Genetic variation and identification of single nucleotide polymorphism of insulin-like growth factor- 1 gene in *Tilapia guineensis*

ESTHER UKENYE[1,♥], IWALEWA MEGBOWON[1], OLALEKAN OGUNTADE[1], TOPE OKETOKI[1], OLUWAFEMI AMUSA[2], ADEOLA USMAN[1], BALKEES SOKENU[1], RAKIYA ADELEKE[1], BAWA JOSEPH[1], CHRISTIAN OMATAH[1]

[1]Department of Biotechnology, Nigerian Institute for Oceanography and Marine Research. Victoria Island, Lagos, Nigeria.
♥email: eekwelem@yahoo.com
[2]Department of Cell Biology and Genetics, University of Lagos. Lagos, Nigeria

Abstract. *Ukenye E, Megbowon I, Oguntade O, Oketoki T, Amusa O, Usman A, Sokenu B, Adeleke R, Joseph B, Omatah C. 2020. Genetic variation and identification of single nucleotide polymorphism of insulin-like growth factor- 1 gene in Tilapia guineensis Biodiversitas 21: 5317-5321.* Tilapia guineensis is an important cichlid species of West African coastal waters with good nutritional, economic, and aquaculture relevance. The knowledge of the genetic basis of variation in growth traits in Tilapia fish is of great importance to support genetic improvement in the context of aquaculture. In this study, regions of the *Tilapia guineensis* IGF-1 gene were sequenced, aligned and compared across populations to identify single nucleotide polymorphism and genetic diversity among four populations of *T. guineensis* in South-west Nigerian coastal waters. A total of thirty-four SNPs were identified across the studied populations and were detected in the forward reaction with twenty-two transversions and twelve transitions. Badagry population showed the highest genetic diversity with the highest molecular diversity indices; number of polymorphic sites, pairwise differences, number of segregating sites and nucleotide diversity while the least diverse population was Pepe. Analysis of molecular variance (AMOVA) revealed that genetic variation was mostly within populations. This finding provides more information regarding variation in insulin growth factor I of *T. guineensis* and will encourage association study for production traits that will inform useful selection marker for breeding program.

Keywords: Igf-1, genetic diversity, SNPs, *Tilapia guineensis*

INTRODUCTION

Tilapia guineensis is among the most exploited fish species of cichlids usually found in brackish water regions of West Africa and has good aquaculture potential. It has been successfully raised in ponds, enclosures, cages, and tanks and accepts the high density necessary for economical culture (Penna-Mendoza 2005). *T. guineensis* has continued to contribute significantly to the nutrition value and economic growth of many developing countries. In spite of these potentials, *T. guineensis* like other Tilapias is still faced with the problem of stunted growth due to its prolific nature. Information on the molecular basis of growth traits for growth and productivity enhancement of *T. guineensis* will increase aquaculture production. Consequently, genetic variation and single nucleotide polymorphism identification from candidate genes associated with growth as a molecular technique to improve its growth is crucial.

In view of the success of genetic improvement programs focused on the selection of faster-growing traits, identification of the genetic variation in candidate genes related to growth and other traits of economic importance is an essential step (Lupchinski Jr et al. 2011). According to Zhang et al. (2018), growth traits of animals are always of primary concern during breeding for its determinant economical value. Fish growth is an important productive trait in aquaculture and identification of single nucleotide polymorphism (SNP) for growth performance is considered as a key step during breeding. Growth traits of fishes and other farmed animals could possibly be improved by selective breeding under specific conditions (Zhang et al. 2018). The contribution of genetic variation as an essential molecular tool in selective breeding for growth enhancement and conservation programs in aquaculture and fisheries cannot be overemphasized. Therefore, the improvement of relatively slow growth rates in Tilapia as reported by many workers is considered necessary with a view to improve aquaculture production.

Insulin-like growth factor I (IGF-1) is one of the candidate genes associated with growth in different species of fish and is considered a highly conserved and crucial growth-regulating gene in Tilapia species (De-Santis and Jerry 2007). It plays important role in fish growth, development and metabolism. It is also a mediator of the endocrine action of hormonal growth and is able to stimulate the synthesis of DNA and subsequent cell division in many cell types. Microsatellites and single nucleotide polymorphism are mainly molecular markers of choice for genetic studies in fishes (Sundaray et al. 2016). Due to the large numbers of SNPs in almost all populations of organisms for genome-wide association studies, SNPs are becoming the most abundant type of DNA sequence polymorphism with important role in genetic studies

(Danzamann et al. 2016). According to Yáñez et al. (2015), SNP has been successfully used for the identification of candidate genes of traits and strains useful in aquaculture. A 7K SNP-chip has been used to study genetic differences between farmed and wild Atlantic salmon in Norway (Karlsson et al. 2011); SNPs were also used to assess the genetic impact of gene flow on wild populations of Atlantic salmon in the Gulf of Finland, Baltic Sea (Ozerov et al. 2016). Cuevas-Rodriguez et al. (2016) successfully identified genetic variation and single nucleotide polymorphism in candidate genes for growth in *Oreochromis niloticus*.

Although, major aquaculture efforts have been focused on *Oreochromis niloticus* particularly in the area of molecular mechanisms of some candidate genes associated with growth and selective breeding to improve its growth and overall productivity, but not much attention has been given to *T. guineensis* species commonly found around us in this research area. Therefore, identifying and validating genetic variation and single nucleotide polymorphisms (SNP) in *IGF1* gene related to growth in *T. guineensis* necessitated this study.

MATERIAL AND METHODS

Fish samples collection and DNA extraction

Forty (40) broodstocks of *T. guineensis* belonging to four different populations (10 per location) namely; Pepe and Ugbonla from Ondo state, Badagry and Epe from Lagos state were used for the present study. The coordinates of the sampling locations are presented in Table 1. Phenol-chloroform protocol Sambrook and Russel (2001) was used to extract genomic DNA from the caudal fin tissue of the experimental fish. The purity check of extracted DNA was done using a Nano-drop spectrophotometer (Shimadzu corporation Japan, MODEL UV-1800, 2000 series) at absorbance of 260/280nm. After checking the quality and quantity of genomic DNA, it was diluted to a final concentration of 50-100 ng/µl in highly purified water for amplification.

Polymerase chain reaction amplification

To identify SNPs within the IGF-1 gene, the specific primer (ALT-IGF-1 forward-CTTGGACGAGTAGGAG GCAAATG and ALT-IGF-1 reverse- GAAATACAAG CAAGCGATAAGAA) of 399bp was designed to amplify regions of the IGF-1gene. The Tilapia specific primer (ALT-IGF-1) was developed from the sequence available from public database and synthesized. DNA amplification was carried out by polymerase chain reaction (PCR) in a thermocycler (Eppendorf Master Cycler) at Inqaba biotechnology laboratory, South Africa. 50 ng DNA, 1.25U Taq DNA polymerase, 0.4 mM dNTPs and $MgCl_2$ at the concentration 12.5mM and the annealing temperatures 55 °C were used for the primer-pair. PCR products were analyzed by gel electrophoresis on 1% agarose gel with ethidium bromide. This was followed by PCR product purification. The primer sequences, annealing temperature and amplicon size of the amplified fragments are shown in Table 2.

Table 1. Geographical location of the sampling stations

Location	Latitude	Longitude	State
Pepe	N06 10[1] 01.3[1]	E02[0] 52.988[1]	Ondo
Ugbonla	N060 08[1] 31.1[1]	E004[0] 47[1] 39.8[1]	Ondo
Badagry	N04[0] 25.012[1]	E02[0] 52.988[1]	Lagos
Epe	N04[0] 27.200[1]	E007[0] 19.618[1]	Lagos

DNA sequencing and SNP identification

Purified PCR products were bidirectional-sequenced in an automatic sequencer (ABI 3500XL Genetic analyzer), at Inqaba biotechnology laboratory, South Africa. Immediately nucleotide sequences were obtained, the sequence of each fragment from each fish sample was aligned and SNP identification was done using Arlequin v. 3.5.2 software.

Sequence analyses

The aligned sequences were utilized to detect genetic diversity indices, single nucleotide polymorphisms, nucleotide diversity, polymorphic sites and mean number of pairwise differences using Arlequin software. The molecular diversity indices measured include number of haplotypes (H), number of segregating sites, average number of nucleotide differences (K), singleton variable site and parsimony-informative site for each population of *T. guineensis*. Percentage of variation among and within-population was also estimated using AMOVA.

RESULTS AND DISCUSSION

Polymorphisms detected

ALT-IGF-1 gene was used to investigate and validate genetic variation and single nucleotide polymorphism associated with growth traits in *T. guineensis*. Overall, 34 polymorphic sites with 34 SNPs were identified in four populations. Out of 34 SNPs, 22 were transversions and 12 were transitions; 33 were revealed in Badagry, 1 in Epe, non in Ugbonla and Pepe populations respectively. The highest number of polymorphism was detected in samples from Badagry with Ugbonla and Pepe samples showing no polymorphism. On the basis of the number of SNP loci in this gene fragment among the four *T. guineensis* populations, the most diversity was observed in the Badagry population while the least diversity was observed in Ugbonla and Pepe (Table 3). A total of 8 indel polymorphisms were detected across the four populations.

The value of expected heterozygosity ranged from 0.002 to 0.021 with the lowest value in Pepe population of *T. guineensis* (Table 3). The polymorphisms and the different population carrying it is shown in table 4. It is observed that Badagry population had 33 SNPs out of the 34 SNPs identified in these populations. However, Pepe and Ugbonla did not show any polymorphism. In table 5, Badagry population has the highest nucleotide diversity, mean number of pairwise differences and number of polymorphic sites. Overall, the populations recorded low nucleotide diversity. Badagry population showed higher

molecular diversity indices Theta S, haplotypes (H), average number of nucleotide differences (K) and the highest nucleotide percentage of cytosine and thymine compared to other populations studied. The AMOVA analysis showed that genetic variation (98%) was mostly within population (Table 6).

Table 2. Primer sequence, annealing temperature and amplicon size, used in amplification of IGF-1 gene.

Primer	Sequence	Annealing temperature (°C)	Amplicon size (bp)
ALT-IGF-1	F-CTTAGGASGAGTAGGAGGCAAATG R-GTGAAATACAAGCAAGCGATAAG	55	399

Table 3. Genetic diversity and SNPs of IGF-1 gene in four *Tilapia guineensis* populations

Population	A	He	Transitions	Transversions	Indels
Badagry	1.101±0.334	0.021±0.072	12	22	3
Ugbonla	1.005±0.071	0.001±0.020	0	0	2
Epe	1.005±0.020	0.001±0.020	0	1	1
Pepe	1.005±0.071	0.002±0.033	0	0	2

A: average number of alleles, He: expected heterozygosity

Table 4. Single nucleotide polymorphism of IGF-1 gene in *Tilapia guineensis* populations

No	Position	SNP	Mutation	Population
1	2	T>G	Transversion	Badagry
2	14	T>A	Transversion	Badagry
3	19	A>C	Transversion	Badagry
4	21	A>C	Transversion	Badagry
5	30	A>T	Transversion	Badagry
6	34	A>C	Transitions	Badagry
7	35	C>A	Transversion	Badagry
8	36	T>C	Transitions	Badagry
9	38	T>C	Transitions	Badagry
10	43	A>T	Transversion	Badagry
11	56	T>A	Transversion	Badagry
12	58	G>C	Transversion	Badagry
13	62	A>T	Transversion	Badagry
14	64	T>C	Transitions	Badagry
15	72	C>A	Transitions	Badagry
16	78	A>T	Transversion	Badagry
17	89	G>T	Transversion	Badagry
18	93	T>A	Transversion	Badagry
19	115	G>C	Transversion	Badagry
20	144	T>A	Transversion	Badagry
21	154	T>A	Transversion	Badagry
22	160	T>A	Transversion	Badagry
23	199	T>G	Transversion	Badagry
24	254	T>A	Transversion	Badagry
25	310	T>G	Transversion	Badagry
26	335	C>T	Transitions	Badagry
27	339	T>C	Transitions	Badagry
28	343	A>C	Transversion	Badagry
29	356	G>A	Transitions	Badagry
30	1	T>G	Transitions	Badagry
31	4	C>A	Transitions	Badagry
32	20	A>T	Transitions	Badagry
33	3	T>A	Transitions	Badagry
34	4	C>A	Tranversion	Epe

Table 5. Summary of molecular diversity Indices of *Tilapia guineensis* populations studied

Indices	Badagry	Ugbonla	Epe	Pepe
Nucleotide diversity	0.021±0.012	0.001±0.001	0.001±0.001	0.002±0.002
No. Poplymorphic sites	33	0	1	0
Mean no. pairwise difference	8.377±4.240	0.556±0.494	0.556±0.494	7.667±3.907
Theta (S)	11.665	0.000	0.353	0.000
Theta (H)	2.403	0.542	0.781	2.122
Theta (K)	3.302	1.052	1.052	1.956
Nucleotide Composition				
Cytosine (C)	22.21%	21.99%	21.96%	22.00%
Thymide (T)	35.88%	35.86%	35.69%	25.84%
Adenine (A)	23.95%	23.96%	23.98%	23.95%
Guanine (G)	18.16%	18.20%	18.19%	18.21%

Note: Theta (S) - number of segregating sites, H- the number of haplotypes, K-average number of nucleotide differences

Table 6. Analysis of molecular variance (AMOVA)

Source of variation	df	Sum of squares	Variance component	Percentage of variation
Among population	3	4.4	0.017	1.24
Within population	36	46.9	1.303	98.76
Total	39	51.3	1.32	

Discussion

In the current study, sequence alignment of *Tilapia guineensis* ALT-IGF-1 gene was used to investigate and validate genetic variation and single nucleotide polymorphism associated with growth traits in tilapia. Previous studies have confirmed this candidate gene approach as a strong technique for the identification of polymorphisms in genes that associate with economic traits (Hemmer-Hansen et al. 2011). The present study revealed the existence of 34 SNPs detected in its coding and noncoding region across the four populations which is an indication of genetic polymorphism in *T. guineensis* populations. This finding is in accordance with the earlier reports of Cuevas-Rodríguez et al. (2016) on the novel single nucleotide polymorphisms in candidate genes for growth in tilapia (*Oreochromis niloticus*). They identified single nucleotide polymorphisms that will allow their association with growth traits in *O. niloticus*. Salem et al. (2012) also identified 22 SNPs associated with growth in rainbow trout. Roman (2018) stated that if an adequate number of SNPs is identified and used, mutations in genes associated with a trait can be detected and used in marker-assisted selection. The identified SNPs facilitate the finding of genotypes useful in aquaculture. Badagry population displayed the most diversity among the four populations followed by Epe population as demonstrated by the number of SNP loci and polymorphic sites. This suggests that there is diversity in IGF-1 gene in the Badagry and Epe populations of *T. guineensis*. A similar observation was made by Ukenye et al. (2016) on genetic diversity assessment of *T. guineensis* using microsatellite markers where Badagry population exhibited high genetic diversity. The observed polymorphisms in Badagry samples conform to a significant difference in body and Caudal Peduncle Depth of *T. guineensis* species from Badagry lagoon reported by Kuton and Adeniyi (2014). These results indicate that the values of each SNP site vary among the four *T. guineensis* populations. The different SNP sites might differentially affect *T. guineensis* growth traits and may be used as breeding selection markers for *T. guineensis*. It is noteworthy that association between single nucleotide polymorphisms detected and phenotypic traits was not investigated in the present study. Therefore, further study could be on the association between single nucleotide polymorphism and size of the specimen. The complete coding DNA sequence of the ALT-IGF-1 gene of the *T. guineensis* species shared 100% identity with *Oreochromis aureus* (GeneBank accession No.: XM031746329.1), Maylandia zebra (GeneBank accession No.: XM012918326.3), and Pundamilia nyererei (GeneBank accession No.: XM013914390.1). The results suggest that conservation of the ALT-IGF-1 gene is very high across different species. Anurodh et al. (2014) made a similar observation in sequence characterization of IGF1 gene in Indian goats. Similarly, Ugbonla and Pepe did not have any polymorphism, which might suggest that the region is relatively conserved in the populations, as this confirms little to no recombination or mutation within the gene region. This might affect the morphology in terms of body size and growth of the *T. guineensis* species in these populations. This finding is similar to that of Babatunde et al. (2016) who identified no polymorphism in IGF-1 gene of the chicken strains studied.

The low nucleotide diversity observed in all the studied populations of *T. guineensis* suggested low mutation rate in this region and indicated a relatively low level of genetic diversity, which may be a result of genetic drift and might have contributed to low growth. Hence, there is urgent need for conservation. This finding is in line with the report of Agbebi et al. (2016) who observed low nucleotide diversity in all the four Tilapia species studied. Babatunde et al. (2016) also reported a low level of nucleotide diversity in Nigerian locally adapted chickens. The SNPs polymorphism revealed a total of 22 transversions and 12 transitions; transversions are higher than transitions. This result suggests that transversions are more easily

substituted than transitions. This finding is contrary to other previous research (Vera et al. 2011).

According to Tian et al. (2014), genetic heterozygosity indicates the proportion of the group with site heterozygotes at some loci. The average heterozygosity of loci reflects the level of variation in the genetic structure, and variability is directly correlated with heterozygosity and the ability to adapt to the environment. Among the four populations, the average expected heterozygosity and molecular diversity indices of the Badagry population were higher than others reflecting greater diversity for Badagry population. Thus, Badagry population represents the best population of *T. guineensis* associated with growth traits in the present study. The least diversity that was observed in Ugbonla and Pepe populations based on number of SNPs may be due to inbreeding or founder effect. Analysis of molecular variance (AMOVA) revealed that genetic variation was mostly within populations suggesting that bulk of the variation seems to be accounted for by differences within populations.

In conclusion, Badagry population that showed greater diversity may represent the best population of *T. guineensis* connected with growth traits among the populations studied. The polymorphisms identified in the current study indicated some level of genetic variability hence; SNPs identified in the *T. guineensis* ALT-IGF-1 gene contributed to the genetic variation in *T. guineensis* populations and should be further investigated through association study to detect their production potential which may inform useful selection marker for breeding program.

ACKNOWLEDGEMENTS

REFERENCES

Agbebi OT, Echeful CJ, Adeosun IO, Ajibade AH, Adegbite EA, Adebambo AO, Ilori MB, Durosaro SO, and Ajibike AB. 2016. Mitochondrial diversity and time divergence of commonly cultured cichlids in Nigeria. British Biotechnol J 13 (2): 17. DOI: 10.9734/BBJ/2016/25470.

Anurodh S, Gautam D, Jayakumar S, Vinita S, Dixit SP. 2014. Sequence characterization and genetic variability analysis of GHR, IGF1, and IGFBP-3 genes in nine Indian goat breeds. J Appl Anim Res 42 (3): 361-365. DOI: 10.1080/09712119.2013.875899.

Babatunde MI, Wheto M, Olutunde SD, Kayode A, Ayotunde OA, Olufunmilayo A. 2016. Polymorphism of IGF-1 Promoter and the UTR Regions of Nigerian Locally Adapted Chickens. J Biol Agric Healthcare 6: 2224-3208.

Cuevas-Rodríguez BL, Sifuentes-Rincón AM, Ambriz-Morales P, García-Ulloa M, Valdez-González FJ, Rodríguez-González H. 2016. Novel single nucleotide polymorphisms in candidate genes for growth in

tilapia (*Oreochromis niloticus*). Revista Brasileira de Zootecnia 45 (6): 345-348.

Danzmann RG, Kocmarek AL, Norman JD, Rexroad CE, Palti Y. 2016. Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. BMC Genomics 17: 60. DOI: 10.1186/s12864-016-2363-5.

De-Santis C, Jerry DR. 2007. Candidate growth genes in finfish - Where should we be looking?. Aquaculture 272 (1-4): 22-38. DOI: 10.1016/j.aquaculture.2007.08.036.

Hemmer-Hansen J, Nielsen EEG, Meldrup D, Mittelholzer C. 2011. Identification of single nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel organism; the Atlantic cod, Gadus morhua. Mol Ecol Resour 11 (1): 71-80. DOI: 10.1111/j.1755-0998.2010.02940.x

Karlsson S, Moen T, Lien S, Glover KA, Hindar K. 2011. Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol Ecol Res 11 (1): 247-253. Mol Ecol Resour 11 (1): 247-253. DOI: 10.1111/j.1755-0998.2010.02959.x.

Kuton MP, Adeniyi BT. 2014. Morphological variations of *Tilapia guineensis* (Bleeker 1862) and *Sarotherodon melanotheron* (Ruppell1852) (Pisces: Cichlidea) from Badagry and Lagos lagoon, South-West, Nigeria. J Fisheries Livest Prod 2: 112.

Lupchinski Jr E, Vargas L, Lopera-Barrero NM, Ribeiro RP, Povh JA, Gasparino E, Braccini G L. 2011. Genetic characterization of three Nile tilapia (*Oreochromis niloticus*) strains. Archivos de Zootecnia 60: 985-995.

Ozerov MY, Gross R, Bruneaux M, Vaha JP, Burimski O, Pukk L, Vasemagi A. 2016. Genomewide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. Mol Ecol 25 (6): 1275-1293. DOI: 10.1111/mec.13570.

Penna-Mendoza, BJL, Gomez IH, Salgado-Ugerte, Ramirez-Nogguera D. 2005. Reproductive biology of *Oreochromis niloticus* (Perciformes: Cichlidae) at Emiliano Zapata dam, Morelos, Mexico. Re Biol Trop 53: 515-522.

Roman W 2018. Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. Aquatic Living Resources 31: 2.

Salem M, Vallejo RL, Leeds TD, Palti Y, Liu S, Sabbagh A, Rexroad 3rd CE, Yao J. 2012. RNA-Seq identifies SNP markers for growth traits in rainbow trout. PLoS One 7 (5): e36264. DOI: 10.1371/journal.pone.0036264.

Sambrook J, Russell DW. 2001. Molecular cloning: A laboratory Manual 3rd edition. Cold Spring Harbour Laboratory Press.

Sundaray JK, Rasal KD, Chakrapani V, Swain P, Kumar D, Ninawe ASNS, Jayasankar P. 2016. Simple sequence repeats (SSRs) markers in fish genomic research and their acceleration via next-generation sequencing and computational approaches. Aquacult Intl 24: 1089-1102. DOI: 10.1007/s10499-016-9973-4.

Tian YG, Yue M, Gu Y, Gu WW, Wang YJ. 2014. Single-nucleotide polymorphism analysis of GH, GHR, and IGF-1 genes in minipigs. Braz J Med Biol Res 47 (9): 753-758. DOI: 10.1590/1414-431X20143945.

Ukenye EA, Taiwo IA, Oguntade OR, Oketoki TO, Usman AB. 2016. Molecular characterization and genetic diversity assessment of *Tilapia guineensis* from some coastal rivers in Nigeria. Afr J Biotechnol 15 (1): 20-28.

Vera M, Antonio J, Dios A, Millán A, Pardo AG, Bouza C, Hermida M, Fernández C Herr n, Molina-Luzón M, Paulino M. 2011. Validation of single nucleotide polymorphism (SNP) markers from an immune Expressed Sequence Tag (EST) turbot *Scophthalmus maximus*. Aquaculture 313: (1-4): 31-41 DOI: 10.1016/j.aquaculture.2011.01.038.

Yáñez JM, Newman S, Houston RD. 2015. Genomics in aquaculture to better understand species biology and accelerate genetic progress. Front Genet 6: 128. DOI: 10.3389/fgene.2015.00128.

Zhang H, Zhao C, Yin S, Li Z, Cao Q, Li X, Xie W, Zhang J, Zhu W, Wang D. 2018. Characterization and identification of single nucleotide polymorphism within the IGF-1R gene associated with growth traits of *Odontobutis potamophila*. J World Aquac Soc 49: 366-379. DOI: 10.1111/jwas.12504.