GENETIC DIFFERENTIATION AND MOLECULAR PHYLOGENETICS OF NORTH AFRICAN CATFISH FROM THREE DISTINCT WATERBODIES

Omoniyi Michael Popoola

Department of Fisheries and Aquaculture Technology, Federal University of Technology, Akure, Nigeria
Aquatic Environmental Biotechnology & Nanotechnology Division, ICAR-Central Inland Fisheries Research Institute, Barrackpore, Kolkata-700 120 West Bengal, India

*Corresponding Author: ompopoola@futa.edu.ng

ARTICLE INFO
Received: 9 March 2022
Accepted: 27 May 2022

ABSTRACT
The population structure and genetic variability of North African catfish Clarias gariepinus (Burchell 1822) were investigated using partial mitochondrial DNA cytochrome b region sequences. Fifty-four (54) samples were investigated from three geographically isolated rivers in Nigeria. The analysis of 53 haplotypes revealed greater haplotype diversity (0.99930) and nucleotide diversity (p) (0.07270). According to an analysis of molecular variance (AMOVA), the genetic diversity of North African catfish within populations is significantly higher than the genetic diversity across populations. The FST scores (0.75000, 0.94792 and 0.95699) indicated that North African catfish populations in three Nigerian freshwater bodies had a strong genetic structure. The phylogenetic reconstruction of unique haplotypes revealed the placement of a haplotype (Ogbese) linked by others from all three groups with a point mutation ranging from 1 to 24 nucleotides. North African catfish populations in the Asejire and Ureje are genetically diverse, as evidenced by a high level of haplotype diversity of 1.0000, low nucleotide diversity spanning from 0.05101 to 0.07889, and high FST values (within-population genetic variation). The common haplotypes between some populations and mixes of haplotypes from different populations within the same genetic cluster demonstrate that the population genetic structure is not distinct.

Keywords:
C. gariepinus
Cytochrome b
Median-joining network
Nucleotide substitution
Genetic distances

How to Cite
INTRODUCTION

Fish are regarded as one of the least expensive sources of quality animal protein and minerals for the world’s millions of underprivileged families (Wu and Yang, 2012). The overall global and total harvest of fish, crustaceans and mollusks has continued to rise, reaching 154 million tonnes in 2011 (FAO, 2012). Simultaneously, due to exploitation, human impacts, pollution and global warming, wild stocks have been declining, resulting in biodiversity loss (Limburg and Waldman, 2009). Since 2001, the capture production of fish has been stable at roughly 90.4 million tonnes (Jagadeesh et al., 2013), despite the rising demand for protein from aquatic animals. Genetic variation is beneficial and important for the long-term survival of natural populations as it ensures the provision of high fitness levels, allowing populations to adjust to new environmental conditions (Frankel and Soule, 1981), and it has resulted in a fascinating phenomenon that is anticipated to be the consequence of mutation or migration to a genetically dissimilar population (Frankel and Soule, 1981). Genetic degradation is caused by a lack of understanding in the field of fish farming. It happens as a consequence of inbreeding, negative selection and hybridization, resulting in an excess of homozygosity in the population and a decline in productivity. The escape of hatchery-bred fish has been reported to cause a loss of genetic diversity in the wild population as a result of hybridization and competition between escaped cultured fish and the wild stock or species (Popoola et al., 2014). Proper fishing resource management is critical for socioeconomic reasons and long-term aquaculture benefits (Omitoyin, 2007). A fall in captured fisheries would require both ecological and genetic approaches, necessitating continued research into the conservation and preservation of fish genetic diversity as a vital ecological resource and life support system (Ekelemu and Zelibe, 2006).

The common methodologies used in the identification and characterization of catfishes have been morphological, i.e. meristic, and morphometrics with anatomical features (Kim et al., 2003a). Despite this, there is significant ambiguity due to morphological resemblance, which has led to some contested species identification hypotheses (Koutrakis and Tsikliras, 2003; Kim et al., 2003). According to Froese (2006), the alternative citations in FishBase suggest the possibility of imprecise identification with respect to several catfish species. Particularly in population dynamics evaluation and fisheries management, precise identification of visually similar species is critical (Haye et al., 2012). Kim et al. (2003a) found that molecular tools can provide valuable data for accurate systematics, confirmation of systematic circumstances and phylogeny (Kim et al., 2003b).

Owing to its higher mutation rate than nuclear DNA (Eyre-walker and Awadalla, 2001), simplified mode of transmission that avoids recombination, small genome size, extremely low probability of parental leakage (Cummins et al., 1997) and maternal mode of inheritance, mitochondrial DNA is a very promising genetic marker system for studying the genetic structure of different species (Menezes et al., 2012). According to Behera et al. (2016), cytochrome b is made of conservative and variable areas, as well as the best-known mitochondrial gene. The variability and delineation observed in the mitochondrial cytochrome b region have been broadly used for population-based studies, phylogeography and genetic diversity in fishes across taxonomic orders (Fontana et al., 2007; Oleinik et al., 2007; Bouza et al., 2008; Murray et al., 2008).

North African catfish Clarias gariepinus (Burchell 1822) is a culturable fish that has a wide range of habitats in Africa, ranging from deep and shallow lakes to rivers and swamps (Vitule et al., 2006; Nwani et al., 2014). Its fast growth rate, strong plasticity to new environments (Barasaab et al., 2014) and ability to dwell in stagnant waterbodies have all contributed to its success (Fagbenro and Davies, 2001). It was commercially imported to Europe, Asia and Latin America (Verreth et al., 1993; Vitule et al., 2006). Although the farming of C. gariepinus has benefited many people, the expansion of the fish has had a negative impact on the growth of native species (Popoola et al., 2014). Many studies on its development, reproduction and physiology have been conducted because of its great economic worth (Nwani et al., 2014), but there have been few investigations into its genetics and genomics, particularly its mitochondria genome (mitogenome).

Barasaab et al. (2014) exploited the partial control area of the mitochondrion to infer D-loop sequence variation, genetic diversity and gene flow among C. gariepinus populations from Lake Victoria and its environs. Mohindra et al. (2007) and Wong et al. (2008) employed DNA barcoding and mitogenome RFLP markers to perform genetic identification and phylogenetic assessment of catfish (Luhariya et al., 2011). Although the genetic structure of the Dutch domesticated C. ariepinus cultured in Nigeria has been described using CYTB in Nigeria (Nwafili, 2014), in Southwest Nigeria, there has been little or no research on C. gariepinus employing mitogenome analysis on wild C. gariepinus. As a result, the goal of this work is to use the cytochrome b gene to perform genetic identification and phylogenetic analysis of African catfish from waterbodies within the Southwest Nigeria.

MATERIALS AND METHODS

Sample Collection

Fifty-four (54) samples of C. gariepinus were randomly obtained from local fishermen at landing sites in the three chosen locations: the River Ogbese, Asejire and Ureje in Ondo, Oyo and Ekiti State, respectively (Figure 1). All specimens were gathered between April and June 2020.
Taxonomic keys were used to confirm species (Idodo-Umeh, 2003; Adesulu and Sydenham, 2007). From each fish sample, 100 mg fin clip was collected and stored in 95-percent ethanol until it was used.

**Extraction of genomic DNA**

Whole genomic DNA was isolated from the ethanol preserved-fin clip with TNES-Urea buffer (8 M urea, 10 mM Tris-HCl pH 7.5, 125 mM NaCl, 10 mM EDTA and 1% sodium dodecyl sulphate), using the method described by Han et al. (2018) with minor modifications. A UV spectrophotometer was used to assess the purity of the DNA, which was then diluted for PCR amplification and stored at -80°C until further use.

**Amplification and sequencing**

The primer CYTB F 5’ TGACTTAAAAACCACCGTTGTA 3’ (Bernatchez and Donzman, 1993) and CYTB R: 5’ CTTCGTATACAGGAC 3’ (Ludwig et al., 2000) were used to amplify a 1227bp (1.2Kbp) fragment from the mitochondrial regulatory region. The polymerase chain reactions (PCRs) were run in a 50 μl reaction volume containing 5μl of 10 Ex Taq TM buffer (Inqaba Inc., South Africa), 0.2 mM of each dNTP, 0.4 mM of each primer, 1.25 units of EX Taq TM polymerase (Inqaba Inc., South Africa), and 5 μl genomic DNA were used in each reaction. The thermal cycling reactions comprised an initial denaturation step of 3 min at 95°C followed by 35 cycles of 1 min at 94°C, annealing at 56°C for 1.30 min and extension at 72°C for 2 min, followed by 4 min and 7 min for the final extension at 72°C using an Eppendorf Nexus gradient master cycler (Model 2022-05, USA). The amplified products were visualized on 2% agarose gels and the products with the expected size were selected for bi-directionally sequencing using an ABI 3730xl capillary sequencer at the sequencing facility.

**Data analysis**

With the BioEdit version 7.0 software (Hall, 1999), the CYTB gene sequences (54) were aligned and proofread. The sequence data was examined using the Maximum Likelihood methods, using 3000 repetitions and 1000 bootstrap values. MEGA 11 software (Tamura et al., 2021) was used for all phylogenetic studies. The dataset was analyzed to find out whether ancestral haplotypes are extinct, based on haplotype networks with the PopArt 1.7 (Bandelt et al., 1999) software package. With the software DNASp version 5.10 (Librado and Rozas, 2009), the number of polymorphic sites and nucleotide diversity (Pi), nucleotide composition and the number of transitions with species transversion (mutation) were determined. Analysis of molecular variance (AMOVA) was used in Arlequin 3.5 (Excoffier and Lischer, 2010) to quantify gene diversity within and among populations. Within the populations, the nucleotide and haplotype diversity, as well as the molecular diversity indices (Baisvar et al., 2015), were calculated using Arlequin 3.5.
RESULTS

The CYTB amplified sequences were validated by percentage similarity in the NCBI's BLASTn web software. A higher percentage similarity (97–100%) against the reference sequence was used to confirm the identity of the species under study.

The amplified CYTB sequence produced 821 bp. There were 628 sites that were constant, 55 sites that were variable, 198 parsimony informative sites and 103 singletons. The transition/transversion bias (R) was estimated to be 0.91. The Kimura two-parameter model was used to estimate substitution patterns and rates (Kumar et al., 2018). The observed nucleotides frequency in all the sequenced samples across all sampling sites was determined (Table 1). There were no indels found. The 54 sequences of CYTB gene, produced 53 haplotypes and were submitted to the GeneBank and awarded Accession numbers OL658600-OL658608 and OM675777-OM675821.

The haplotypes were divided into three primary groups by the CYTB dataset, with substantial clusters corresponding to sampling areas (Fig. 2).

### Table 1. Substitution Matrix for Maximum Likelihood Estimate

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>G</th>
<th>Nucleotide frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-</td>
<td>8.83</td>
<td>9.93</td>
<td>7.48</td>
<td>A 28.55%</td>
</tr>
<tr>
<td>T</td>
<td>9.42</td>
<td>-</td>
<td>6.31</td>
<td>4.83</td>
<td>T 26.75%</td>
</tr>
<tr>
<td>C</td>
<td>9.42</td>
<td>5.61</td>
<td>-</td>
<td>4.83</td>
<td>C 30.07%</td>
</tr>
<tr>
<td>G</td>
<td>14.59</td>
<td>8.83</td>
<td>9.93</td>
<td>-</td>
<td>G 14.63%</td>
</tr>
</tbody>
</table>

The probability of substitution (r) from one base (row) to another base is given for each entry (column). The Tamura-Nei (1993) model was used to estimate the substitution pattern and rates. The rates of various transitional substitutions are bolded, whereas those of transversional substitutions are italicized.

The haplotype median connecting network revealed a significant degree of similarity amongst haplotypes, separated by 1–24 mutational steps (Figure 2). Based on the sampling rivers, haplotypes show a degree of similarity (Figure 2). Shared haplotypes occupy the middle area of the network, while unique haplotypes radiate out from the center.

![Fig 2. C. gariepinus minimum spanning haplotype network. Each circle represents a distinct haplotype, with circle size according to frequency. Parentheses represent the mutational phases](image-url)
**Table 2.** Percentage nucleotide composition of CYTB gene sequence of *C. gariepinus* from three populations

<table>
<thead>
<tr>
<th>S. no.</th>
<th>Populations</th>
<th>Nucleotide composition</th>
<th>Haplotype diversity</th>
<th>No of the haplotypes h</th>
<th>Nucleotide diversity Pi</th>
<th>Fst values</th>
<th>Number of segregating sites, S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ureje</td>
<td>54.5</td>
<td>45.5</td>
<td>1.000</td>
<td>18</td>
<td>0.07889</td>
<td>0.75000</td>
</tr>
<tr>
<td>2</td>
<td>Ogbese</td>
<td>54.1</td>
<td>45.9</td>
<td>0.994</td>
<td>17</td>
<td>0.06329</td>
<td>0.94792</td>
</tr>
<tr>
<td>3</td>
<td>Asejire</td>
<td>53.7</td>
<td>46.3</td>
<td>1.000</td>
<td>18</td>
<td>0.05101</td>
<td>0.95699</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>0.999</td>
<td>53</td>
<td>0.07270</td>
<td>0.15952</td>
</tr>
</tbody>
</table>

**Table 3.** Analysis of molecular variance (AMOVA) of genetic variation among and within two *C. gariepinus* populations using 628CYTB

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>Variance components</th>
<th>Percentage of variation</th>
<th>Fixation index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among pops</td>
<td>2</td>
<td>331.59</td>
<td>9.06 Va</td>
<td>75.73</td>
<td>0.75726</td>
</tr>
<tr>
<td>Within pops</td>
<td>51</td>
<td>147.94</td>
<td>2.90 Vb</td>
<td>24.27</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>479.53</td>
<td>11.96</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Stat</td>
<td>P (rand &gt;= data)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PhiPT</td>
<td>0.079</td>
<td>0.010</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nm (Haploid)</td>
<td>5.850</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance (1000 permutations): Phi_ST: Pr(random value > observed Phi_ST) < 0.001

This formation produces a star-shaped profile, indicating population growth. The dominant haplotypes in the clusters are E1, O1 and A8, with ten individuals, O1 had the highest distribution frequency. The central haplotypes in the haplotype network are linked to the others (Asejire and Ureje), indicating that these are the ancestral populations (Figure 2). Furthermore, three separate mitochondrial lineages (networks) were discovered to correspond to geographically defined sampling sites.

For all three groups, nucleotide diversity (p) was 0.05101–0.07889, while haplotype diversity (Hd) was 0.994–1.000 (Table 2). At the Ureje sampling site, the highest haplotype (1.000) and nucleotide (0.07889) diversity was discovered. The mean nucleotide diversity (p) was 0.07270, and haplotype diversity (Hd) was 0.999 (Table 2). The haplotype O1 from the Ureje population was the largest, and it was interconnected with haplotypes from other sampling locations (Ogbese) (Figure 1).

AMOVA for all fifty-four sequences in three groups (based on three different sampling sites) found considerable variance among groups ($P <0.001$) and major variation among populations within groups (75.73 %, $P <0.001$) and within populations (24.27 %, $P <0.001$) (Table 3).

The Maximum Likelihood technique was used to infer the evolutionary history of *C. gariepinus* gathered from three river bodies. Figure 3 depicts the best tree with a sum of branch length of 0.067. Next to the branches are the percentages of duplicate trees in which the related taxa clustered together in the bootstrap test (1000 repetitions). The evolutionary distances were calculated using the Maximum Composite Likelihood technique and are in base substitutions per site unit (Table 4).

**Table 4.** Estimates of genetic distance over Sequence Pairs between populations

<table>
<thead>
<tr>
<th>Populations</th>
<th>Asejire</th>
<th>Ogbese</th>
<th>Ureje</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asejire</td>
<td></td>
<td>0.0519</td>
<td></td>
</tr>
<tr>
<td>Ogbese</td>
<td></td>
<td>0.0518</td>
<td>0.0086</td>
</tr>
<tr>
<td>Ureje</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The median-joining network (Figure 3) had a broad topology that matched the ML tree (Fig. 3), with three significant clusters detected. The three populations showed primarily two distinct groups, confirming the resilience of the study’s species-specific molecular markers. Individuals from the same sampling sites are grouped together, indicating the outcome of the median network analysis. However, the trees produced revealed that the sampled fish belong to the same genera, with the Asejire samples belonging to a separate clade and the Ogbese samples with the Ureje samples belonging to a separate clade, accurately indicating the evolutionary relationship for the *C. gariepinus* species under study (Figure 3).
DISCUSSION

From individual fitness to ecosystem function, genetic diversity is critical for ecological and evolutionary processes (Rezvani et al., 2012). Heterogeneity is an important evolutionary signal for predicting population dynamics and survival (Reed, 2009). In contrast to Rupesh et al. (2011) and Behera et al. (2016), this study's investigation of the CYTB gene found a high degree of genetic variation among the populations and a low amount of genetic diversity within the population. The populations’ total Fst score of *C. gariepinus* was 0.115664, showing minimal genetic diversity within the populations. In contrast to the Ogbese and Ureje populations, the estimates of genetic distance between populations were 0.0519 between the Asejire and Ogbese, which was high but non-significant from each other because the sampling sites are physically apart from one another. As a result, estimating the population's genetic diversity could indicate a substantially closer link between the sampling sites.

When compared to other places, haplotype diversity in the Ogbese was lower, but the high value could indicate that the population under study has advanced. According to Grant and Bowen (1998), high haplotype diversities indicate a rapid population increase with the accumulation of novel mutations following a period of low effective population size.

Haplotype O1 is detected in all Ogbese samples, as well as major haplotypes from the Ureje samples (E1). This likeness may have evolved independently from various traits in their common ancestor due to homoplasy, when convergence or parallelism occurs and the perceived similarity evolves (Daniel et al., 1995). The high frequency of O1 in most sampling sites, as well as its extensive geographical distribution, suggests that all haplotypes within the *C. gariepinus* CYTB gene mitochondrial lineage could be monophyletic.

Fst is a genomic structure-based measure of population differentiation. In distinguishing populations, a Fst value larger than 0.15 can be regarded as important (Frankham et al., 2002). According to the Fst values obtained, there was a significant divergence within each of the *C. gariepinus* populations. However, the two groups had a low Fst value (0.75), indicating that there was little genetic difference between them (Table 3). This is supported by the AMOVA results, which showed that within-subpopulation variations accounted for 76% of the total variation, whereas among-subpopulation variations accounted for just 24% of the total variation.

Wright (1965) stated that Nm values of less than one indicate limited gene exchange among subpopulations, whereas the Nm value of 5.85 in this study was high, implying that high genetic exchange or gene flow (Eltaher et al., 2018) may have occurred, resulting in low genetic differentiation between populations.
The genetic distance between populations is an essential measure of evolutionary events and mutation. The genetic structure of *C. gariepinus* populations was found to be rather poor. The fish samples from the Asejire River are notably distinct from the other populations, according to the population genetic structure based on pair-wise values (Ogbese and Ureje River). This is consistent with the findings of Yodsiri et al. (2017) who found that random sampling of an allele from the colonizer's source population can result in genetic divergence. Similarly, Jang et al. (2017) found that river branches induce ecophysiological changes that restrict gene flow between populations.

Gene flow, on the other hand, is thought to be feasible when a species migrates from one habitat to another over a long period of time (Farrag et al., 2015). Geographic remoteness is a powerful tool for increasing genetic diversity and influencing gene flow (Pramono et al., 2019). Within-population variation was measured by nucleotide diversity and haplotype. The Ureje and Asejire populations have the highest and lowest haplotype diversity and nucleotide diversity, respectively (Table 3).

These values have long been thought to be a sensitive indicator of a population's genetic diversity (Nei and Li, 1979). The haplotype diversity of the Asejire population is comparable to that of *Chitala chitala* (Mandal et al., 2012) and *Labeo rohita* (Mandal et al., 2012; Luhariya et al., 2011). High haplotype and nucleotide diversity numbers indicate a high level of population variety, as seen in the Ureje population. Several variables have been linked to the genetic divergence of aquatic animal groups. Geographic segregation and environmental variations may cause divergence, but river tides, migration and human activities may cause divergence to diminish (Lehmann et al., 1998). The Asejire River is a large river that runs for a long distance and passes through important cities. This could explain the poor diversity index readings. Furthermore, a quick decline in a colony's population could lead to genetic divergence (Hedrick, 1999). In comparison to other groups, the Ureje population lives in relatively undisturbed habitats with only limited human-caused damage. This could explain why the Ureje population has higher genetic diversity. Nevertheless, sampling error could have affected the detected variability in the investigated river populations. Nonetheless, the current work could serve as a starting point for a thorough investigation of the diversity of *C. gariepinus* in Nigerian waterbodies.

The topologies generated by the phylogenetic tree had the three populations in separate groupings. Similar samples in the population clustered under the same nodes, while different fish samples clustered under separate nodes, establishing a phylogenetic link among the sampled fish. In the CYTB sequence data, there is certainly some phylogenetic signal. Ward et al. (2005) found that congeneric species, as well as confamilial species, always clustered together. Bineesh et al. (2015) and Saraswat et al. (2013) found similar results in fish species from the Indian seas. Two significant clades of fish samples were identified (figure 3). In one clade, the Asejire sample (group 1) was clearly separated, whilst other groups were discovered in the same clade. Species were clustered separately based on their location, implying that there is a possibility of interbreeding or sharing parentage linkage among species within collection rivers. This is consistent with the findings of Swain et al. (2014) who found that geographic separation occurs during the early developmental stages in *C. gariepinus* species, implying that geographic delineation is visible in *C. gariepinus* species.

The usefulness of the CYTB gene in distinguishing *C. gariepinus* has been proven in this study. The aquaculture of *C. gariepinus* necessitates prior knowledge of the genetic composition of potential source populations, and the findings of this study may aid fish farmers in identifying the best source of parental stock for improved performance through long-term selective breeding programmes for *C. gariepinus* fish. This finding, which was based on partial CYTB gene sequences, demonstrated the stock differentiation and genetic diversity of *C. gariepinus* in three river bodies in Southwest Nigeria and provided important information for the development of conservation and management strategies for wild fish populations.

**ACKNOWLEDGMENT**

The author wishes to express his gratitude to the Head of the Department of Fisheries and Aquaculture Technology at the Federal University of Technology, Akure, Nigeria, as well as the staff of the National Centre for Genetic Resources and Biotechnology in Ibadan, Nigeria for their assistance and encouragement in achieving success in this study.

**GENETSKA DIFERENCIJACIJA I MOLEKULARNA FILOGENETIKA AFRIČKOG SOMA IZ TRI RAZLIČITE VODE**

**SAŽETAK**

Populacijska struktura i genetska varijabilnost afričkog soma istražena je korištenjem djelomičnih sekvenci mitochondrijske DNA citokroma b. Pedeset četiri (54) uzorka su istražena iz tri zemljopisno izolirane rijeke u Nigeriji. Analiza 53 haplotipa pokazala je veću raznolikost haplotipova (0,99930) i nukleotidnu raznolikost (p) (0,07270). Prema analizi molekularne varijance (AMOVA), genetska raznolikost afričkog soma unutar populacija značajno je veća od genetske raznolikosti među populacijama. FST rezultati (0,75000, 0,94792 i 0,95699) pokazuju da populacije afričkog soma u tri nigerijska
REFERENCES


Han, Z., Sun, J., Lv, A., Yeong-Yik, S., Xueliang, S., Hongyue, S., Xiucai, H., Anli, W., Kezhi, X. (2018): A modified method for genomic DNA extraction from the fish intestinal microflora. AMB Express, 8 (1), 52.


