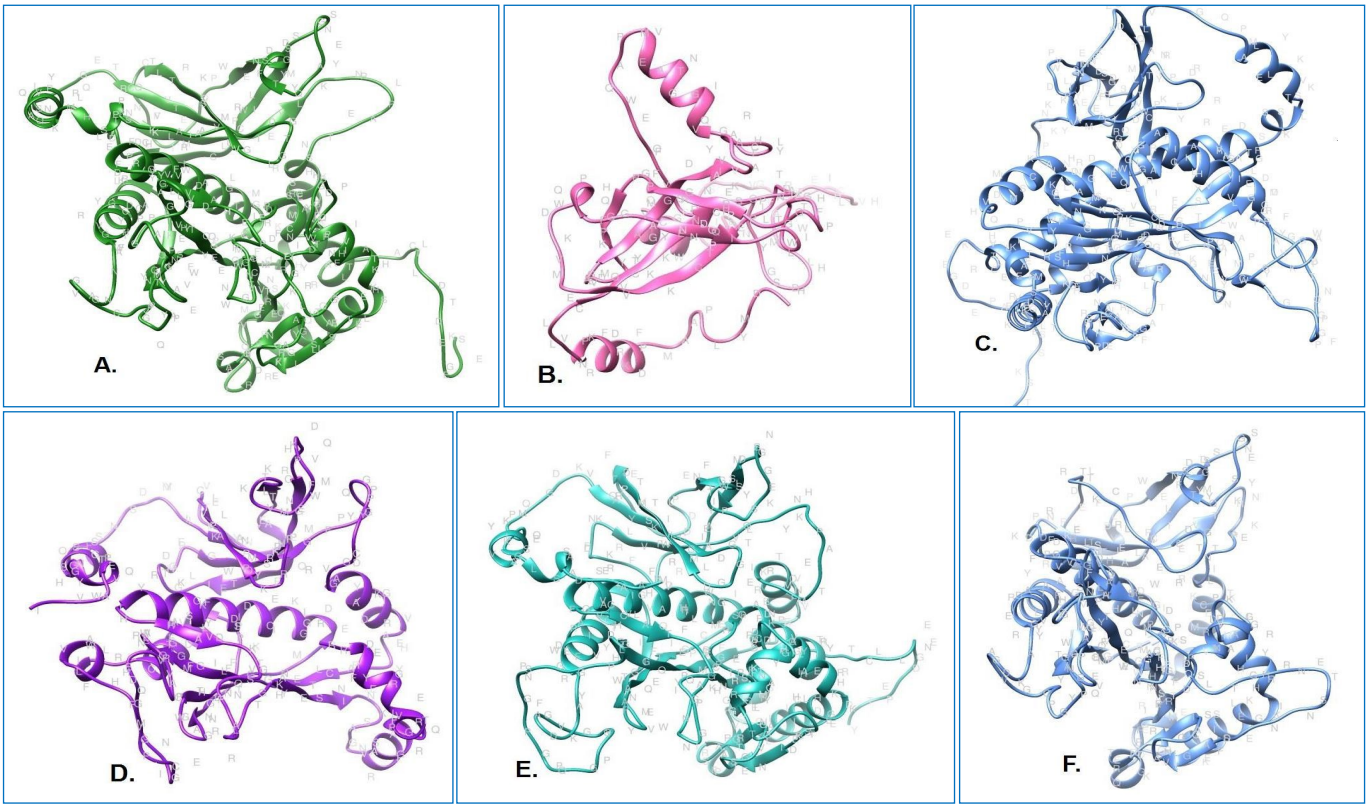

Fig. 9- The predicted homology model of GS structure, as displayed by UCSF Chimera. **(A)** GS-I of *M. cuchia*, **(B)** GS-I of *M. albus*, **(C)** GS-II of *M. cuchia*, **(D)** GS-II of *M. albus*, **(E)** GS-III of *M. cuchia*, **(F)** GS-III of *M. albus*.

The theoretical structure of GS-III of *M. cuchia* has 2 sheets, 10 helices, 32beta turns, 5 gamma turns, 5 helix-helix interacs, 4 beta hairpins,3 beta bulges and 11 strands on the other hand GS-III for *M. albus* has 11 helices, 10 helix-helix interacs, 41 beta turns,6 gamma turns, 3 sheets, 5 beta hairpins, 3 beta bulges and 14 strands [Fig-9](E-F); [Table-7]. The verification performed by ER-RAT had revealed that the overall quality factor for the predicted tertiary structures of GS-I, GS-II and GS-III is around 95% [Fig-10]. ProCheck verification revealed that more that 90% of the amino acid residues in the predicted 3D structures of GS are in the range of most favoured region, which confirms the validity and usefulness of the 3D structures [Fig-11]. The functional annotation results are listed in the [Table-8].

Table 7- Structural characteristics of predicted tertiary structures of GS

Protein name	Taxon	No. of helices	No. of helix-helix interacts	No. of sheets	No. of beta hairpins	No. of beta turns	No. of gamma turns
GS-I	*Monopterus cuchia*	5	2	1	3	23	3
	Monopterus albus	11	11	2	5	39	5
GS-II	*Monopterus cuchia*	9	6	3	4	27	6
	Monopterus albus	11	10	3	5	36	7
GS-III	*Monopterus cuchia*	10	5	2	4	32	5
	Monopterus albus	11	10	3	5	41	6

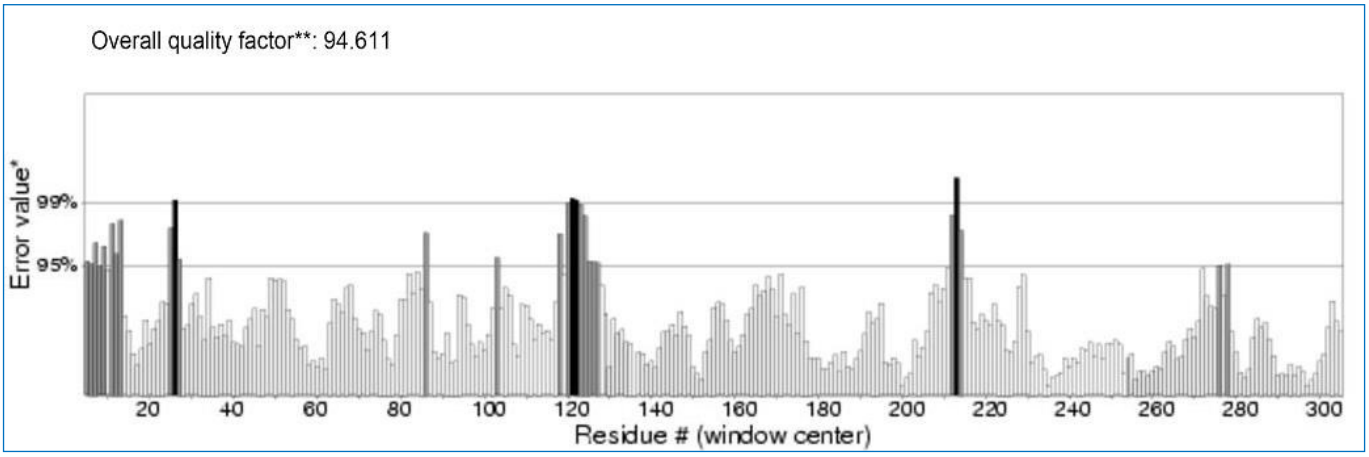

Fig. 10- Structure validation results showing Overall quality of 3D structure of GS (ERRAT2 Verification).

Table 8- Summary of predicted function with ProFunc score (shown within parenthesis)

Taxon	Protein name terms	Gene Ontology (GO) terms		
		Cellular component	Biological process	Biochemical function
GS-I				
M. cuchia	glutamine synthetase, (23.71); glutamine synthetase fragment (2.20); sv=1 (2.07)	intracellular (23.29);cytoplasm (23.29); cell part (23.29)	primary metabolic process (36.87); cellular process (36.04);cellular metabolic process (35.19)	catalytic activity (34.44); nucleotide binding (24.14);ATP binding (23.13)
M. albus	glutamine synthetase (54.48); glutamine synthetase fragment (7.41); phosphate (6.21)	intracellular (28.19);cytoplasm (28.19);cell (28.19);cell part (28.19)	cellular metabolic process (53.33);primary metabolic process (53.33);metabolic process (52.76)	nucleotide binding (53.59); ATP binding (53.59); purine nucleotide binding (53.59)
GS-II				
M. cuchia	glutamine synthetase (32.01); phosphate (5.21); human (3.29); pe=2 (2.84); sulfoximine (2.70); sulfoximine phosphate (2.36); sv=1 (2.21)	cell (30.23); cell part (30.23); intracellular (27.39); intracellular part (27.39)	cellular process (43.92); primary metabolic process (42.13); cellular metabolic process (42.12)	catalytic activity (44.18); nucleotide binding (36.70); purine nucleotide binding (35.84)
M. albus	glutamine synthetase (31.57); phosphate (5.43); fragment (3.24); mycobacterium (2.85); pe=2 (2.83); sulfoximine (2.70); human (2.53)	cell (28.50); cell part (28.50); intracellular (26.55); intracellular part (26.55)	primary metabolic process (38.15); cellular process (37.31); cellular metabolic process (37.31)	catalytic activity (38.76); nucleotide binding (37.62); purine nucleotide binding (36.62)
GS-III				
M. cuchia	glutamine (45.27); glutamine synthetase (41.72); phosphate (7.61); sulfoximine (3.31); salmonella typhimurium (3.15); sulfoximine phosphate (2.97); pe=2 (2.84)	intracellular (29.38); cytoplasm (29.38); cell (29.38); cell part (29.38)	cellular process (55.32); cellular metabolic process (55.32); cellular biosynthetic process (53.24)	catalytic activity (56.98); ligase activity (54.93); ligase activity, forming carbon-nitrogen bonds (54.93); binding (54.58)
M. albus	glutamine synthetase (37.46); phosphate (7.14); adp (3.16); sulfoximine (3.10); human (2.87); sulfoximine phosphate (2.76)	cell (30.21); cell part (30.21); intracellular (29.25); intracellular part (28.44)	cellular process (47.22); cellular metabolic process (46.36); cellular biosynthetic process (43.82)	catalytic activity (50.12); nucleotide binding (45.69); purine nucleotide binding (43.95)

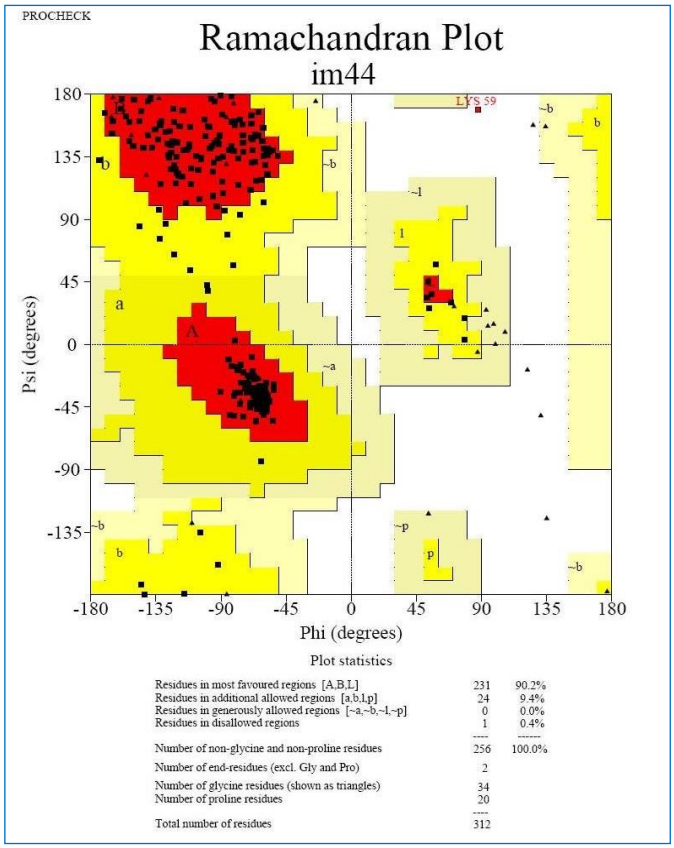

Fig. 11- Ramachandran analysis of the backbone dihedral angles PSI (y) and PHI (f) for the final structure of GS (from *Monopterus spp.*). Red region represents the most favored region, yellow = allowed region, light yellow = generously allowed region, white = disallowed region [ProCheck].

Discussion

Efficient identification of the two eel species of the present study is critical for aquaculture management as well as for eel conservation [39]. Thus, identification of *M. cuchia* and *M. albus* needs to be supported by molecular characterization instead of conventional methods [40]. The present study had revealed an interesting point of identification that the sequences of *M. albus* were found to be A:T rich for all the three gs gene sequences. On the other hand, the gs genes sequences of *M. cuchia* were rich in G:C frequency than A:T frequency. The GS of both *M. cuchia* and *M. albus* are rich in Glycine.

gs genes are differentially expressed in different tissues and therefore possibly involved in different metabolic pathways. GS activity is typically high in the brain [41], although liver can also be an important site of ammonia detoxification [42]. Sequencing and analysis of *gs* genes from all possible types of cells of its occurrence will revail crucial information leading to detailed understanding of detoxification mechanism in freshwater air-breathing fishes.

The presence of two GS genes in zebrafish and fugu and four in trout suggests that gene duplication events of GS have occurred within bony fishes in multiple copies [3]. The Instability index in GS-I (25.19) and GS-II (39.63) of *M. cuchia* is in the range of stable molecule where Instability index value in all the three GS in *M. albus* along with GS-III of *M. cuchia* showed that GS-I, GS-II, GS-III in *M. albus* and GS-III in *M. cuchia* are unstable proteins.

The hydropathicity plot [Fig-3] revealed that the GS protein is hydrophilic in nature. The phylogenetic analysis of *gs01* gene revealed the *M. cuchia* and *M. albus* are sister taxa followed by *Heteropnestes fossilis* and *Lepisosteus oculatus* as their successive sister taxa. ProFunc analysis has revealed that GS have several functional properties which include primery cellular metabolic process, purine nucleotide binding, ATP binding, catalytic activity, synthetase activity etc. [Table-8].

The GS amino acid sequences of *Monopterus* along with other fish species and those of amphibians and mammals are highly conserved [43]. The extraordinary capacity of *M. albus and M. cuchia* to increase glutamine synthesis and accumulation for cell volume regulation is probably a consequence of the lack of functional gills [44], which could have developed as an extension of its ability to increase glutamine synthesis to detoxify ammonia during emersion [45], exposure to ammonia in environmental condition or aestivation in mud [46,47]. The findings of a previous study demonstrate that rainbow trout have a considerable reserve capacity to prevent brain ammonia toxicity by inhibition of glutamine synthetase [48]. A study on *Bostrichthys sinensis* revealed that exposure to ammonia results in significant increases in GSase activity, GSase protein and GSase mRNA levels in all tissues [49].

The present study clearly emphasizes the need to fully identify all the possible isoformes of *gs* genes coding for the Glutamate Synthetase enzyme prior to the interpretation of data showing changes in the levels of mRNA expression and suggests a complex interaction study of the gene products of duplicated loci in multimeric Glutamate Synthetase isoforms.

Acknowledgment: The authors gratefully acknowledge the Department of Biotechnology, Govt. of India for the financial assistance to establish the Bioinformatics Infrastructure Facility (BIF) at the Zoology Department, Gauhati University, which has been utilized in the present study.

Conflicts of Interest: None declared.

References

[1] Walsh P.J., Meyer G.D., Medina M., Bernstein M.L., Barimo J.F. & Mommsen T.P. (2003) *The Journal of Experimental Biology,* 206, 1523-1533.

[2] Mommsen T.P., Busby E.R., von Schalburg K.R., Evans J.C., Booth H.L. & Elliot M.E. (2003) *Journal of Comparative Physiology,* 173B, 419-427.

[3] Murray B.W., Busby E.R., Mommsen T.P. & Wright P.A. (2003) *The Journal of Experimental Biology,* 206, 1511-1521.

[4] Korsgaard B., Mommsen T.P. & Wright P.A. (1995) *Nitrogen Metabolism and Excretion*, CRC Press, Boca Raton, FL, 259-287.

[5] Ip Y.K., Chew S.F. & Randall D.J. (2001) *Fish Physiology*, 20, Nitrogen Excretion, Academic Press, San Diego, 109-148.

[6] Gharbi K., Murray B.W., Moghadam H.K., Ferguson M.M., Wright P.A. & Danzmann R.G. (2007) *Cytogenetic and Genome Research*, 116(1-2), 113-115.

[7] Zemla A., Venclovas Č., Moult J. & Fidelis K. (1999) *Proteins: Structure, Function, and Bioinformatics*, 37(S3), 22-29.

[8] Boeckmann B., Bairoch A., Apweiler R., Blatter M.C., Estreicher A., Gasteiger E., Martin M.J., Michoud K,. O'Donovan C., Phan

I., Pilbout S. & Schneider M. (2003) *Nucleic Acids Research*, 31, 365-370.

[9] Apweiler R., Bairoch A., Wu C.H., Barker W.C., Boeckmann B., Ferro S., Gasteiger E., Huang H., Lopez R., Magrane M., Martin M.J., Natale D.A., O'Donovan C., Redaschi N. & Yeh L.S. (2004) *Nucleic Acids Research*, 32, D115-119.

[10] Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D. & Bairoch A. (2005) *The Proteomics Protocols Handbook,* Humana Press, 571-607.

[11] Thompson J.D., Higgins D.G. & Gibson T.J. (1994) *Nucleic Acids Research*, 22, 4673-4680.

[12] Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S. (2013) *Molecular Biology and Evolution,* 30, 2725-2729.

[13] Eck R.V. & Dayhoff M.O. (1966) *Atlas of protein sequence and structure*, National Biomedical Research Foundation, Silver Spring, Maryland, USA.

[14] Jones D.T., Taylor W.R. & Thornton J.M. (1992) *Computer Application in Biosciences*, 8, 275-282.

[15] Posada D. & Crandall K.A. (1998) *Bioinformatics*, 14, 817-818.

[16] Altschul S.F. (1991) *Journal of Molecular Biology*, 219(3), 555-565.

[17] Pearson W.R. (1991) *Genomics*, 11(3), 635-650.

[18] Berman H., Henrick K., Nakamura H. & Markley J.L. (2007) *Nucleic Acids Research,* 35, D301-D303.

[19] Lassmann T. & Sonnhammer E.L. (2005) *Nucleic Acids Research*, 33(22), 7120-7128.

[20] Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G. & Gibson T.J. (1994) *Nucleic Acids Research*, 22, 4673-4680.

[21] Fiser A., Do R.K. & Sali A. (2000) *Protein Science*, 9, 1753-1773.

[22] Martí-Renom M.A., Yerkovich B. & Sali A. (2002) *Protein Science*, 1, 2.9.1-2.9.22.

[23] Fiser A. & Sali A. (2003) *Bioinformatics*, 18, 2500-2501.

[24] Eswar N., Webb B., Marti-Renom M.A., Madhusudhan M.S., Eramian D., Shen M.Y., Pieper U. & Sali A. (2006) *Current Protocols in Protein Science*, 2.9.1-2.9.31.

[25] Sali A. & Blundell T.L. (1993) *Journal of Molecular Biology*, 234, 779-815.

[26] Giorgetti A., Raimondo D., Miele A.E. & Tramontano A. (2005) *Bioinformatics,* 21(2), 72-76.

[27] Colovos C. & Yeates T.O. (1993) *Protein Science*, 2, 1511-1519.

[28] Laskowski R.A., Watson J.D. & Thornton J.M. (2005) *Nucleic Acids Research*, 33, W89-W93.

[29] Tiziana C., Paolo D'Onorio D.M., Domenico C., Talamo I.G. & Tramontano A. (2006) *Nucleic Acids Research*, 34, D306-D309.

[30] Sayle R.A. & Milner-White E.J. (1995) *Trends in Biochemical Sciences*, 20(9), 374-376.

[31] Laskowski R.A., Watson J.D. & Thornton J.M. (2005) *Journal of Molecular Biology*, 351, 614-626.

[32] Zdobnov E.M. & Apweiler R. (2001) *Bioinformatics*, 17, 847-848.

[33] Felsenstein J. (1985) *Evolution*, 39, 783-791.

[34] Nei M. & Kumar S. (2000) *Molecular evolution and phylogenetics*, Oxford University Press, Oxford.

[35] Tamura K. & Nei M. (1993) *Molecular Biology and Evolution*, 10, 512-526.

[36] Tamura K. (1992) *Molecular Biology and Evolution*, 9(4), 678-687.

[37] Kimura M. (1980) *Journal of Molecular Evolution*, 16, 111-120.

[38] Kyte J. & Doolittle R. (1982) *Journal of Molecular Biology*, 157, 105-132.

[39] Dudu A., Georgescu S.E. & Costache M. (2010) *Archiva Zootechnica*, 13(1), 53-59.

[40] Huang J.P., Han Y.S. & Tzeng W.N. (2001) *Acta Zoologica Taiwanica*, 12(2), 41-49.

[41] Wang Y. & Walsh P.J. (2000) *Aquatic Toxicology,* 50, 205-219.

[42] Iwata K., Kajimura M. & Sakamoto T. (2000) *Journal of Experimental Biology*, 203, 3703-3715.

[43] Pesole G., Bozzetti M.P., Lanave C., Preprata G. & Saccone C. (1991) *Proceedings of National Academy of Sciences USA*, 88, 522-526.

[44] Graham J.B. (1997) *Air-breathing Fishes: Evolution, Diversity and Adaptation*, Academic Press, London.

[45] Tay A.S.L., Chew S.F. & Ip Y.K. (2003) *Journal of Experimental Biology*, 206, 2473-2486.

[46] Chew S.F., Poothodiyil, N.K., Wong W.P. & Ip Y.K. (2006) *Journal of Experimental Biology*, 209, 484-492.

[47] Ip Y.K., Tay A.S.L., Lee K.H. & Chew S.F. (2004) *Physiological and Biochemical Zoology*, 77, 390-405.

[48] Sanderson L.A., Wright P.A., Robinson J.W., Ballantyne J.S. & Bernier N.J. (2010) *Journal of Experimental Biology*, 213, 2343-2353.

[49] Anderson M., Broderius M.A., Fong K.C., Tsui K.N.T., Chew S.F. & Ip Y.K. (2002) *Journal of Experimental Biology*, 205, 2053-2065.