Phylogeny and genetic divergence of three phenotypic variants of the ornamental goldfish, *Carassius auratus* (Linnaeus, 1758) based on CO1 gene

Ann Rose Bright, Naveen Sathyan, Swapna P. Antony, T. J. James, Rosamma Philip

1Department of Genomic Science, School of Biological Science, Central University of Kerala, Kerala, India
2Department of Marine Biology, Microbiology and Biochemistry, School of Marine Sciences, CUSAT, Kochi-16, Kerala, India
3Department of Zoology, Sacred Heart College, Thevara, Kerala, India

Received 17 March 2013; accepted 30 March 2013

Abstract

*Carassius auratus* is one of the most popular aquarium fishes having wide distribution in Eurasian continents. Artificial selection has led to the formation of a number of morphological variants of this fish. The present study is aimed at determining the genetic variation among the morphological variants of *C. auratus* based on cytochrome oxidase subunit-I (CO1), mitochondrial gene. The intraspecific and interspecific genetic divergence for *C. auratus* based on CO1 gene was also assessed. Analysis of the CO1 nucleotide sequences obtained from three morphological variants of goldfish confirmed that the specimens belonged to species *C. auratus* and no genetic divergence exist among the phenotypic variants of goldfish. The mean interspecific genetic distances (Kimura 2 parameter) calculated for *C. auratus* was found to be 36 fold higher than the intraspecific mean distance calculated for Cyprinidae fishes. The phylogenetic tree constructed (NJ method) based on CO1 gene showed *C. auratus* to align with the Cyprinidae clade. The mean interspecific distance calculated for Cyprinidae fishes indicated that with regard to genetic variations, family Cyprinidae is more stable in comparison to other Cypriniformes families. High efficiency in species identification and phylogenetic analysis was exhibited by CO1 gene in the present study highlighting the proficiency of mitochondrial genome based molecular markers in taxonomic and evolutionary studies.

Keywords: *Carassius auratus*; Ornamental Goldfish; DNA barcoding; Phylogenetic analysis; Genetic divergence

1. Introduction

Southeast Asia harbors rich biodiversity with respect to freshwater fishes. Many of these freshwater fish species have been exploited for ornamental fish trade [1]. There is immense scope for utilizing CO1 based barcoding in better regulation of International trade of ornamental fishes, which, if not regulated properly might pose risks to biodiversity via invasive alien species and exotic pathogens. High diversity of ornamental fresh water fishes and their dramatic phenotypic changes during development has made the morphology based taxonomic identification of species a difficult task. Expanding the current CO1 based barcode reference libraries and assessing barcode similarity with morphological identifications will provide a better option for precise identification of aquarium fishes [2].

Goldfish, *Carassius auratus* is primarily a freshwater fish regarded to be one of the most popular pet fishes of the world. It is widely distributed in the Eurasian continents and belongs to family Cyprinidae of order Cypriniformes. *C. auratus* is an exceedingly variable organism which includes a range of morphological varieties. Phenotypic traits in goldfish have evolved together with human culture owing to rigorous artificial breeding. Komiyama et al. [3] has shown that all goldfish varieties have originated from a single *Carassius auratus gibelio* ancestor commonly found in Chinese waters. Their work also revealed that the process of artificial selection in goldfish began with the losing of dorsal fin followed by diversification of other morphological characters like eyes. The present study was conducted to test the efficiency of CO1 gene in determining the species identity, phylogeny and genetic divergence in goldfish, *C. auratus*. Three morphological variants of *C. auratus* viz., Common goldfish, Fantail goldfish and Comet goldfish were examined for studying genetic variation with respect to CO1 gene. An in-depth analysis of intraspecific and interspecific genetic divergence has been carried out in the present study.
2. Materials and Methods
2.1. Sample collection
Live specimens of *C. auratus* were purchased from an ornamental fish aquarium in Kochi (Kerala, India). Three phenotypic varieties of *C. auratus* were considered for the study. These include Fantail goldfish (Voucher No. S-5A), Common goldfish (Voucher No. S-5C) and Comet goldfish (Voucher No. S-5D) (Fig. 1). In each case muscle tissue was considered for DNA isolation.

![Fig. 1: Phenotypic variants of Carassius auratus considered for the study. a) Fantail goldfish (S-5A); b) Common goldfish, (S-5C); c) Comet goldfish-Lateral view (S-5D); d) Comet goldfish-Dorsal view (S-5D).](image)

2.2. DNA extraction, PCR amplification and sequencing
Total genomic DNA was isolated from muscle tissue (10 mg) taken from fresh specimens. Tissue was digested by incubating with proteinase K/SDS solution at 37°C for two hours. DNA isolation was carried out following standard phenol: chloroform extraction and ethanol precipitation technique [4]. Purity and quality of DNA was checked on 0.8% agarose gel. The concentration of dissolved DNA was estimated using a UV spectrophotometer (Hitachi U-2900).

DNA was diluted so as to obtain a final concentration of 100 ng/µl. PCR amplification was carried out to obtain sequences of the partial mitochondrial COI gene, in a total volume of 25 µl containing 1x standard Taq buffer (10 mM Tris-HCl, 50 mM KCl, pH 8.3), 3.5 mM MgCl₂, 200 µM dNTPs, 0.4 µM each primer, 1U Taq DNA polymerase (Fermentas, Inc.) and 1µl DNA template (100 ng). The primer pairs used were COI- F (5'-taaccaacacacaaga cattgccac-3') and COI- R (5'-tagacttctgggctgeaag aatca-3'). The thermal profile used was 94°C for 2 min followed by 35 cycles of 94°C for 15 sec, 55°C for 30 sec and 68°C for 30 sec and a final extension at 68°C for 10 min. 10 µl of the amplified PCR product was analyzed by electrophoresis in 1.5% agarose gel in TBE buffer, stained with ethidium bromide and visualized under UV light.

Purified PCR products were sequenced with CO1 primers using ABI Prism Sequencing kit (BigDye Terminator Cycle) at SciGenom Sequencing Facility, India.

2.3. Sequence data analysis
The homologue searching of the nucleotide sequence (using blastn suite) and deciphered amino acid sequence (using blastp suite) were performed with the Basic Local Alignment Search Tool (BLAST) through NCBI server (http://www.ncbi.nlm.nih.gov/blast). Nucleotide sequence was translated to protein with standard vertebrate mitochondrial code (codon 2) using DNA to Protein translation tool (http://insilico.ehu.es_TRANSLATE/).

Nucleotide sequences of COI genes of related fishes were retrieved from the NCBI Genbank. The sequences were imported to BioEdit v.7.0.9. [5] and aligned using CLUSTALW [6]. Phylogenetic tree was constructed by the Neighbour-Joining (NJ) method and Maximum Likelihood (ML) method based on nucleotide sequence of cytochrome oxidase subunit-1, using MEGA version 5.05 [7] and bootstrap analysis was carried out using 100 and 1,000 replicates with MEGA version 5.05. Kimura 2-parameter (K2P) model [8] was used to construct NJ and ML tree based on nucleotide sequence. Pairwise intraspecific and interspecific nucleotide sequence divergences under the K2P model were calculated using MEGA 5.05. Further statistical analysis of the nucleotide sequences of *C. auratus* was carried out in DnaSP v.5 software [9].

3. Results
The morphological variants of ornamental goldfish, S-5A, S-5C and S-5D (Fig.1) were subjected to genotypic identification based on CO1 gene sequence. Partial COI readable sequence of 530bp, 618bp and 636bp were obtained for S-5A, S-5C and S-5D amplicons respectively. BLAST analysis of the nucleotide sequences of S-5A, S-5C and S-5D showed 100% similarity to that of *Carassius auratus* for query coverage of 100%. The result confirmed the identity of S-5A, S-5C and S-5D as *C. auratus*. The nucleotide sequences obtained was found to represent the expected 5' region of the mitochondrial CO1 gene following BLAST analysis with previously published mitochondrial CO1 sequence data. BLASTp analysis of the deduced amino acid sequences of CO1 gene confirmed the presence of CO1 super family domain, possessing hemebinding sites in the translated sequence of S-5A (Frame-1), S-5C (Frame-1) and S-5D (Frame-1). The obtained sequences were deposited in GenBank database (GenBank ID: JF752338 for S-5A, JF752339 for S-5C and JF752340 for S-5D).

---

![Image](image)
All the three examined nucleotide sequences shared 100% similarity with each other indicating complete absence of genetic divergence among the three morphological variants (S-5A, S-5C and S-5D) of goldfish with respect to CO1 gene. Since all three sequences were similar to each other, only one nucleotide sequence (S-5D) was considered for further analysis. The nucleotide and deduced amino acid sequences is shown in Fig. 2. Nucleotide frequency of CO1 gene for S-5D was found to be A: 26.26%, C: 29.09%, G: 16.67% and T: 27.99% (Fig. 3). The A+T content in case of S-5D was found to be on a higher side (54.25%). This is in agreement with the nucleotide sequences of all CO1 genes analyzed in the present study.

The NJ tree representing the phylogenetic relationship of C. auratus (S-5D) to 48 closely related fishes belonging to order Cypriniformes, constructed based on the CO1 gene nucleotide sequences is shown in Fig. 3. CO1 gene nucleotide sequences of five elasmobranchs were included as an out-group. The CO1 nucleotide sequence based NJ tree branched into five clades, excluding the clade formed by the elasmobranch out-group. Each family of the order Cypriniformes viz. Cyprinidae, Catostomidae, Gyrinocheilidae, Balitoridae and Cobitidae were distributed into separate clade. S-5D was found to be closely related to related CO1 nucleotide sequences of all Cypriniformes fishes. Though orders Cypriniformes include six families, the GenBank database contained CO1 nucleotide sequences of representatives from only five families viz. Cyprinidae, Catostomidae, Gyrinocheilidae, Balitoridae and Cobitidae. CO1 nucleotide sequences for fishes belonging to family Psilorhynchidae was not represented in GenBank database.

Based on the similarity in BLAST results, nucleotide sequences of CO1 gene of closely related fishes belonging to order Cypriniformes were downloaded from GenBank database. Only a single representative from each species was considered for phylogenetic analysis. Though orders Cypriniformes include six families, the GenBank database contained CO1 nucleotide sequences of representatives from only five families viz. Cyprinidae, Catostomidae, Gyrinocheilidae, Balitoridae and Cobitidae. CO1 nucleotide sequences for fishes belonging to family Psilorhynchidae was not represented in GenBank database. Even in case of Gyrinocheilidae, CO1 sequence of only a single representative (Gyrinocheilus aymonieri) was found in GenBank database. Phylogenetic relationship of S-5D to other Cypriniformes was established based on the nucleotide sequence comparisons of CO1 gene. Phylogenetic relationship of S-5D was found to be virtually identical in NJ tree and ML tree.

The NJ tree representing the phylogenetic relationship of C. auratus (S-5D) to 48 closely related fishes belonging to order Cypriniformes, constructed based on the CO1 gene nucleotide sequences is shown in Fig. 3. CO1 gene nucleotide sequences of five elasmobranchs were included as an out-group. The CO1 nucleotide sequence based NJ tree branched into five clades, excluding the clade formed by the elasmobranch out-group. Each family of the order Cypriniformes viz. Cyprinidae, Catostomidae, Gyrinocheilidae, Balitoridae and Cobitidae were distributed into separate clade. S-5D was found to be closely related to related CO1 nucleotide sequences of all Cypriniformes fishes. Though orders Cypriniformes include six families, the GenBank database contained CO1 nucleotide sequences of representatives from only five families viz. Cyprinidae, Catostomidae, Gyrinocheilidae, Balitoridae and Cobitidae. CO1 nucleotide sequences for fishes belonging to family Psilorhynchidae was not represented in GenBank database. Even in case of Gyrinocheilidae, CO1 sequence of only a single representative (Gyrinocheilus aymonieri) was found in GenBank database. Phylogenetic relationship of S-5D to other Cypriniformes was established based on the nucleotide sequence comparisons of CO1 gene. Phylogenetic relationship of S-5D was found to be virtually identical in NJ tree and ML tree.

In BLAST analysis 45 related CO1 gene sequences of C. auratus (excluding the complete/partial sequences of whole mitochondrial genome) were found in GenBank database. These sequences were downloaded from the database. The overall mean K2P intraspecific distance was calculated using S-5D and 45 downloaded sequences. Overall mean K2P distance was found to be 0.0038. Overall mean interspecific K2P distance was calculated using all 48 sequences (including S-5D and excluding elasmobranch out-group) considered for construction of NJ tree was calculated to be 0.1898. K2P mean distance of Cyprinidae clade to that of Gyrinocheilidae, Catostomidae, Bothidae and Cobitidae was estimated to be 0.1707, 0.1806, 0.2053 and 0.2267 respectively. Gyrinocheilidae was found to be closest to Cyprinidae. K2P within group mean distance was calculated for each clade and it was found to be 0.1237 for Cyprinidae, 0.1252 for Catostomidae, 0.1543 for Bothidae and 0.2287 for Cobitidae. No result was obtained for Gyrinocheilidae as the clade is represented by a single fish species. K2P pairwise distance was calculated and the distance of S-5D to other fishes of Cyprinidae clade is presented in Table. 1. C. gibelio was found to be the most closely related species to C. auratus and Barbonymus was found to be the most closely related genera to Carassius. The CO1 gene nucleotide sequences of S-5D along with 45 downloaded sequences for C. auratus was analyzed in
DNAp software which estimated the number of variable sites, S, among 46 sequences to be 13. Total number of mutations, Eta was also calculated to be 13. Ten haplotypes were defined from among 46 sequences. Haplotype (gene) diversity, Hd was calculated to be 0.732 (Standard Deviation: 0.050). Variance of Haplotype diversity was estimated to be 0.00249. Nucleotide diversity (per site), Pi was found to be 0.00377 (Standard Deviation: 0.00035). Average number of nucleotide differences, k was calculated to be 2.28753. Statistical significance of Tajima's D test (0.0152; P > 0.10) [11] and Fu and Li's D test (-1.45155; P > 0.10) [10] were included as an out group. Each of five families viz. Catostomidae, Gyrinocheilidae, Balitoridae and Pseudogyrinocheilidae is represented by a single genus (4 species) and one species in GenBank database while no representatives were found for family Psilorhynchidae. Gyrinocheilidae is represented by a single genus (4 species) of algae feeding mountain carps found in Southeast Asia. Family Psilorhynchidae include a group of freshwater mountain carps (2 genus; 5 species) found in mountain streams of India, Nepal and Burma. Though the study is mainly about phylogenetic analysis of goldfish, these two families of mountain carps which were poorly represented in GenBank database has captured our attention as they reflect the fact that mountain streams around Indian subcontinent has not been explored greatly with respect to biodiversity. Application of molecular tools will greatly increase our knowledge on the biodiversity of the mountain streams along these regions and it also offers immense scope for identifying many new species, totally unknown to scientific world, inhabiting the waters of these mountain streams.

4. Discussion

The mitochondrial cytochrome oxidase gene has been proposed to be the most effective barcoding gene for metazoans. The total length of CO1 gene in vertebrates is about 1545 bp in length and the region nominated as “barcode” comprises of a standardized 648 bp segment commencing from 58 to 705 bp region from 5' terminus. The segment has been defined using mouse mitochondrial genome as reference [14]. The DNA barcoding is gaining despread popularity because of its simplicity and accuracy. Taxonomic identification and phylogenetic analysis using molecular tools facilitate rapid assessment, from field collection to report presentation, while attaining species level resolution [15]. In the present study, we examined the mitochondrial CO1 gene nucleotide sequence obtained from three phenotypic variants of *C. auratus* for assessing intraspecific and interspecific genetic divergence and tracing its phylogenetic history.

BLAST analysis of the nucleotide and amino acid sequences confirmed the three morphological variants to be of *C. auratus* species and sequences to be of CO1 super family. No genetic divergence was observed in all three examined phenotypic variants of goldfish as they exhibited 100% similarity with each other. This indicates that artificial selection which may have resulted in the formation of these phenotypic variants has not enforced any genetic variation in CO1 gene of *C. auratus*. During BLAST analysis of CO1 gene of goldfish it was observed that family Gyrinocheilidae was represented by a single fish species in GenBank database while no representatives were found for family Psilorhynchidae.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>GenBank ID/ Organism</th>
<th>Nucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gb</td>
<td>EF609306.1</td>
</tr>
<tr>
<td>2</td>
<td>gb</td>
<td>JQ979143.1</td>
</tr>
<tr>
<td>3</td>
<td>gb</td>
<td>HQ961040.1</td>
</tr>
<tr>
<td>4</td>
<td>gb</td>
<td>HQ536319.1</td>
</tr>
<tr>
<td>5</td>
<td>gb</td>
<td>JX074196.1</td>
</tr>
<tr>
<td>6</td>
<td>gb</td>
<td>HQ645094.2</td>
</tr>
<tr>
<td>7</td>
<td>gb</td>
<td>JX074208.1</td>
</tr>
<tr>
<td>8</td>
<td>gb</td>
<td>HM536924.1</td>
</tr>
<tr>
<td>9</td>
<td>gb</td>
<td>JX074186.1</td>
</tr>
<tr>
<td>10</td>
<td>gb</td>
<td>GU086572.1</td>
</tr>
<tr>
<td>11</td>
<td>gb</td>
<td>JX074169.1</td>
</tr>
<tr>
<td>12</td>
<td>gb</td>
<td>JQ623926.1</td>
</tr>
<tr>
<td>13</td>
<td>gb</td>
<td>HM536890.1</td>
</tr>
<tr>
<td>14</td>
<td>gb</td>
<td>HM536895.1</td>
</tr>
<tr>
<td>15</td>
<td>gb</td>
<td>HM536915.1</td>
</tr>
<tr>
<td>16</td>
<td>gb</td>
<td>HM536907.1</td>
</tr>
<tr>
<td>17</td>
<td>gb</td>
<td>HM536894.1</td>
</tr>
<tr>
<td>18</td>
<td>gb</td>
<td>JQ667502.1</td>
</tr>
</tbody>
</table>
S-5D along with other members of genus Carassius aligned in a small cluster within the clade formed by Cyprinidae fishes. C. auratus was found to be more closely related to C. gibelio than to C. cuvieri and C. carassius. This is in agreement with the findings of Komiyama et al. [3], according to which C. auratus diverged from C. gibelio. Gyrinocheilidae was found to be the closest relative of family Cyprinidae, though it showed more affinity towards Catostomidae than to Cyprinidae. Cobitidae represented the most distant relative of family Cyprinidae. NJ tree seems to suggest that Cobitidae was the first family to have diverged from the common Cyriniforms ancestor. Balitoridae was the next in line. Cyprinidae then diverged from Gyrinocheilidae and Catostomidae. Analysis of the NJ tree confirmed that goldfishes belong to Cyprinidae family and share a common lineage with modern Cypridae fishes. Intraspccific and Interspecific K2P mean pairwise distance was calculated for S-5D. The K2P interspecific mean distance for 46 sequences of C. auratus was calculated to be 0.0034 while mean interspecific distances for order Cypriniformes and clade Cyprinidae was calculated to be 0.1898 and 0.1237 respectively. The interspecific mean distance calculated for clade Cyprinidae was 36 folds greater to that of intraspecific distance calculated for the CO1 nucleotide sequences of goldfish deposited in GenBank from over the world. This clearly reflects that no considerable genetic variation exists in goldfish with regard to the CO1 gene. Comparison of K2P mean interspecific distances for each clade indicate that Cyprinidae is the most stable clade with respect to genetic variations while Cobitidae was found to be the most unstable group of Cypriniformes. The genetic stability of goldfish, to an extent, can be considered as a characteristic feature of family Cyprinidae as it exhibits least interspecific distance in comparison to other families. A detailed study on other species of family Cyprinidae would be interesting as it might reveal whether or not other Cyprinidae fishes follow the same trend of goldfish. The K2P mean distances of Cypriniformes to other clades reveal that Gyrinocheilidae is its closest relative. The K2P pairwise interspecific distance calculated for clade Cypriniformes confirms C. gibelio to be the closest relative of C. auratus. In our study we discovered genus Barbonymus to be the next of kin to genus Carassius.

Statistical analysis of CO1 gene nucleotide sequences of S-5D along with 45 downloaded sequences for C. auratus was carried out in DnaSP software. Ten haplotypes were defined for the 46 sequences analyzed, though the probability of ten haplotypes was found to be very low, as estimated using Fu's Fs statistic and Strobeck's S statistic. Results of the neutrality tests conducted for genetic variations clearly indicated that no significant genetic polymorphism / divergence exist among the tested CO1 nucleotide sequences. All these results clearly indicate that phenotypic variants of C. auratus which exhibit wide range of morphological variations share a similar gene pool with minimum genetic variations. Artificial selection and geographical isolation has not enforced much genetic variation in goldfish. One of the possible reasons could be the International trade of these ornamental fishes which provide chances of gene pool mixing beyond boundaries.

Conclusion
In the present study we conducted an in-depth analysis of nucleotide sequences of CO1 gene from C. auratus for interspecific and intraspecific genetic variations. On the basis of our results, we report that negligible genetic variation exists among the phenotypic variants of goldfish C. auratus. Carassius gibelio was found to be the closest relative of C. auratus. Family Cyprinidae which include Carassius diverged from Gyrinocheilidae and Catostomidae and was found to be more closely related to Gyrinocheilidae. In comparison to other Cypriniformes families, Cyprinidae was found to be more stable with regard to genetic variations. The effectiveness and accuracy of CO1 gene sequence in assigning “species” to unknown specimens, determining phylogeny of an organism and estimating the genetic divergence has been well represented in the study. Though the importance of morphology based traditional taxonomy remains unchallenged, we would like to emphasize that morphological data should be supplemented with genetic data while describing a new species as specimens seemingly different in morphology might represent a single species. The lack of intraspecific genetic divergence among goldfish suggests that the favorite fish companion of humans has successfully defended its genetic identity against all odds of geographical distance and artificial selection.

Acknowledgement
The authors are grateful to Prof. I. S. Bright Singh, Coordinator, National Centre for Aquatic Animal Health, Kochi, India for providing necessary facilities. The authors also thank the Head, Department of Zoology, Sacred Heart College, Thevara, India for the support in the successful completion of the work.

References
6. J.D. Thomson, D.G. Higgins, T.J. Gibson, CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight

Source of support: Nil; Conflict of interest: None declared