# African Journal of Biotechnology

Volume 15 Number 1, 6 January 2016

ISSN 1684-5315

Academic Journals

# ABOUT AJB

The **African Journal of Biotechnology (AJB)** (ISSN 1684-5315) is published weekly (one volume per year) by Academic Journals.

**African Journal of Biotechnology (AJB),** a new broad-based journal, is an open access journal that was founded on two key tenets: To publish the most exciting research in all areas of applied biochemistry, industrial microbiology, molecular biology, genomics and proteomics, food and agricultural technologies, and metabolic engineering. Secondly, to provide the most rapid turn-around time possible for reviewing and publishing, and to disseminate the articles freely for teaching and reference purposes. All articles published in AJB are peer-reviewed.

**Contact Us**

**Editorial Office:** ajb@academicjournals.org

**Help Desk:** helpdesk@academicjournals.org

**Website:** http://www.academicjournals.org/journal/AJB

**Submit manuscript online** http://ms.academicjournals.me/

African Journal of Biotechnology

Table of Content: Volume 15 Number 1 6 January, 2016

## ARTICLES

academicJournals

Vol. 15(1), pp. 1-6, 6 January, 2016
DOI: 10.5897/AJB2015.14996
Article Number: 1BB691C56979
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

*Full Length Research Paper*

# Isolation and characterization of polymorphic microsatellite loci from transcriptome sequence of *Ruditapes philippinarum*

**Jia Li[1,2], Hongtao Nie[1,2], Depeng Zhu[1], Yanjie Qin[1], Zhongming Huo[1] and Xiwu Yan[1,2]***

[1]Engineering and Technology Research Center of Shellfish Breeding in Liaoning Province, Dalian Ocean University, Dalian 116023, China.
[2]Key Laboratory of Mariculture and Stock Enhancement in North China's Sea, Ministry of Agriculture, Dalian Ocean University, Dalian, 116023, China.

Received 18 September, 2015; Accepted 18 November, 2015

**In this study, 38 novel polymorphic microsatellite markers derived from transcriptome sequence of *Ruditapes philippinarum* were reported. The polymorphisms of these markers were detected in a natural population of *R. philippinarum*. The number of alleles per locus ranged from 2 to 10 with an average of 4.8. The observed and expected heterozygosity per locus ranged from 0.000 to 0.939 and 0.086 to 0.832, with an average of 0.255 and 0.542, respectively. Eighteen loci significantly deviated from Hardy-Weinberg equilibrium after Bonferroni correction ($P < 0.0013$) and two pairwise combinations of four loci (RpT178/RpT146 and RpT200/RpT249) were significant after Bonferroni correction. These loci will provide useful information for the studies on genetic diversity and structure, construction of genetic linkage maps and the effective management of *R. philippinarum*.**

**Key words:** *Ruditapes philippinarum*, microsatellite, polymorphism, genetic diversity.

## INTRODUCTION

The Manila clam, *Ruditapes philippinarum*, is an economically-important marine bivalve species of the China aquaculture industry and is widely distributed in the coasts of China. The world production of this species was 3.9 million metric tons in 2013. China is the first largest country in the world in terms of production of the Manila clam, producing about 3.0 million metric tons annually, which accounts for about 90% of global production (Zhang and Yan, 2010). In aquaculture, genetic diversity is the fundamental resource on which stock improvements rely. However, many aquaculture practices, such as producing large numbers of offspring from a few parents, inbreeding and using broodstocks derived from hatchery seed, are likely to reduce genetic diversity and thereby diminish disease resistance and reduce the population's ability to adapt to new environments (Allendorf and Phelps, 1980). In recent years, the wild resources of *R. philippinarum* have experienced dramatic population declines due to over-exploitation and the deterioration of coastal environment.

*Corresponding author. E-mail: yanxiwu@dlou.edu.cn. Tel: 86-411-84763026. Fax: 86-411-84763026.

The significant decline of wild *R. philippinarum* stocks makes people to pay close attention to its genetic variation and population structure which will provide essential information on maintenance and management of the clam resources (An et al., 2012; Mura et al., 2012; Xing et al., 2014).

Microsatellite or simple sequence repeat (SSR) markers, which are inherited in a Mendelian fashion as codominant markers, have been increasingly popular in genetic studies because of their high levels of allelic variability, wide dispersal and abundance throughout the genome (Chistiakov et al., 2006). Until recently, microsatellite markers have been developed in the *R. philippinarum* derived from both expressed sequence tag (EST) and anonymous genomic sequence (Yasuda et al., 2007; An et al., 2009; Nie et al., 2014). Molecular markers can be divided into type I (coding) markers which are associated with genes of known functions and type II (noncoding) markers which are associated with anonymous genomic sequences (O'Brien, 1991). As Type I markers that are associated with genes of known function, the EST-SSRs are superior to anonymous genomic SSR in functional diversity assessment and interspecific transferability (Pashley et al., 2006), but genomic SSRs usually are more polymorphic than EST-SSRs (Ellis and Burke, 2007).

About 60 microsatellite markers were developed in *R. philippinarum* until now (Yasuda et al., 2007; An et al., 2009; Hu et al., 2014), including 36 genomic SSRs and 25 expressed sequence tag derived SSRs (EST-SSRs). These markers provide sufficient information to evaluate wild and cultured genetic resources, but are still deficient for the development of genetic linkage map and marker-assisted selection (MAS). Therefore, the objective of the present study was to develop more polymorphic microsatellite markers using the transcriptome data derived from 454 pyrosequencing.

## MATERIALS AND METHODS

### Samples collection and DNA extraction

Thirty five individuals of *R. philippinarum* in total were collected from Jinzhou, Dalian, Liaoning province China. Genomic DNA of each specimen was isolated from muscle tissues following the standard phenol-chloroform method (Li et al., 2006) with some modifications. The adductor muscle was removed from fresh specimens and preserved in 100% ethanol until DNA preparation. Tissue was homogenized in 500 µL of extraction lyses buffer together with 0.5 µg/mL proteinase K and incubated at 55°C. Following phenol: chloroform: isoamyl alcohol (25:24:1) extractions, the supernatants were precipitated by the addition of 2 volumes of absolute ethanol. DNA was washed with 70% ethanol, dissolved in TE and stored at -20°C.

### Microsatellite primer design and PCR

Microsatellite sequences were screened from transcriptome sequences derived from *R. philippinarum* in our laboratory. Microsatellite sequences were screened from a total of 9450 ESTs in the 454 database using the software SSRHUNTER 1.3 (Li and Wan, 2005). Microsatellite primers were designed using Primer Premier 5.0 software (http: //www. premierbiosoft.com/ primerdesign/). From 9450 sequences, 324 were identified with microsatellite motifs, and a set of 105 microsatellite primer pairs were designed and synthesized. The major parameters for primer design were set as follows: primer length from 19 to 25 nucleotides, the size of PCR product from 100 to 350 bp, and annealing temperature at 50-65°C. The primers were synthesized by Sangon Company (Shanghai). ESTs containing SSRs were then annotated using BLAST software as described by Maneeruttanarungroj et al. (2006). The BLAST results were classified into 3 groups: known gene products, hypothetical proteins and unknown genes.

### Polymorphism assessment for primers

The polymorphisms of microsatellite primers were tested in 35 individuals of *R. philippinarum*. Of the 105 potential microsatellite markers, 39 were not easily amplified, and 38 were found to be polymorphic among 8 individuals of *R. philippinarum*. Then, thirty eight microsatellite markers were selected to test polymorphic and genetic diversity of natural population of *R. philippinarum* in Jin Zhou, Dalian, China.

Polymerase chain reaction (PCR) was performed in 10-µl volumes containing 0.5 U easy Taq DNA polymerase (TransGen, Beijing), 1× PCR buffer, 0.2 mM dNTP, 0.4 µM of each primer set, 1.5 mM MgCl2, and about 25 ng template DNA. The reactions were performed using the following parameters: 3 min at 94°C, followed by 35 cycles of 45 s at 94°C, 45 s at the annealing temperature listed in Table 1 and 45 s at 72°C, then a final extension of 5 min at 72°C. Amplification products were resolved on a 8% polyacrylamide gel and visualized by silver staining.

### Data analysis

The number of alleles, and observed ($H_O$) and expected ($H_E$) heterozygosities were estimated by MICROSATELLITE ANALYSER software (Dieringer and Schlötterer, 2003). Tests for linkage disequilibrium (LD) and deviations from Hardy-Weinberg equilibrium (HWE) were performed by GENEPOP 4.0 (Rousset, 2008). Sequential Bonferroni corrections (Rice, 1989) were applied for all multiple tests ($P < 0.05$). MICRO-CHECKER (Van Oosterhout et al., 2004) was employed to infer the most probable technical cause of HWE departures, including null alleles, mis-scoring due to stuttering and allelic dropout due to short allele dominance.

## RESULTS AND DISCUSSION

Microsatellites, which are inherited in a Mendelian fashion as codominant markers, have been increasingly popular in genetic studies because of their high level of heterozygosity, wide dispersal and abundance throughout the genome and transferability across different strains. For *R. philippinarum*, numerous microsatellites have been developed recently (Yasuda et al., 2007; An et al., 2009; Hu et al., 2014). But, the pace of development has been limited by the time-consuming and labor intensive requirement to construct, enrich and sequence genomic libraries (Edwards et al., 1996). Recently, identification of microsatellites from expressed sequences has been

**Table 1.** Characterization of thirty eight microsatellite markers in the Manila clam, *Ruditapes philippinarum*.

| Locus | GenBank accession No. | Repeat motif | Primer sequence (5'-3') | $T_a$ (°C) | Number of alleles | Size range (bp) | $H_O$ | $H_E$ | *P* value | Putative function | Primer location |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RpT238 | KP178950 | $(CAAA)_4$ | F:RCGCCTTCTGTGCTTTGATT | 60.0 | 8 | 259-273 | 0.6286 | 0.7023 | 0.0670 | Unknown | Unknown |
| | | | R:TGCGGCTGTGCGAAATAG | | | | | | | | Unknown |
| RpT190 | KP178935 | $(AT)_5$ | F:CGGTTGATAGGCTAATGC | 50.0 | 9 | 241-269 | 0.9394 | 0.8322 | 0.9679 | 4-aminobutyrate | CDS |
| | | | R:TTGCTGTTTGTGGGTTGT | | | | | | | aminotransferase | CDS |
| RpT171 | KP178929 | $(TG)_5$ | F:TAATTCCCAGTGGCAAGAT | 56.0 | 3 | 300-304 | 0.0000 | 0.5486 | 0.0000* | hypothetical protein | CDS |
| | | | R:CCTCAGCATTCCATTCAACT | | | | | | | | CDS |
| RpT209 | KP178944 | $(AT)_5$ | F:CTGTCTGCTTATTTGGTCGTT | 50.0 | 4 | 302-310 | 0.0370 | 0.5807 | 0.0000* | Unknown | Unknown |
| | | | R:CATTCGTCCTTGTGATACCTG | | | | | | | | Unknown |
| RpT196 | KP178938 | $(CA)_5$ | F:TGGTTGTAGCAGGTGTAGTTGT | 43.6 | 2 | 312-314 | 0.1250 | 0.1190 | 1.0000 | Unknown | Unknown |
| | | | R:GTTCTCGGATTCGTTGTTCT | | | | | | | | Unknown |
| RpT193 | KP178936 | $(TA)_6$ | F:TTGTTCGGAGTTGTCTGGT | 60.0 | 2 | 112-114 | 0.0313 | 0.1463 | 0.0044 | Unknown | Unknown |
| | | | R:TTCGTCAGTGTGAATCTTGC | | | | | | | | Unknown |
| RpT206 | KP178942 | $(AT)_5$ | F:TACCAACGCTCCTACAACTGAT | 60.0 | 2 | 190-192 | 0.4000 | 0.3578 | 0.8926 | hypothetical protein | CDS |
| | | | R:TCCCATTCACTTTCCAGCA | | | | | | | | CDS |
| RpT268 | KP178956 | $(TG)_5$ | F:GCAAATATGGTGCCTGAT | 50.0 | 3 | 270-306 | 0.1333 | 0.3367 | 0.0178 | Unknown | Unknown |
| | | | R:TAGCGTTCAGCTAACAAATC | | | | | | | | Unknown |
| RpT227 | KP178948 | $(TA)_5$ | F:TCAGGGACATGAATACGAC | 60.0 | 3 | 380-384 | 0.0357 | 0.6370 | 0.0000* | Unknown | Unknown |
| | | | R:AAATGATGCTGTCTGCTTG | | | | | | | | Unknown |
| RpT182 | KP178933 | $(TA)_5$ | F:CTTGTTATTGGAGGAGTGGT | 60.0 | 3 | 340-344 | 0.0000 | 0.5469 | 0.0000* | mitofusin 2 | CDS |
| | | | R:TTTACTGGTGGCAGAAGACT | | | | | | | | CDS |
| RpT241 | KP178951 | $(AT)_5$ | F:CAGGACTGGACCAAAGTG | 58.0 | 5 | 176-190 | 0.3438 | 0.4876 | 0.0380 | Unknown | Unknown |
| | | | R:AAGGCACCATAATATGTCAG | | | | | | | | Unknown |
| RpT195 | KP178937 | $(TA)_5$ | F:GCAGGGAATTTAATCAGC | 58.0 | 3 | 358-362 | 0.4242 | 0.5520 | 0.0151 | Unknown | Unknown |
| | | | R:AGTAAAACTTGGGACGGAT | | | | | | | | Unknown |
| RpT228 | KP178949 | $(AT)_5$ | F:AGTCTCGCTTTGACAGGA | 56.0 | 4 | 194-202 | 0.1212 | 0.5744 | 0.0001* | hypothetical protein | CDS |
| | | | R:CCCAAGAAGGGAGTTTATG | | | | | | | | CDS |
| RpT281 | KP178958 | $(AT)_5$ | F:GGTCCATTAACAGGCACAT | 56.0 | 3 | 274-284 | 0.1212 | 0.3170 | 0.0000* | Unknown | Unknown |
| | | | R:GAAAGCACGAGCAACCAT | | | | | | | | Unknown |
| RpT249 | KP178954 | $(AT)_5$ | F:TGAATCCAATGAAGTGCTG | 58.0 | 3 | 305-309 | 0.2000 | 0.4627 | 0.0020 | phosphate transport | Unknown |
| | | | R:CAGATGATGCTCCTGAACT | | | | | | | system permease protein | Unknown |
| RpT242 | KP178952 | $(TC)_5$ | F:CGCCTATTGCTGGATGTT | 48.0 | 3 | 290-294 | 0.2059 | 0.5158 | 0.0048 | Unknown | Unknown |
| | | | R:TGGAGGAAGACCGATTGAC | | | | | | | | Unknown |
| RpT282 | KP178959 | $(AT)_5$ | F:TGCATCGAGTATGGCAGAT | 50.0 | 3 | 170-174 | 0.0000 | 0.5397 | 0.0000* | Unknown | Unknown |
| | | | R:CAGGCTCCAAGTTTGTTTG | | | | | | | | Unknown |
| RpT208 | KP178943 | $(TA)_5$ | F:GACTAAAGTCATTGTGGGAAC | 50.0 | 5 | 402-410 | 0.2500 | 0.7158 | 0.0009* | hypothetical protein | CDS |
| | | | R:GACGCTGACATGGATTTGT | | | | | | | | CDS |
| RpT258 | KP178955 | $(TA)_5$ | F:TTGCATTCATTGCAGGAC | 48.0 | 3 | 301-305 | 0.0000 | 0.5669 | 0.0000* | Unknown | Unknown |
| | | | R:CAAGTAACCGTATGCGTGT | | | | | | | | Unknown |
| RpT214 | KP178945 | $(TG)_5$ | F:GGTAGCGTGACTCTTGGAT | 52.0 | 6 | 296-308 | 0.4688 | 0.5630 | 0.0450 | synaptotagmin-15-like | CDS |

**Table 1.** Contd.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | R:TGCCCTGTGAACTGTTTCT | | | | | | | | CDS |
| RpT183 | KP178934 | $(CT)_5$ | F:TGCCTGGTGTAAAGGTTGT | 55.5 | 2 | 309-311 | 0.0857 | 0.1346 | 0.1428 | ankyrin repeat | CDS |
| | | | R:CTAGCAGCCCAGATAAGTG | | | | | | | domain-containing protein | CDS |
| RpT197 | KP178939 | $(TA)_5$ | F:AGTGCGGTCTTAAAATCATC | 58.0 | 8 | 410-432 | 0.2727 | 0.6872 | 0.0000* | Unknown | Unknown |
| | | | R:AATACGGCTTGGTGTCAG | | | | | | | | Unknown |
| RpT223 | KP178946 | $(AT)_5$ | F:GTCAGGGACTTGGTCTTTCT | 58.0 | 5 | 361-369 | 0.5294 | 0.7164 | 0.0580 | tripartite motif-containing | CDS |
| | | | R:TCTTTAGCAACGGGATAGC | | | | | | | protein | CDS |
| RpT224 | KP178947 | $(CA)_5$ | F:GCACCGATACAGTCAATCTG | 58.0 | 2 | 390-392 | 0.0882 | 0.0856 | 1.0000 | hypothetical protein | CDS |
| | | | R:TGTAAGGGTTCAACAGGTTTC | | | | | | | | CDS |
| RpT201 | KP178941 | $(TA)_5$ | F:TTGCCCATCCATTGTCTC | 58.0 | 4 | 206-216 | 0.1429 | 0.7325 | 0.0000* | Unknown | Unknown |
| | | | R:GAAAGCCAATAAAGCTATCCAG | | | | | | | | Unknown |
| RpT244 | KP178953 | $(TA)_5$ | F:CCTTGCCACCAGTTTGAT | 58.0 | 3 | 184-190 | 0.0857 | 0.1362 | 0.0802 | Roquin-1, partial | CDS |
| | | | R:GCGGGATCTTGGTATTGA | | | | | | | | CDS |
| RpT274 | KP178957 | $(AC)_5$ | F:CGGTGGCGAGTATAAATAAATG | 56.0 | 10 | 148-178 | 0.3667 | 0.7316 | 0.0000* | hypothetical protein | CDS |
| | | | R:GAGGTGCAAACGGTAGAGATG | | | | | | | | CDS |
| RpT175 | KP178930 | $(AT)_5$ | F:TATCCAAGCAGTAGCGAAGT | 52.0 | 2 | 383-385 | 0.0667 | 0.4271 | 0.0000* | cysteine protease | CDS |
| | | | R:AAGAGTTTGCCGTTGTAGAG | | | | | | | ATG4A-like | CDS |
| RpT200 | KP178940 | $(TA)_5$ | F:GTGCTGCTTGCGATTTGT | 58.0 | 7 | 321-333 | 0.1935 | 0.7853 | 0.0000* | Unknown | Unknown |
| | | | R:TCATTTGCTCAACAGACCAAC | | | | | | | | Unknown |
| RpT169 | KT438736 | $(AT)_6$ | F:CAATTATCTGGCCTGTCA | 51.0 | 8 | 170-184 | 0.1875 | 0.7659 | 0.0000* | Unknown | Unknown |
| | | | R:TCTGGGTTCTTGCTGTCT | | | | | | | | Unknown |
| RpT150 | KP178926 | $(AT)_6$ | F:CCCCAGTTGTTCTTTGCT | 50.0 | 7 | 390-402 | 0.5588 | 0.7871 | 0.0046 | Unknown | Unknown |
| | | | R:CCTGTGAAGTTTGGAGGC | | | | | | | | Unknown |
| RpT37 | KP178923 | $(AC)_8$ | F:CAGATTTGACTGGCTTGG | 46.0 | 9 | 346-362 | 0.5625 | 0.8065 | 0.0238 | hypothetical protein | CDS |
| | | | R:TCCGATGAGAAACCCTTA | | | | | | | | CDS |
| RpT138 | KP178924 | $(GTC)_6$ | F:TCCAATGGCGACAACTAA | 38.0 | 7 | 346-360 | 0.2333 | 0.7638 | 0.0000* | lysosome-associated | CDS |
| | | | R:AGTCCCGAGGGTTCTTAT | | | | | | | membrane glycoprotein | CDS |
| Rpt178 | KP178931 | $(AAT)_5$ | F:GAATGTCCCGTTTCTATG | 41.0 | 9 | 340-356 | 0.4857 | 0.7979 | 0.0001* | Unknown | Unknown |
| | | | R:CAACAATCTAAGCCTCGT | | | | | | | | Unknown |
| RpT146 | KP178925 | $(TCA)_5$ | F:AGGCTTCATTCTCGTTAG | 33.0 | 8 | 446-460 | 0.8286 | 0.7375 | 0.8740 | Unknown | Unknown |
| | | | R:GTGGTGGATTTATGGATAT | | | | | | | | Unknown |
| RpT179 | KP178932 | $(TCA)_5$ | F:TAACGGAGGTAATGGACG | 38.1 | 2 | 180-182 | 0.0606 | 0.2163 | 0.0160 | hypothetical protein | CDS |
| | | | R:AATGATGCTGCTATGGGT | | | | | | | | CDS |
| RpT151 | KP178927 | $(AACT)_{26}$ | F:TCGGAGCAGATCACATGG | 38.1 | 6 | 148-160 | 0.3429 | 0.4437 | 0.1174 | Unknown | Unknown |
| | | | R:GAGATTGACGCTGACACG | | | | | | | | Unknown |
| RpT170 | KP178928 | $(TA)_5$ | F:AAACGCTCGTCTATCTCAG | 53.0 | 7 | 330-340 | 0.1212 | 0.7227 | 0.0000* | Unknown | Unknown |
| | | | R:CCACTTTAAGGCTTTCCA | | | | | | | | Unknown |

$T_a$: Annealing temperature of each primer pair, $N_a$: observed number of alleles, $H_O$: observed heterozygosity, $H_E$: expected heterozygosity. *Indicates significant departure from Hardy-Weinberg equilibrium after sequential Bonferroni correction ($P < 0.05/38$).

extensively used as an alternative strategy.

In addition to requiring less time and money to develop, expressed sequence tags (EST)-derived microsatellites have a number of intrinsic advantages. They tend to be more widely transferable between species, and even genera (Bouck and Vision, 2007). More importantly, because they represent genes, they serve as type I markers, which are more valuable for comparative gene mapping (Liu et al., 1999). In this study, a total of 324 microsatellite-containing sequences were identified from 9,450 transcriptome sequences in the *R. philippinarum* 454 database. Of the 324 sequences, 105 were selected for microsatellite marker optimization because of repetition times and flaking sequence priority. Of the 105 potential microsatellite markers, 39 were not easily amplified, 28 were monomorphic, and 38 were found to be polymorphic among 35 individuals of *R. philippinarum*.

GenBank (BLAST) searches indicated that 17 of the 38 EST-SSRs matched genes of known functions at E values less than $10^{-4}$, whereas the other 21 had no significant matches to known genes (Table 1). Of the 105 primer pairs developed, 38 microsatellite loci (36.2%) showed polymorphism in the population of *R. philippinarum* (Table 1). The 38 polymorphic loci (36.2%) yielded 2 to 10 alleles per locus with an average of 4.8 (Table 1). The observed and expected heterozygosities ranged from 0.000 to 0.939 and from 0.086 to 0.832, with an average of 0.255 and 0.542, respectively (Table 1). Significant linkage disequilibrium was detected between 12 pairs of loci (RpT282/RpT258, RpT209/RpT282, RpT193/RpT197, RpT208/RpT223, RpT268/RpT201, RpT200/RpT249, RpT182/RpT150, RpT175/RpT150, RpT238/RpT178, RpT178/RpT146, RpT138/RpT179 and RpT258/RpT170) ($P$ <0.01) before sequential Bonferroni correction for multiple tests (Rice, 1989); however, only 3 pairwise combinations of 6 loci (RpT209/RpT282, RpT178/RpT146 and RpT200/RpT249) were significant after Bonferroni correction.

Twenty loci conformed to HWE, while the remaining 18 loci showed significant deviation from HWE after Bonferroni correction at 5% significance level (Table 1). MICRO-CHECKER (Van Oosterhout et al., 2004) was used to estimate the most probable cause of departures from HWE. Micro-Checker analysis suggested that there was no evidence for scoring error due to stuttering and no evidence for large allele dropout. All the 18 loci were prone to null alleles ($P$ < 0.01). Widespread null alleles have been reported for approximately 51.9% of loci in the Pacific oyster (Li et al., 2003) and 52.2% of loci in the Zhikong scallop (Zhan et al., 2009). A high percentage of primers containing variable nucleotides, such as base substitutions or deletions at the PCR-priming sites in the flanking region of the microsatellites, may be responsible for the widespread appearance of null alleles in bivalves (Hedgecock et al., 2004).

The results obtained in this study indicated that these SSRs developed from EST in the Manila clam will be a useful tool for the genetic research such as population variation, parentage analysis, stock enhancement evaluation and the establishment of effective conservation strategy of *R. philippinarum*.

## Conflict of interests

The author has not declare any conflict of interest.

## ACKNOWLEDGEMENTS

This study was supported by the National Natural Science Foundation of China (31302183), the Program for Liaoning Excellent Talents in University (LJQ2014076), the Cultivation Plan for Youth Agricultural Science and Technology Innovative Talents of Liaoning Province (2014004), the Modern Agro-industry Technology Research System (CARS-48) and the National High Technology Research and Development Program (2012AA10A410-2).

### REFERENCES

Allendorf FW, Phelps SR (1980). Loss of genetic variation in a hatchery stock of cutthroat trout. Trans. Am. Fish. Soc. 109:537-543.

An HS, Kim EM, Park JY (2009). Isolation and characterization of microsatellite markers for the clam *Ruditapes philippinarum* and cross-species amplification with the clam *Ruditapes variegate*. Conserv. Genet. 10:1821-1823.

An HS, Park WJ, Cho KC, Han HS, Myeong J-I (2012). Genetic structure of Korean populations of the clam *Ruditapes philippinarum* inferred from microsatellite marker analysis. Biochem. Syst. Ecol. 44:186-195.

Bouck A, Vision T (2007). The molecular ecologist's guide to expressed sequence tags. Mol. Ecol. 16: 907-924.

Chistiakov DA, Hellemans B, Volckaert FAM (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255:1-29.

Dieringer D, Schlötterer C (2003). Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes 3:167-169.

Edwards KJ, Barker JH, Daly A, Jones C, Karp A (1996). Microsatellite libraries enriched for several microsatellite sequences in plant. Biotechniques 20:758-760.

Ellis J, Burke J (2007). EST-SSRs as a resource for population genetic analyses. Heredity 99:125-132.

Hedgecock D, Li G, Hubert S, Bucklin K, Ribes V (2004). Widespread null alleles and poor cross-species amplification of microsatellite DNA loci cloned from the Pacific oyster, *Crassostrea gigas*. J. Shellfish Res. 23:379-385.

Hu GW, Yan XW, Zhu DP, Nie HT (2014). Isolation and characterization of fourteen polymorphic microsatellite loci in the Manila clam (*Ruditapes philippinarum*). Conserv. Genet. Resour. 6:251-253.

Li G, Hubert S, Bucklin K, Ribes V, Hedgecock D (2003). Characterization of 79 microsatellite DNA markers in the Pacific oysters *Crassostrea gigas*. Mol. Ecol. Notes 3:228-232.

Li Q, Wan JM (2005). SSRHUNTER: development of a local searching software for SSR sites. Hereditas 27:808-810.

Li Q, Yu H, Yu RH (2006). Genetic variability assessed by microsatellites in cultured populations of the Pacific oyster (*Crassostrea gigas*) in China. Aquaculture 259:95-102.

Liu ZJ, Tan G, Li P, Dunham RA (1999). Transcribed dinucleotide microsatellites and their associated genes from channel catfish,

*Ictalurus punctatus*. Biochem. Biophys. Res. Commun. 259:190-194.
Maneeruttanarungroj C, Pongsomboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J, Li Y, WhanV, Chu KH, Li CP, Tong J, Glenn K, Rothschild M, Jerry D, Tassanakajon A (2006). Development of polymorphic expressed sequence tag-derived microsatellites for the extension of the genetic linkage map of the black tiger shrimp (*Penaeus monodon*). Anim. Genet. 37:363-368.
Mura L, Cossu P, Cannas A, Scarpa F, Sanna D, Dedola GL (2012). Genetic variability in the Sardinian population of the manila clam, *Ruditapes philippinarum*. Biochem. Syst. Ecol. 41:74-82.
Nie HT, Zhu DP, Yang F, Zhao LQ, Yan XW (2014). Development and characterization of EST-derived microsatellite makers for Manila clam (*Ruditapes philippinarum*). Conserv. Genet. Resour. 6:25-27.
O'Brien SJ (1991). Molecular genome mapping lessons and prospects. Curr. Opin. Genet. Dev. 1:105-111.
Pashley CH, Ellis JR, McCauley DE, Burke JM (2006). EST Databases as a source for molecular markers: Lessons from Helianthus. J. Hered. 97(4):381-388.
Rice RW (1989). Analyzing tables of statistical tests. Evolution 43:223-225.
Rousset F (2008). GENEPOP'O007: a complete reimplementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8:103-106.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes. 4:535–538.
Xing K, Gao ML, Li HJ (2014). Genetic differentiation between natural and hatchery populations of Manila clam (*Ruditapes philippinarum*) based on microsatellite markers. Genet. Mol. Res. 13(1):237-245.
Yasuda N, Nagai S, Yamaguchi S, Lian CL, Hamaguchi M (2007). Development of microsatellite markers for the Manila clam *Ruditapes philippinarum*. Mol. Ecol. Notes 7:43-45.
Zhan A, Hu J, Hu X, Hui M, Wang M, Peng W, Huang X, Wang S, Lu W, Sun C, Bao Z (2009) Construction of microsatellite-based linkage maps and identification of size-related quantitative trait loci for Zhikong scallop (*Chlamys farreri*). Anim. Genet. 40:821-831.
Zhang GF, Yan XW (2010). Clam aquaculture. Science Press, Beijing, China.

academicJournals

Vol. 15(1), pp. 7-19, 6 January, 2016
DOI: 10.5897/AJB2015.14878
Article Number: 7C928B856980
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

*Full Length Research Paper*

# Simple sequence repeats (SSR) and interspersed sequence repeats (ISSR) markers for genetic diversity analysis among selected genotypes of *Gossypium arboreum* race *'bengalense'*

**Khushboo Sethi[1], Priyanka Siwach[1]* and Surender Kumar Verma[2]**

[1]Deparment of Biotechnology, Chaudhary Devi Lal University, Sirsa, Haryana, India.
[2]Central Institute of Cotton Research, Regional Station, Sirsa, Haryana, India.

Received 22 July, 2015; Accepted 19 November, 2015

**Genetic diversity among 65 selected genotypes of *Gossypium arboreum* race bengalense was explored using 62 simple sequence repeats (SSR) and 73 interspersed sequence repeats (ISSR) markers. The SSR primers produced a total of 170 alleles (all polymorphic), while ISSRs yielded 281 bands of which only 94.3% were polymorphic. Utility of various markers were evaluated by calculating different parameters like polymorphic information content (PIC), marker index (MI), and discriminative ability (D), on the basis of which 21 SSR and 53 ISSRs primers were found very efficient for genetic diversity analysis. ISSR outperformed the SSR for discriminative ability as it yielded higher number of banding patterns (ISSR-658, SSR-175), greater numbers of polymorphic bands/assay (ISSR-3.63 and SSR-2.7) and higher D values (ISSR-0.862 and SSR-0.442). Values of I (SSR-0.740 and ISSR-0.421) and He (SSR-0.433 and ISSR-0.262) indicated SSRs as more suitable for characterizing the species in terms of abundance and evenness of alleles. A slight difference was observed in terms of MI values of the SSR (1.20) and ISSR (MI-1.38), showing an edge for ISSR in detecting overall polymorphism among given genotypes. Phylogenetic analysis was carried out by SSR, ISSR as well as combined datasets of markers. The highest value of cophenetic correlation coefficient was obtained for ISSR (r=0.94), followed by combined datasets (r=0.91) and SSR markers (r=0.87).**

**Key words:** Molecular markers, marker index, polymorphism information content, genetic diversity, *Gossypium arboreum,* discrimination coefficient.

## INTRODUCTION

Cotton (*Gossypium* spp.) is one of the principal cash crops, providing most of the world's natural textile fiber. The genus *Gossypium* (family *Malvaceae*) comprises nearly 45 diploid and 5 allotetraploid species. Spinnable

*Corresponding author. E-mail: Psiwach29@gmail.com

fibers are obtained only from four species; two allotetraploids or new world cotton (*Gossypium hirsutum* and *Gossypium barbedense*) and two diploids or Asiatic/old world cotton (*Gossypium herbaceum* and *Gossypium arboreum*).

India is the original home of domestication, diversification and development of Asiatic cultivated cottons. From 1500 BC to 1700 AD, India was recognized as the cradle of the cotton industry. The Indian monopoly in cotton muslins was broken up by the industrial revolution in England; new world cotton largely replaced the Asiatic cotton (Mohan et al., 2006). The major cause for this change was the unsuitability of diploid cotton fibers for mechanized spinning because of short length (<23 mm), high coarseness (>5.0 micronaire) and poor strength (<20 g/tex at 3.2 mm gauge) (Kulkarni et al., 2009). At present, tetraploid cotton (dominantly *G. hirsutum*) occupies a major fraction (>90%) of world cotton cultivation due to its suitability to mechanized harvesting and spinning. However, in marginal and drought-prone environments of Asia, diploid cottons are still popularly cultivated. This is because of certain inherent traits (which the tetraploids lack) like drought and salinity tolerance (Tahir et al., 2011); resistance to several pests including bollworms (Dhawan et al., 1991), aphids and leafhoppers (Nibouche et al., 2008); and diseases like rust, fungal (Wheeler et al., 1999) and viral (Akhtar et al., 2010).

Of the two diploid cultivated species, *G. arboreum* is more popular due to its suitability to a wider range of environments, and better fiber and plant features (Mohan et al., 2006). From its origin, dispersal and domestication of *G. arboreum* germplasm in different directions resulted in six races- indicum, burmanicum, cernuum, sinense, bengalense and soudanase. India is the only country where all six races are cultivated, the major share of which is contributed by 'bengalense' (cultivated commonly across central and North India).

*G. arboreum* germplasm constitutes an indispensable gene pool for modern cotton improvement programs. However, due to continuous selective breeding and selection during the last few decades, the germplasm is facing the constraints of narrow genetic base. Knowledge of genetic variation among *G. arboreum* germplasm is essential for future developments. Equally essential are the efficient tools which enable the detection of higher levels of genetic diversity (Ulloa et al., 2007). During the last two decades, various molecular markers have been extensively used for genetic diversity studies across species. *G. arboreum* germplasm has been explored with markers like randomly amplified polymorphic DNA (RAPD) (Deosarkar et al., 2010), interspersed sequence repeats (ISSR) (Bardak and Balek, 2012), simple sequence repeats (SSRs) (Noormohammadi et al., 2013a) etc.; and all studies report low polymorphism. Considering the edge of SSR and ISSR markers in cultivar fingerprinting and diversity studies, the present study was planned to evaluate the utility of these two methods for assessing genetic diversity as well as phylogenetic analysis among elite genotypes of *G. arboreum* race 'bengalense'.

## MATERIALS AND METHODS

### Plant materials and DNA extraction

Seeds from 65 elite genotypes belonging to race 'bengalense' of *G. arboreum* (Table 1) were procured from the Central Institute of Cotton research (CICR), Regional Station, Sirsa, Haryana, India. The cotton plants were cultivated in two rows of 6 m length with 30 cm interplant distance in the experimental field of CICR, Sirsa, in a completely randomized design (CRD) with three replications. Fresh and young leaves of randomly selected single plants of each genotype were subjected to total genomic DNA extraction using the cetyltrimethylammonium bromide (CTAB) method (Saghai et al., 1984) with certain modifications. The quality and quantity of extracted DNA was examined by agarose gel (0.8%) electrophoresis and ultra violet (UV)-spectrophotometry, respectively.

### SSR amplification

One hundred microsatellite primer pairs were obtained from Brookhaven National Laboratory (BNL), MGHES (M for Mississipi, GH- *G. hirsutum*, E- EST, S-SSR), CIR (CIRAD), JESPR (named after the names of Principal Investigators), Nanjing Agricultural University (NAU), and MUSS (M- Microsatellite, U- Last name of Principal Investigator, SS- Simple Sequences). Out of 100 primers, only 62 gave polymorphism and reproducible banding patterns and hence were selected for the present study (Table 2). The sequence information of these SSRs is available at http://www.cottonmarker.org.

Polymerase chain reaction (PCR) amplification was performed in a volume of 20 µl containing 2 µl of DNA (50 ng/µl), 0.5 µM of each primer (Sigma-Aldrich), 200 µM of dNTPs (Sigma-Aldrich), 0.5 U *Taq* polymerase (Sigma-Aldrich) and 1X PCR buffer (Sigma-Aldrich). Thirty five (35) cycles, each consisting of 1 min denaturation at 95°C, 2 min at annealing temperature (optimized separately for each primer pair, generally Tm-5°C) and 1 min polymerization at 72°C, were performed in a thermocycler (Bio-Rad, USA). The PCR products were separated by electrophoresis in a horizontal gel system at 100 V for 4 h in a 4% metaphor agarose gel. A 100 bp ladder (Thermo Scientific) was used for size determination of amplified products. Polymorphism was visualized by staining the gel with ethidium bromide, and it was photographed with the gel documentation system (Bio-Rad, USA).

### ISSR amplification

One hundred ISSR primers were used for initial screening, out of which 73 primers gave informative banding patterns with good reproducibility. The selected 73 primers were 15-20-mers which included 54.7% di-nucleotide repeat motif, 31.5% tri-nucleotide repeat motif, 8.21 % tetra-nucleotide repeat motif and 5.47% penta-nucleotide repeat motif (Table S1). These were anchored at 5' end or 3' end by zero nucleotides or by one to three partially degenerated selective nucleotides.

PCR amplification was performed in a volume of 20 µl containing 2 µl of DNA (50 ng/µl), 0.4 µM of each primer (Sigma-Aldrich), 200 µM of dNTPs (Sigma-Aldrich), 0.5 U Taq polymerase (Sigma-

**Table 1.** The selected genotypes of *G. arboreum*-race 'bengalense'

| Accession | Source | S/N | Accession | Source |
|---|---|---|---|---|
| CISA-6-187 | Sirsa | 34 | DLSA-1006 | Karnataka |
| CISA-6-123 | Sirsa | 35 | LD-960 | Punjab |
| CISA-6-209 | Sirsa | 36 | LD-909 | Punjab |
| CISA-6-214 | Sirsa | 37 | FDK-124 | Punjab |
| CISA-6-256 | Sirsa | 38 | PAIG-8/1 | Maharashtra |
| CISA-6-295 | Sirsa | 39 | DAS-802 | Karnataka |
| CISA-6-350 | Sirsa | 40 | CCA-4 | Tamilnadu |
| CISA-614 | Sirsa | 41 | RAAS-931 | Karnataka |
| CISA-6 | Sirsa | 42 | GBaV-105 | Gujarat |
| CISA-7 | Sirsa | 43 | GBaV-120 | Gujarat |
| CISA-8 | Sirsa | 44 | ARBHA-0853 | Karnataka |
| CISA-9 | Sirsa | 45 | ARBAS-104 | Karnataka |
| CISA-10 | Sirsa | 46 | RAAS-36 | Karnataka |
| CISA-294 | Sirsa | 47 | RAAS-8 | Karnataka |
| CISA-64 | Sirsa | 48 | GAM-158 | Gujarat |
| CISA-310 | Sirsa | 49 | AKA-0106 | Maharashtra |
| LD-327 | Punjab | 50 | CINA-369 | Maharashtra |
| LD-733 | Punjab | 51 | CAN-1006 | Maharashtra |
| ARBAS-105 | Karnataka | 52 | HD-485 | Hisar |
| TKA-9102/03 | Tamilnadu | 53 | GAM-150 | Gujarat |
| MDL-2617 | Karnataka | 54 | JTAPTI-007 | Madhya pradesh |
| GBaV-107 | Gujarat | 55 | CCA-8 | Tamilnadu |
| PA-532 | Maharashtra | 56 | LD-694 | Punjab |
| PA-686 | Maharashtra | 57 | RG-8 | Rajasthan |
| RG-526 | Rajasthan | 58 | HD-123 | Hisar |
| RG-540 | Rajasthan | 59 | PA-255 | Maharashtra |
| RG-541 | Rajasthan | 60 | LD-987 | Punjab |
| RG-514 | Rajasthan | 61 | RG-579 | Rajasthan |
| FDK-118 | Punjab | 62 | LD-919 | Punjab |
| TKA-9102 | Tamilnadu | 63 | LD-936 | Punjab |
| KWP-902 | Madhya pradesh | 64 | LD-1010 | Punjab |
| DLSA-17 | Karnataka | 65 | RG-595 | Rajasthan |
| DLSA-1005 | Karnataka | | | |

* All the 65 genotypes were collected from C.I.C.R, Regional station, Sirsa (Haryana), India, which in turn procured from respective source in India mentioned above.

**Table 2.** Different SSR and ISSR primers used for present study.

| Marker type | Names |
|---|---|
| SSR (The sequence information of these SSRs is available at http://www.cottonmarker.org.) | 1.BNL-530, 2.BNL-686, 3.BNL-852, 4.BNL-1694, 5.BNL-1679, 6.BNL-3408, 7.BNL-256, 8.BNL-1030, 9.BNL-834, 10.BNL1317, 11.BNL-1414, 12.BNL-252, 13.BNL-1053, 14.BNL-3279, 15.BNL-3649, 16.BNL-2960, 17.BNL-1707b, 18.BNL-448, 19.BNL-1434, 20.BNL-3140, 21.BNL-1231, 22.MGHES-13, 23.MGHES-14, 24.MGHES-32, 25.MGHES-58, 26.MGHES-70, 27.MGHES-7, 28.MGHES-52, 29.MGHES-50, 30.MGHES-46, 31.MGHES-45, 32.NAU-2035, 33.NAU-1047, 34.NAU-1231, 35.NAU-1068, 36.NAU-3675, 37.NAU-3519, 38.NAU-2317, 39.NAU-3008, 40.NAU-1218, 41.NAU-980, 42.NAU-2083, 43.NAU-862, 44.NAU-3418, 45.NAU-923, 46.NAU-1233, 47.NAU-3260, 48.MUSS-563, 49.MUSS422, 50.MUSS-257, 51.MUSS-300, 52.MUSS-321, 53.MUSS-88, 54.MUSS-20, 55.MUSS-49, 56.MUSS-121, 57.MUSS-439, 58.JESPR-127, 59.JESPR-65, 60.JESPR-307, 61.JESPR-297, 62.CIR-070 |
| ISSR | 73 Primers (The sequence information along with their annealing temperature is given in Supplementary Table 1) |

Aldrich) and 1X PCR buffer (Sigma-Aldrich). After a pre-denaturation step of 5 min at 95°C, amplification reactions were cycled forty times at 95°C for 1 min, at the annealing temperature (optimized separately for each primer pair, generally Tm-5°C) for 2 min and polymerization at 72°C for 1 min in a thermocycler (Bio-Rad, USA). The PCR products were visualized by running on 2% agarose gel, followed by staining with ethidium bromide. Finally, the gel was photographed as above.

### Data analysis

#### *Evaluating efficiency of different primers within each marker systems for diversity analysis*

Within each marker system, the efficiency of each assay unit (that is, primer) was studied by: a) the number of scorable bands (NSB); b) the number of polymorphic bands (NPB); c) polymorphism information content (PIC); d) marker index (MI); e) the number of patterns (Tp); and, f) discrimination power (D). The formulas used for the above calculations are as follows:

The number of scorable bands (NSB) represents the average number of DNA fragments amplified/detected per genotype using a marker system. Of these, some loci (fragments or bands) may be polymorphic (NPB).
PIC for SSR markers was calculated according to Anderson et al. (1993). For ISSR markers, PIC of a band (PIC*i*) was calculated as follow: $\mathbf{PIC}_i = 1 - \sum_j f_{ij}^{\ 2}$, where *fij* is the frequency of the $j^{th}$ pattern of the $i^{th}$ band (note that dominant markers have two patterns for a band as being present and absent). Then PIC of each ISSR primer was calculated as: $\mathbf{PIC} = 1/n \sum_{i=1}^{n} \mathbf{PIC}_i$, where *n* is NPB for that primer.

The utility of a given marker system is a balance between the level of polymorphism detected and the extent to which an assay can identify multiple polymorphisms. Marker index is the product of PIC and effective multiplex ratio (EMR) (Powell et al., 1996). EMR is estimated as: EMR= NSB X ß, where ß is the fraction of polymorphic markers and is estimated after considering the polymorphic loci (np) and non-polymorphic loci (nnp) as ß = np / (np + nnp). Tp and D were calculated according to Tessier et al. (1999).

#### *Comparison of two marker systems for diversity analysis*

To compare the discriminating capacity of the SSR and ISSR markers, the following statistical calculations were performed manually according to Belaj et al. (2003): a) the number of assay units (U); b) the number of polymorphic bands ($n_p$); c) the number of monomorphic bands ($n_{np}$); d) the average number of polymorphic bands/assay unit ($n_p$/U); e) the number of Loci (L); f) number of loci/assay unit ($n_u$); g) the number of banding patterns ($T_p$); h) the average number of patterns/assay unit (I); i) average confusion probability (C); j) average discriminating power (D); and k), the average limit of discriminating power ($D_L$).

Several other genetic diversity parameters viz. effective number of allele ($N_e$), Shannons index (I) and expected heterozygosity ($H_e$) were determined using GenAlex 6.5.

#### *Cluster analysis*

For this analysis, each amplified band was treated in terms of binary code, based on the presence (1) and absence (0) of bands. To analyze data obtained from binary matrices, the NTSYS-pc ver 2.2 statistical package (Rohlf, 2000) was used. Three data sets were utilized, viz. SSR, ISSR and combined datasets of SSR and ISSR. The binary qualitative data matrices were then used to construct similarity matrices based on Jaccard similarity coefficients (Jaccard, 1908). The similarity matrices were then used to construct a dendrogram using the unweighted pair group method with arithmetic average (UPGMA). To compare SSR and ISSR based dendrograms, cophenetic matrices were derived from dendrograms using COPH (cophenetic values) program, and the goodness-of-fit of the clustering to the 2 data matrices was calculated by comparing the original similarity matrices with the cophenetic value matrices using the Mantel matrix correspondence test (Mantel, 1967) in the MXCOMP program. Similarly, a dendrogram was also constructed for combined dataset of SSR and ISSR markers.

## RESULTS

### SSR analysis

The 62 primers detected a total of 170 alleles (all polymorphic), and the number of alleles per locus varied from 2 to 6 with an average of 2.7 alleles per locus$_T$ in all 65 genotypes examined. The size of the alleles ranged from a minimum of 90 bp (loci JESPR-297) to maximum of 720 bp (loci MGHES-14). The PIC values ranged from 0.030 (MUSS-439, NAU-923, NAU-3675, BNL-1434, BNL-1694) to 0.809 (NAU-3008), giving an average of 0.38. Of the 62 SSR loci, 21 loci yielded a PIC value of ≥ 0.5 (Table 3) and produced quite distinct bands in the metaphor gels (Figure S1). These included 5 BNLs, 3 MGHES, 6 NAUs and 6 MUSS and 1 JESPR SSR loci, which were considered as highly informative markers (Table 3). Among these 21 loci, 52.1% had di-nucleotide motifs, 30.4% had tri-nucleotide motifs, while the remaining 17.5% had tetra/penta/hexa-nucleotide motifs.

Marker index (MI), considered to be an overall measure of the efficiency to detect polymorphism, was obtained in the range of 0.06-4.85 (average 1.20). The 21 informative primers, designated so on the basis of high PIC, also exhibited a high marker index value (more than 1.5). Primer NAU-3008 yielded highest MI value (4.85), which was obvious because it had the highest PIC and EMR.
The discriminating power (D) of a primer depends on the number of fragments it generates as well as the frequency of the banding patterns. In the present study, the maximum value of discrimination power (D) observed was 0.927 (NAU-3008) while the lowest was 0.44 (NAU-3675, BNL 1434, BNL 1694, NAU 923 and MUSS 439); with an overall mean value of 0.442 for the 62 SSR loci. The 21 above-mentioned informative primers pairs also exhibited high discrimination power (values of D more than 0.6) and thus these 21 primers were categorized as highly informative and discriminative primers (Table 3).

### ISSR analysis

Across the 65 genotypes, the 73 ISSR primer pairs yielded a total of 281 reproducible bands, of which 265

**Table 3.** Description of 21 selected SSR markers for all the studied genotypes of *G. arboretum.*

| Primer name | NSB | NPB | PIC | MI | Tp | D |
|---|---|---|---|---|---|---|
| BNL-834 | 3 | 3 | 0.51 | 1.53 | 4 | 0.66 |
| BNL-252 | 3 | 3 | 0.55 | 1.65 | 5 | 0.77 |
| BNL-2960 | 4 | 4 | 0.57 | 2.31 | 4 | 0.78 |
| BNL-448 | 3 | 3 | 0.58 | 1.74 | 3 | 0.67 |
| BNL-3140 | 4 | 4 | 0.61 | 2.45 | 4 | 0.7 |
| MGHES-50 | 4 | 4 | 0.68 | 2.74 | 4 | 0.74 |
| MGHES-46 | 3 | 3 | 0.59 | 1.77 | 3 | 0.69 |
| MGHES-45 | 3 | 3 | 0.57 | 1.73 | 3 | 0.69 |
| NAU-3519 | 3 | 3 | 0.51 | 1.55 | 3 | 0.60 |
| NAU-2317 | 3 | 3 | 0.57 | 1.73 | 3 | 0.68 |
| NAU-3008 | 6 | 6 | 0.80 | 4.85 | 9 | 0.92 |
| NAU-1218 | 3 | 3 | 0.58 | 1.75 | 3 | 0.67 |
| NAU-3418 | 4 | 4 | 0.64 | 2.56 | 6 | 0.82 |
| NAU-3260 | 4 | 4 | 0.70 | 2.8 | 4 | 0.75 |
| MUSS-257 | 3 | 3 | 0.57 | 1.71 | 3 | 0.66 |
| MUSS-300 | 3 | 3 | 0.58 | 1.74 | 3 | 0.67 |
| MUSS-321 | 3 | 3 | 0.55 | 1.67 | 3 | 0.65 |
| MUSS-20 | 4 | 4 | 0.69 | 2.77 | 4 | 0.77 |
| MUSS-49 | 3 | 3 | 0.57 | 1.73 | 3 | 0.72 |
| MUSS-121 | 4 | 4 | 0.70 | 2.8 | 4 | 0.76 |
| JESPR-65 | 4 | 4 | 0.70 | 2.8 | 4 | 0.79 |

NSB, Number of scorable bands; NPB, number of polymorphic bands (NPB); PIC, polymorphic information content; MI, marker index; Tp, number of banding patterns; D, discriminative ability.

(94.3%) were polymorphic. The number of loci (or bands) scored varied from 2 (ISSR-17, 19, 38, 47, 49, 59, 62, 69, 74, 81, 82, 84, 87, 96 and 103) to 15 (ISSR-18), with an average of 3.84 bands/loci per primer. The PIC values for 73 ISSR primer pairs ranged from 0.3 to 0.5, with an average of 0.38 per primer; and the highest PIC (0.5) was obtained for ISSR-82. Marker index (MI), calculated for each primer pair, was found in the range of 0.16 (ISSR-17, 19, 49, 81) to 5.28 (ISSR-18). The highest value of discrimination power was observed for primer ISSR-18 (0.998), while primer ISSR-17, 19, 49 and 81 yielded the lowest values of D (0.316).

On the basis of higher values of MI (more than 1), D (more than 0.8) and PIC (more than 0.3), 53 primer pairs were identified as very efficient for the present genetic diversity analysis (Table 4). Further, in addition to these 53 primer pairs, 7 more primer pairs viz. ISSR-15, 16, 59, 80, 87, 96 and 103 exhibited higher values of D (more than 0.8), though MI values were considerably low for some.

## Comparison of marker systems

Performance of the two marker systems was compared based on two main aspects: The discriminating capacity (that is, efficiency of discrimination) between any two genotypes at random from the studied genotypes; and, the overall efficiency in detecting polymorphisms in all the studied genotypes.

Overall, SSR markers were more polymorphic (100% polymorphic bands) than ISSR (94.3% polymorphic bands), however, the number of polymorphic bands per assay unit was higher in ISSR (3.63) as compared to SSR (2.7). SSR markers are locus specific so only 1 loci was analyzed per assay, and 62 loci overall. ISSR primer pairs produced 281 bands, with each band considered as one locus, resulting in an average of 3.84 loci per assay unit. ISSR produced a higher number of banding patterns (658) than SSR (175) and so the average number of banding pattern per assay unit was also higher for ISSR (9.01) than for SSR (2.8).

The number of effective alleles (Na) in all 65 genotypes examined was higher in SSR (2.112) than in the ISSR assay (1.397), while the average discriminating capacity (D) was distinctly higher for ISSR (0.862), compared to SSR (0.442) (Figure 1). The average limits of discriminating powers ($D_L$) for both the markers were found to be very close to the actual value of the discriminating powers (D) of both.

A higher value for the Shannon index (I) was obtained for SSR (0.740), ISSR yielding a comparatively low value of I (0.421) (Figure 1). The average expected hetero-zygosity (He) values calculated for SSR and ISSR came

**Table 4.** Description of 53 ISSR markers for all the studied genotypes of *G. arboretum.*

| Name | NSB | NPB | PIC | MI | Tp | D | Name | NSB | NPB | PIC | MI | Tp | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ISSR1 | 3 | 3 | 0.45 | 1.35 | 7 | 0.89 | ISSR-70 | 3 | 3 | 0.43 | 1.29 | 7 | 0.94 |
| ISSR-2 | 4 | 4 | 0.38 | 1.52 | 12 | 0.96 | ISSR-71 | 4 | 4 | 0.40 | 1.6 | 12 | 0.96 |
| ISSR-3 | 4 | 4 | 0.31 | 1.24 | 10 | 0.94 | ISSR-72 | 4 | 4 | 0.40 | 1.6 | 12 | 0.96 |
| ISSR-4 | 3 | 3 | 0.34 | 1.02 | 6 | 0.92 | ISSR-73 | 4 | 4 | 0.37 | 1.48 | 9 | 0.95 |
| ISSR-5 | 7 | 7 | 0.32 | 2.24 | 17 | 0.90 | ISSR-75 | 4 | 4 | 0.35 | 1.4 | 11 | 0.89 |
| ISSR-6 | 3 | 3 | 0.43 | 1.29 | 7 | 0.93 | ISSR-76 | 3 | 3 | 0.41 | 1.23 | 7 | 0.94 |
| ISSR-7 | 4 | 4 | 0.33 | 1.32 | 11 | 0.95 | ISSR-77 | 4 | 4 | 0.41 | 1.64 | 11 | 0.95 |
| ISSR-10 | 4 | 4 | 0.42 | 1.68 | 14 | 0.97 | ISSR-78 | 3 | 3 | 0.45 | 1.35 | 7 | 0.94 |
| ISSR-18 | 15 | 13 | 0.41 | 5.28 | 33 | 0.99 | ISSR-83 | 4 | 4 | 0.43 | 1.72 | 13 | 0.96 |
| ISSR-27 | 5 | 5 | 0.30 | 1.5 | 14 | 0.98 | ISSR-85 | 3 | 3 | 0.42 | 1.26 | 7 | 0.92 |
| ISSR-28 | 3 | 3 | 0.38 | 1.14 | 5 | 0.94 | ISSR-86 | 4 | 4 | 0.42 | 1.68 | 7 | 0.92 |
| ISSR-29 | 8 | 8 | 0.37 | 2.96 | 26 | 0.97 | ISSR-88 | 4 | 4 | 0.41 | 1.64 | 10 | 0.93 |
| ISSR-31 | 12 | 10 | 0.38 | 3.78 | 32 | 0.95 | ISSR-89 | 3 | 3 | 0.42 | 1.26 | 7 | 0.93 |
| ISSR-34 | 5 | 5 | 0.31 | 1.55 | 13 | 0.96 | ISSR-90 | 4 | 4 | 0.43 | 1.72 | 7 | 0.92 |
| ISSR-35 | 9 | 9 | 0.30 | 2.7 | 27 | 0.87 | ISSR-91 | 3 | 3 | 0.42 | 1.26 | 6 | 0.91 |
| ISSR-36 | 4 | 4 | 0.39 | 1.56 | 12 | 0.93 | ISSR-92 | 4 | 4 | 0.42 | 1.68 | 6 | 0.92 |
| ISSR-40 | 9 | 8 | 0.32 | 2.53 | 24 | 0.93 | ISSR-93 | 5 | 5 | 0.42 | 2.10 | 6 | 0.91 |
| ISSR-45 | 3 | 3 | 0.34 | 1.02 | 7 | 0.94 | ISSR-94 | 3 | 3 | 0.43 | 1.29 | 6 | 0.92 |
| ISSR-50 | 4 | 4 | 0.30 | 1.2 | 12 | 0.98 | ISSR-95 | 4 | 4 | 0.43 | 1.72 | 6 | 0.90 |
| ISSR-60 | 4 | 4 | 0.34 | 1.36 | 12 | 0.97 | ISSR-97 | 3 | 3 | 0.43 | 1.29 | 7 | 0.93 |
| ISSR-61 | 4 | 4 | 0.41 | 1.64 | 13 | 0.95 | ISSR-98 | 3 | 3 | 0.43 | 1.29 | 7 | 0.93 |
| ISSR-63 | 4 | 4 | 0.37 | 1.48 | 11 | 0.97 | ISSR-99 | 4 | 4 | 0.43 | 1.72 | 7 | 0.93 |
| ISSR-64 | 4 | 4 | 0.39 | 1.56 | 10 | 0.92 | ISSR-100 | 3 | 3 | 0.43 | 1.29 | 7 | 0.92 |
| ISSR-65 | 4 | 4 | 0.30 | 1.2 | 10 | 0.96 | ISSR-101 | 5 | 5 | 0.43 | 2.15 | 7 | 0.93 |
| ISSR-66 | 3 | 3 | 0.40 | 1.2 | 7 | 0.94 | ISSR-102 | 3 | 3 | 0.43 | 1.29 | 7 | 0.93 |
| ISSR-67 | 5 | 5 | 0.40 | 2.0 | 20 | 0.98 | ISSR-104 | 3 | 3 | 0.43 | 1.29 | 7 | 0.93 |
| ISSR-68 | 4 | 4 | 0.40 | 1.6 | 13 | 0.96 | | | | | | | |

NSB, Number of scorable bands; NPB, number of polymorphic bands (NPB); PIC, polymorphic information content; MI, marker index; Tp, number of banding patterns; D, discriminative ability.

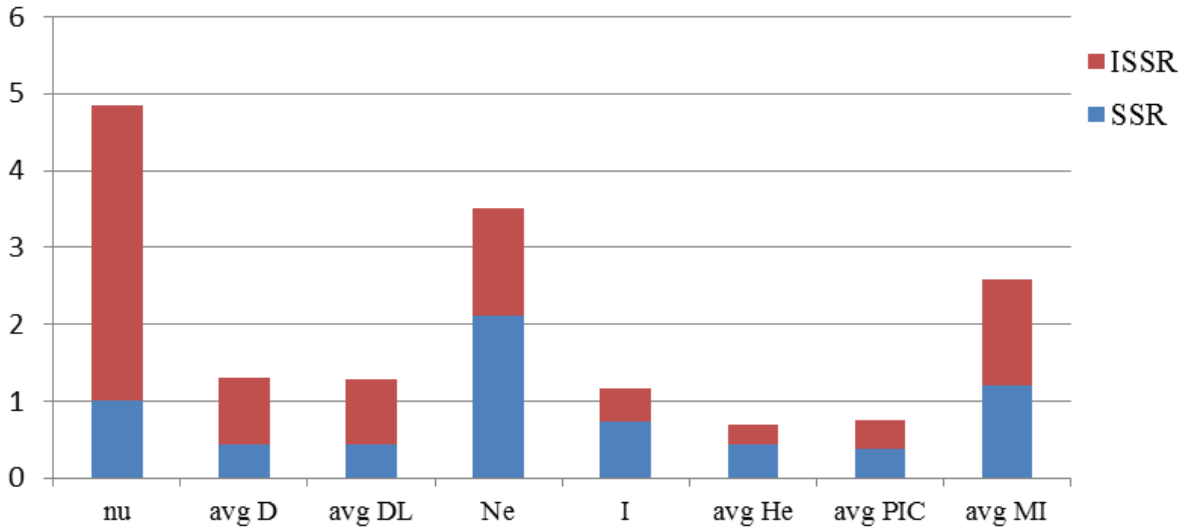


**Figure 1.** Comparative description of SSRs and ISSRs for parameters-number of loci/assay unit (nu), average discriminating power (avg D), average limit of discrimination power (avgDL), number of effective alleles (Ne), Shannon's index (I), average expected heterozygosity (avg He), average PIC (avg PIC), average MI (avg MI).

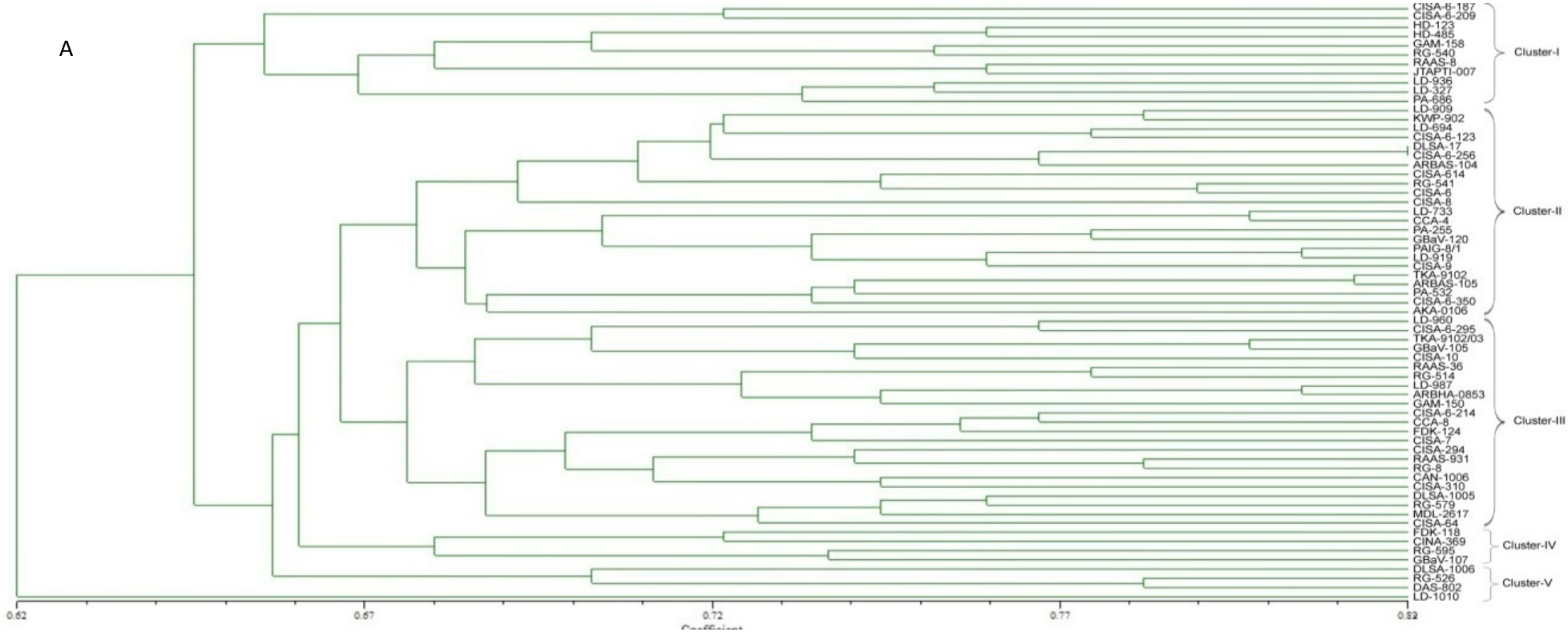


**Figure 2.** Dendrograms of 65 *G. arboreum* genotypes based on the similarity coefficient values calculated using: **(a)** SSRs, **(b)** ISSRs and **(c)** joined data set of markers.

out to be 0.433 and 0.264 respectively. The average PIC was found to be the same for both SSR and ISSR markers (0.38) in the studied genotypes, while the Average MI was slightly higher for ISSR compared to SSR markers.

## Cluster analysis

A dendrogram obtained using the UPGMA method based on SSR, ISSR and SSR + ISSR data set (Figure 2) clearly distinguished all the genotypes of the race 'bengalense' of *G. arboreum*. Genetic similarity coefficients were obtained in the range of 0.62-0.82 for the SSR marker, 0.56-0.86 for ISSR markers and 0.59-0.80 for the combined data of SSR and ISSR markers. Five main clusters were formed in all three dendrograms. Each cluster consists of a different number of genotypes with different genetic similarity coefficients.

In dendrograms based on SSR, ISSR and SSR + ISSR, the first cluster consists of 11, 14 and 9 genotypes, respectively, in which CISA-6- 187 have been found to be more distant than the other genotypes in all three dendrograms. The second cluster consists of 23, 21 and 21 genotypes, respectively, showing almost similar groupings of genotypes but with some differences in the similarity coefficient between different genotypes. For example, with SSR markers, DLSA-17 and CISA-6-256 exhibited a maximum similarity coefficient value of 0.82, while with ISSR the maximum value (0.86) was for CISA-6 and CISA-8. For combined datasets, a maximum similarity coefficient (0.785) within cluster 2 was obtained

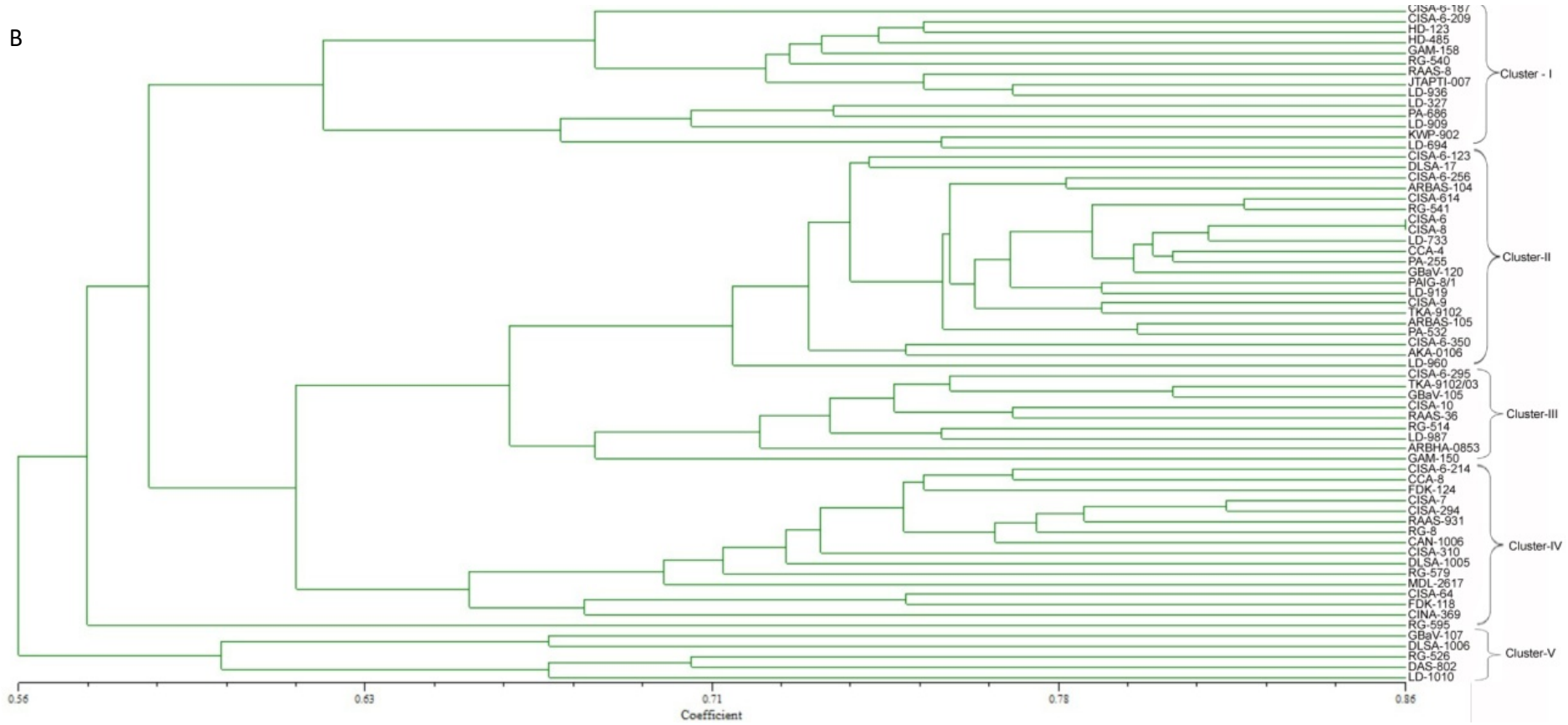


**Figure 2.** Contd.

for CISA-614 and RG-541. The third cluster of SSR-based dendrograms consisted of 23 genotypes, while in dendrograms based on ISSR and combined data, the third cluster consisted of nine genotypes each. Similar observations were made for cluster four, which consisted of 4, 16 and 19 genotypes in SSR, ISSR and SSR+ISSR based dendrograms, respectively; CISA-7 and CISA-294 were found to be closer than the rest of the genotypes in the case of ISSR and SSR + ISSR dendrograms; but, in the case of SSR, these two genotypes were present in cluster 3. Cluster five consisted of almost similar number of genotypes, that is, 4, 5 and 7, in the three dendrograms formed. In this cluster, LD-1010 was found to be more distant than the rest of genotypes in all three of the dendrograms obtained.

Cophenetic correlation coefficients for individual techniques based on genetic similarity value matrices were obtained using the Mantel matrix correspondence test. High correlation coefficient values were obtained for ISSR markers (r = 0.94), for combined data set (SSR + ISSR) marker (r = 0.91) and for SSR markers (r = 0.87). All three dendrograms showed almost similar groupings with some differences in the genetic similarity

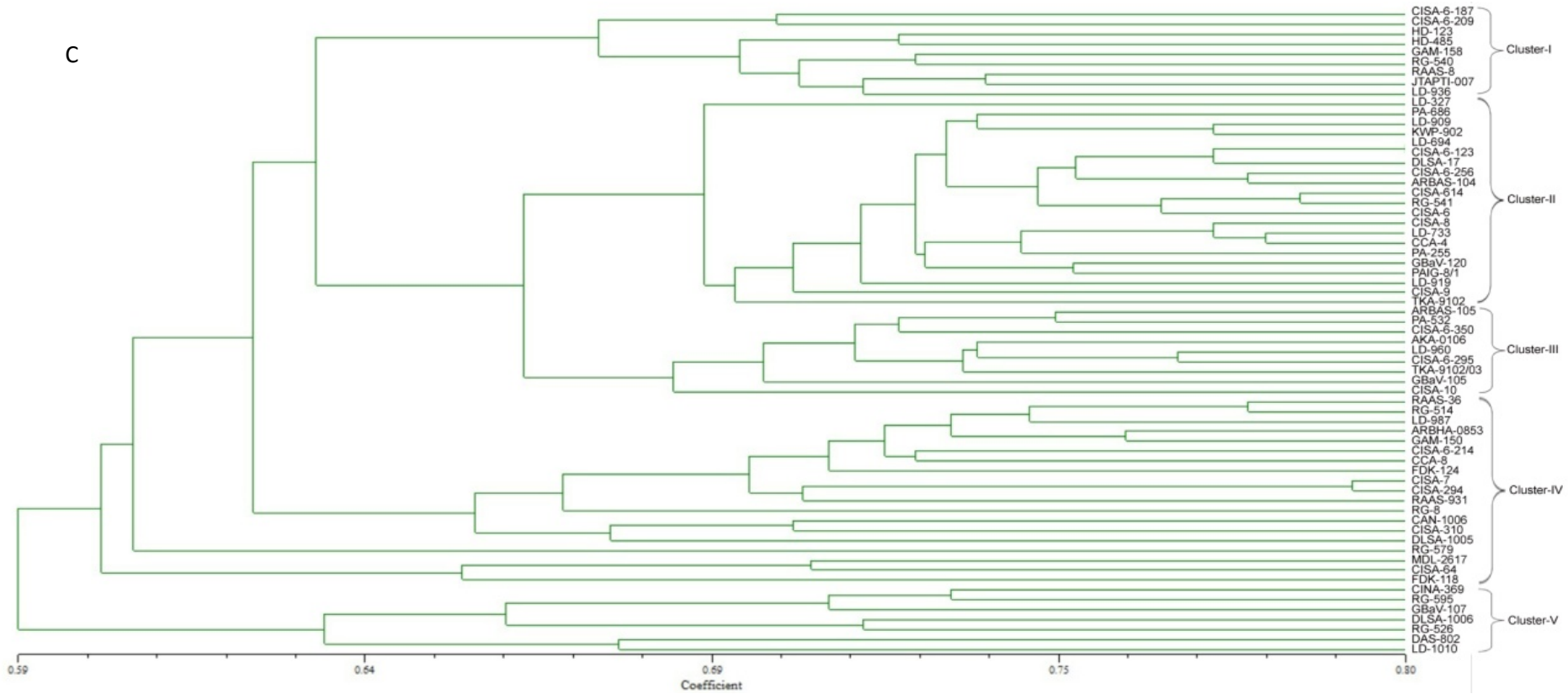


**Figure 2.** Contd.

coefficient, as discussed above.

## DISCUSSION

During the past few decades, molecular markers have been commonly used for assessing genetic diversity, which is the basis for the genetic improvement of any given species. The important criteria of selecting the right molecular marker depends on the specific application, presumed level of polymorphism, presence of sufficient technical facilities, time constraints and financial limitations (Kumar et al., 2009). Sometimes the combined use of two or more markers for the study of genetic diversity has been found to be better than respective individual markers (Anna Serra et al., 2007). In the past, a variety of molecular markers like RAPD, ISSR and SSR have been used for estimating the genetic diversity in *G. arboreum* (Dongre et al., 2011; Bardak and Bolek, 2012; Noormohammadi et al., 2013a). SSR are locus specific, co-dominant markers, and are considered ideal for fingerprinting; while ISSR are multi-locus, dominant markers, and have been found very efficient for diversity analysis. Therefore, the present study documents the comparative utility of these maker types for genetic diversity studies in accessions belonging to *G. arboreum* race bengalense.

**Marker polymorphism**

Both SSR and ISSR markers were found to reveal a similar level of polymorphism as revealed by the same average value of PIC (0.38) obtained for each. The average PIC value for SSR markers obtained during the present study was less than that obtained by Kantartzi et al. (2009) (average PIC 0.42) while genotyping various *G. arboreum* genotypes with SSR markers, though their highest PIC obtained was less (0.75) than that obtained during the present study (0.809). ISSRs are dominant markers and therefore a maximum PIC value of 0.50 can be expected for a given ISSR loci. During the present investigation, for one marker ISSR-82, this threshold was reached while values very close to the threshold were obtained for ISSR-1, ISSR-62, ISSR-69, and ISSR-78 (0.45, 0.45, 0.47 and 0.45, respectively). A PIC range of 0.00 to 0.5 with an average of 0.321 was also obtained previously in another study using ISSR markers for some tetraploid cotton (Noormohammadi et al., 2013b).

In addition to PIC, certain other parameters such as MI and D have been documented as very useful for evaluating the efficiency of molecular markers (Belaj et al., 2003; Myskow et al., 2010). The utility of any given marker is found in a balance between the level of polymorphism it can detect and its capacity to identify multiple polymorphisms (Powell et al., 1996). The MI is considered to be an overall measure of the efficiency of a marker to detect polymorphism, and is related to EMR value. Discriminating power is considered as a good estimator of the efficiency of a primer or locus. It depends not only on the number of patterns generated, but also on their relative frequency (Tessier et al., 1999). On the basis of these factors, a core set of 21 SSR primers (Table 3) were identified as highly informative markers with high PIC, very good discriminative power and MI. Likewise, 54 ISSR primer pairs could be identified on the basis of higher MI values. Multi locus marker systems like ISSR are expected to produce higher EMR and MI than single locus SSRs (Belaj et al., 2003). Markers with higher EMR and MI values are better for analysis of both interspecific and intraspecific genetic diversity (Singh et al., 2014). Several studies report such identification of a core set of highly polymorphic and discriminative markers to be very helpful for varietal identification and genetic diversity assessment (Masi et al., 2003; Jain et al., 2004; Kantartzi et al., 2009).

**Comparative utility of marker system**

The selection of a particular type of molecular marker is important and critically depends on the intended use (Gupta et al., 2002). The discriminative abilities of both marker systems were compared using certain selected parameters which have also been used earlier for such purposes in some studies (Mukherjee et al., 2013). The presence of rare bands/alleles can produce low frequency of patterns and result in lower D values. ISSR markers exhibited considerably higher number of banding patterns, more polymorphic bands/assays and higher discriminative powers compared to SSR during the present investigation. The similar edge of ISSR over SSR in terms of discriminative capability for a given set of genotypes has also been observed in certain other studies (Singh et al., 2014).

SSR markers are locus specific, multi-allelic and co-dominant in nature. These have been found to detect higher levels of polymorphism and so, generally, are the markers of choice in plant genetics and breeding (Kantartzi et al., 2009). ISSR are bi-allelic (hence supposed to be less informative) and are locus un-specific, but are more randomly distributed throughout genome than SSR (Kumar et al., 2009). This abundance of ISSR sometimes compensates for their bi-allelic nature and may make them very informative for a given germplasm (Vijayan, 2005). Further, the low development and running cost makes ISSR more suitable than SSR (Vijayan, 2005).

During the present study, SSR markers outperformed the ISSR in terms of Ne, I and He parameters. Ne represents the number of equally frequent alleles it would take to achieve a given level of gene diversity. The Shannon index (I) is a diversity index that is used to characterize species diversity and is an indicator of both the abundance and evenness of the species present. The reason for high heterozygosity in case of SSR markers is due to its co-dominant nature, which permits the detection of a high number of alleles per locus as these are multi-allelic as compared to ISSR markers, which are biallelic in nature (Belaj et al., 2003).

During the present study, the average PIC value for SSR was on the lower side (0.38) as SSR, being co-dominant, yielded PIC values in the range of 0 to 1.0. On the other hand, ISSR markers yielded a higher value of average PIC (0.38), while for dominant markers the range is 0 to 0.5. Further, ISSR also showed better utility in detecting multiple polymorphisms as revealed by high MI and high EMR (Table 4).

**Phylogenetic relationships in examined germplasm**

The present study has reported that both SSR and ISSR techniques, along with proper statistical tools, could be successfully applied to assess genetic diversity and perform phylogenetic analysis in *G. arboreum.* Although SSR and ISSR markers showed differences in detecting polymorphism and discriminating capacity, they showed similar groupings in dendrograms on the basis of similarity matrices. A high significant correlation coefficient was obtained for all the three dendrograms. The correlation coefficient between genetic similarity values depends not only on the kind of molecular technique and species examined, but also upon the range of discovered diversity. Noormohammadi et al. (2013b) and Sheidai et al. (2012) reported higher values (r=0.87-95) by using

different molecular techniques in cotton. High *r* values and identical topologies of dendrograms suggest that each method of molecular marker development, used independently, could be a reliable source of information about the relationships between analyzed germplasm (Myskow et al., 2010). In our study, ISSR and SSR+ISSR markers depicted better topology and high correlation coefficient than SSR markers.

In conclusion, although the average PIC is the same for both markers, there are certain parameters in which SSR exceeds like Ne, I and He, and in the rest of the parameters - MI, EMR, and D - ISSR was found better than SSR. So, a combination of both markers would be highly efficient in detecting genetic diversity and phylo-genetic analysis between genotypes of race 'bengalense' of *G. arboreum*.

## Conflict of interests

The authors have not declared any conflict of interest.

## Abbreviations

**ISSR,** Interspersed sequence repeats; **SSR,** simple sequence repeats; **PIC,** polymorphic information content.

## REFERENCES


Akhtar KP, Haidar S, Khan MKR, Ahmad M, Sarwar N (2010). Evaluation of *Gossypium* species for resistance to cotton leaf curl Burewela virus. Ann. Appl. Biol. 157:135-147.

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorells ME (1993). Optimizing parental selection for genetic linkage maps. Genome 36:181-186.

Anna Serra I, Procaccini G, Carmela Intrieri M, Migliaccio M, Mazzuca S, Maria Innocenti A (2007). Comparison of ISSR and SSR markers for analysis of genetic diversity in sea grass (*Posidonia aceanica*). Mar. Ecol. Prog. Ser. 338:71-79.

Bardak A, Bolek Y (2012). Genetic diversity of diploid and tetraploid cottons by SSR and ISSR markers. Turk. J. Field Crops 17:139-144.

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I (2003). Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theor. Appl. Genet. 107:736-744.

Deosarkar VD, Gupta DB, Gaikwad DB (2010). Genetic diversity studies in intra-specific *desi* cotton (*G. arboreum*) through DNA marker. J. Cotton Res. Dev. 24:133-137.

Dhawan AK, Simwat GS, Sidhu AS (1991). Field reaction of some varieties of Asiatic Cotton (*Gossypium arboreum* L.) to sucking and boll worm pests. J. Res. Punjab Agric. Univ. 28: 57-62.

Dongre AB, Ramteke DM, Bhandarkar MR, Raut MP, Meshram KJ (2011). Genetic diversity analysis of *Gossypium arboreum* (diploid cotton) cultivars revealed by PCR based molecular markers. J. Cotton Res. Dev. 25:1-8.

Gupta PK, Varshney RK, Prasad M (2002). Molecular markers: principles and methodology. In: SM Jain, BS Ahloowalia, DS Brar, Eds. Molecular techniques in crop improvement. Kluwer Academic Publishers, Netherlands, pp. 9-54.

Jaccard P (1908). Nouvelles recherche sur la distribution florale. Bulletin de la societe vaudoise des sciences naturelles 44:223-270.

Jain S, Jain RK, McCouch S (2004). Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluoroscently labeled microsatellite markers. Theor. Appl. Genet. 109:965-977.

Kantartzi SK, Ulloa M, Sacks E, Stewart JM (2009). Assessing genetic diversity in *Gossypium arboreum* L. cultivars using genomic and EST-derived microsatellites. Genetica 136:141-147.

Kulkarni VN, Khadi BM, Manjula S, Lalitadas M, Deshapande A, Narayanan SS (2009). The Worldwide Gene Pools of *Gossypium arboreum* L. and *G. herbaceum* L. and their Improvement, In: AH Paterson ed, Genet. Genomics Cotton. Vol 3, pp. 69-100.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009). Potential of molecular markers in Plant Biotechnology. Plant Omics J. 2:141-162.

Mantel N (1967). The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209-220.

Masi P, Spagnoletti Zeuli PL, Donini P (2003). Development and analysis of multiplex microsatellite markers sets in common bean (*Phaseolus vulgaris* L.). Mol. Breed. 11:303-313.

Mohan P, Mukewar PM, Singh VV, Singh P, Khadi BM, Amudha J, Deshpande VG (2006). Identification of sources of resistance to grey mildew disease (*Ramularia areola*) in diploid cotton (*Gossypium arboreum*). CICR Technical Bulletin No: 34, pp. 1-26.

Mukherjee A, Sikdar B, Ghosh B, Banerjee A, Ghosh E, Bhattacharya M, Roy SC (2013). RAPD and ISSR analysis of some economically important species, varieties, and cultivars of the genus *Allium* (Alliaceae). Turk J. Bot. 37:605-618.

Myskow B, Milczarski and Masoje P (2010). Comparison of RAPD, ISSR and SSR markers in assessing genetic diversity among rye (*Secale cereale* L.) inbred lines. Plant Breed. Seed Sci. 62:107-115.

Nibouche S, Brevault T, Klassou C, Dessauw D, Hau B (2008). Assessment of the resistance of cotton germplasm (*Gossypium* spp.) to aphids (Homoptera, Aphididae) and leafhoppers (Homoptera: Cicadellidae, Typhlocybinae): methodology and genetic variability. Plant Breed. 127:376-382.

Noormohammadi Z, Farahani YHA, Sheidai M, Baraki SG, Alishah O (2013b). Genetic diversity analysis in opal cotton hybrids based on SSR, ISSR and RAPD markers. Genet. Mol. Res. 12:256-269.

Noormohammadi Z, Taghavi E, Foroutan M, Sheidai M, Alishah O (2013a). Structure analysis of genetic diversity in tetraploid and diploid cotton genotypes. Int. J. Plant Animal Environ. Sci. 3:79-90.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey J, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2:225-238.

Rohlf FJ (2000). NTSYS-pc: Numerical Taxonomy System. Ver. 2.1. Exeter Software, Setauket, NY, USA. pp. 29-34.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. U.S.A 81:8014-8018.

Sheidai M, Noormohammadi Z, Shojaei-Jesvaghani F, Baraki SG, Farahani F, Alishah O (2012). SSR (Simple sequence repeat) and ISSR (Inter simple sequence repeat) analysis of genetic diversity in tissue regenerated plants of cotton. Afr. J. Biotechnol. 11(56):11894-11900.

Singh A, Dikshit HK, Jain N, Singh D, Yadav RN (2014). Efficiency of SSR, ISSR and RAPD markers in molecular characterization of mungbean and other vigna species. Ind. J. Biotechnol. 13:81-88.

Tahir MS, Khan NUI, Sajid Ur R (2011). Development of an interspecific hybrid (Triploid) by crossing *Gossypium hirsutum* and *G. arboreum*. Cytologia 76:193-199.

Tessier C, David J, This P, Boursiquot JM, Charrier A (1999). Optimisation of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theor. Appl. Genet. 98:171-177.

Ulloa M, Brubaker C, Chee P (2007). Cotton. In: Kole C (ed) Genome mapping & molecular breeding. vol 6: technical crops. Springer, New York, pp.1-49.

Vijayan K (2005). Inter simple sequence repeat (ISSR) polymorphism and its application in mulberry genome analysis. Int. J. Ind. Entomol. 10:79-86.

Wheeler TA, Gannaway JR, Keating K (1999). Identification of resistance to *Thielaviopsis basicola* in diploid cotton. Plant Dis. 83:831-833.

**Table S1.** List of 73 ISSR primers used in the present study, with sequence information and annealing temperature.

| S/N | ISSR name | Sequence | Annealing Temp. (°C) | S/N | ISSR name | Sequence | Annealing Temp. (°C) |
|---|---|---|---|---|---|---|---|
| 1 | ISSR-1 | $(AGC)_5GA$ | 52 | 38 | ISSR-69 | $(AC)_8TA$ | 45 |
| 2 | ISSR-2 | $(AGC)_5GG$ | 55 | 39 | ISSR-70 | $(GT)_8TA$ | 49 |
| 3 | ISSR-3 | $(AGC)_5GT$ | 52 | 40 | ISSR-71 | $(GT)_8CG$ | 50 |
| 4 | ISSR-4 | $(AGC)_5GC$ | 55 | 41 | ISSR-72 | $(CAG)_5$ | 50 |
| 5 | ISSR-5 | $(CA)_7AT$ | 43 | 42 | ISSR-73 | $(CAA)_5$ | 55 |
| 6 | ISSR-6 | $(CA)_7AC$ | 44 | 43 | ISSR-74 | $(GATA)_4$ | 55 |
| 7 | ISSR-7 | $(CA)_7GT$ | 45 | 44 | ISSR-75 | $(TGGA)_4$ | 43 |
| 8 | ISSR-10 | $(CA)_7AA$ | 42 | 45 | ISSR-76 | $(CA)_7AG$ | 45 |
| 9 | ISSR-15 | $(GT)_7AT$ | 42 | 46 | ISSR-77 | $(ACTG)_4$ | 47 |
| 10 | ISSR-16 | $(GT)_7AC$ | 45 | 47 | ISSR-78 | $(GA)_8CG$ | 55 |
| 11 | ISSR-17 | $(GA)_9T$ | 49 | 48 | ISSR-79 | $CCAG(GT)_7$ | 50 |
| 12 | ISSR-18 | $(GA)_9A$ | 49 | 49 | ISSR-80 | $(GACAC)_4$ | 49 |
| 13 | ISSR-19 | $(GA)_9C$ | 51 | 50 | ISSR-81 | $(TG)_8TT$ | 49 |
| 14 | ISSR-27 | $(CA)_9GT$ | 52 | 51 | ISSR-82 | $(TGT)_5$ | 52 |
| 15 | ISSR-28 | $CAG(GA)_7$ | 47 | 52 | ISSR-83 | $(AGC)_5$ | 55 |
| 16 | ISSR-29 | $GCT(GA)_7$ | 47 | 53 | ISSR-84 | $(GAA)_5$ | 52 |
| 17 | ISSR-31 | $T(AG)_7$ | 45 | 54 | ISSR-85 | $GT(CAC)_7$ | 55 |
| 18 | ISSR-34 | $G(CA)_7$ | 42 | 55 | ISSR-86 | $CT(CAC)_7$ | 43 |
| 19 | ISSR-35 | $C(CA)_7$ | 42 | 56 | ISSR-87 | $CAG(CT)_8$ | 44 |
| 20 | ISSR-36 | $A(CA)_7$ | 45 | 57 | ISSR-88 | $CGT(CA)_8$ | 45 |
| 21 | ISSR-38 | $A(CT)_8$ | 45 | 58 | ISSR-89 | $AGG(CA)_8$ | 42 |
| 22 | ISSR-40 | $C(CT)_8$ | 47 | 59 | ISSR-90 | $(CAC)_5GT$ | 42 |
| 23 | ISSR-45 | $(TG)_7C$ | 42 | 60 | ISSR-91 | $(CAC)_5CT$ | 45 |
| 24 | ISSR-47 | $(GACA)_4$ | 45 | 61 | ISSR-92 | $(CAG)_5AT$ | 49 |
| 25 | ISSR-49 | $T(GA)_8$ | 47 | 62 | ISSR-93 | $(CAG)_5GT$ | 49 |
| 26 | ISSR-50 | $C(GA)_8$ | 47 | 63 | ISSR-94 | $TC(GACA)_4$ | 51 |
| 27 | ISSR-58 | $(CTC)_6$ | 46 | 64 | ISSR-95 | $G(TGGGG)_5$ | 52 |
| 28 | ISSR-59 | $(GGGTG)_3$ | 45 | 65 | ISSR-96 | $C(CAG)_5$ | 47 |
| 29 | ISSR-60 | $AGT(AG)_7$ | 47 | 66 | ISSR-97 | $G(CAG)_5$ | 47 |
| 30 | ISSR-61 | $GCG(GA)_7$ | 52 | 67 | ISSR-98 | $GT(GACA)_4$ | 45 |
| 31 | ISSR-62 | $AAG(GT)_7$ | 55 | 68 | ISSR-99 | $(GCTTC)_3$ | 42 |
| 32 | ISSR-63 | $CAC(TG)_7$ | 45 | 69 | ISSR-100 | $(AAG)_5$ | 42 |
| 33 | ISSR-64 | $AAG(CT)_7$ | 45 | 70 | ISSR-101 | $(AAG)_5GT$ | 47 |
| 34 | ISSR-65 | $(TC)_8A$ | 52 | 71 | ISSR-102 | $(AAG)_5GC$ | 45 |
| 35 | ISSR-66 | $(TC)_8G$ | 47 | 72 | ISSR-103 | $T(AAG)_5$ | 55 |
| 36 | ISSR-67 | $(AC)_8C$ | 54 | 73 | ISSR-104 | $G(AAG)_5$ | 55 |
| 37 | ISSR-68 | $(AC)_8CT$ | 55 | | | | |

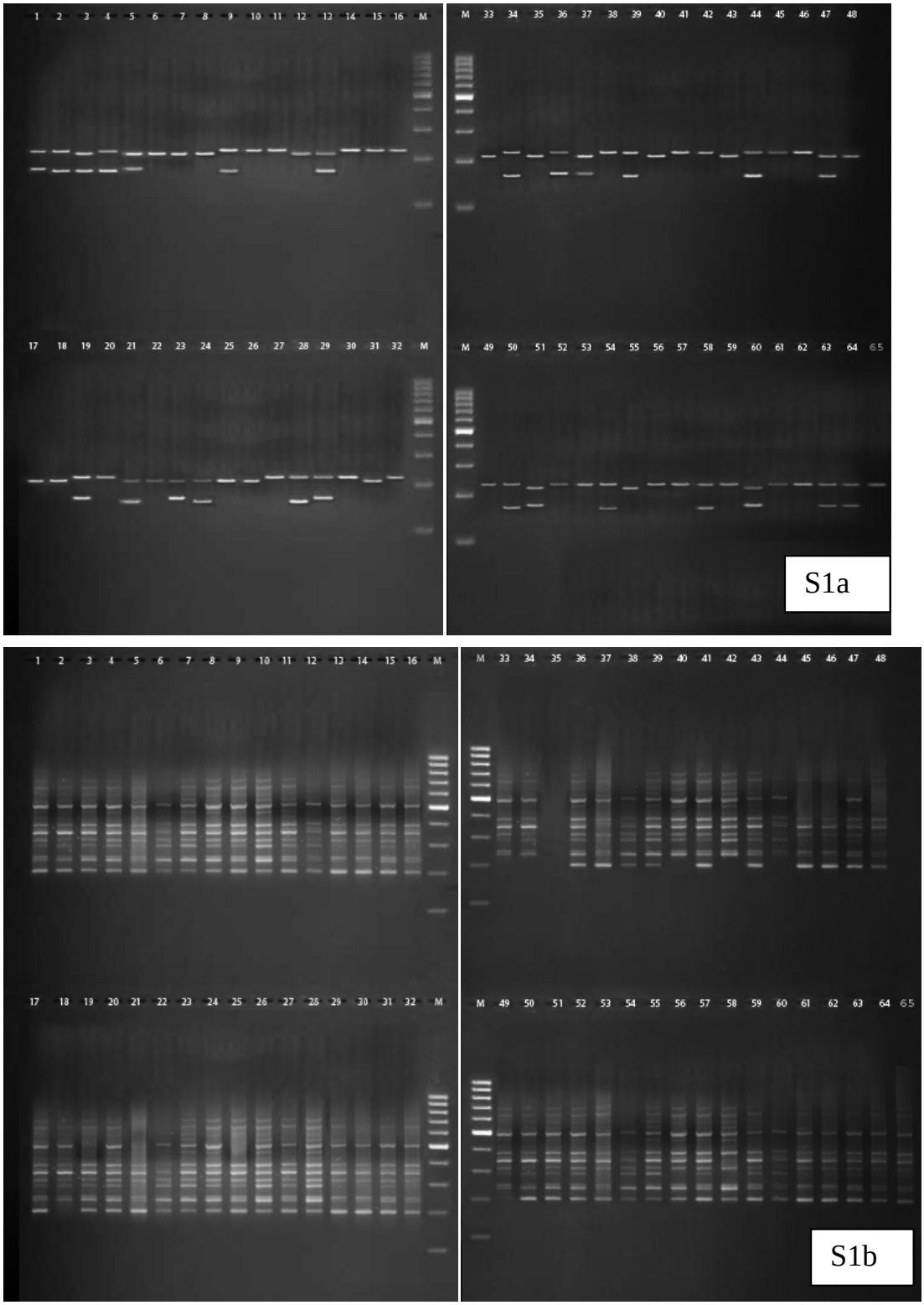


**Figure S1.** a) SSR profile with primer BNL2960, b) ISSR profile with primer ISSR 40, of the selected 65 genotypes (numbers are as per Table 1).

academicJournals

Vol. 15(1), pp. 20-28, 6 January, 2016
DOI: 10.5897/AJB2015.14599
Article Number: 85B1A2756982
ISSN 1684-5315

http://www.academicjournals.org/AJB

African Journal of Biotechnology

*Full Length Research Paper*

# Molecular characterization and genetic diversity assessment of *Tilapia guineensis* from some coastal rivers in Nigeria

**Ukenye E. A.[1]*, Taiwo I. A.[2], Oguntade O. R.[1], Oketoki T. O.[1] and Usman A. B.[1]**

[1]Department of Biotechnology, Nigerian Institute for Oceanography and Marine Research, Victoria Island, Lagos, Nigeria.
[2]Department of Cell Biology and Genetics, University of Lagos, Lagos, Nigeria.

Received 26 March, 2015; Accepted 12 November, 2015

***Tilapia guineensis*** **is an important economically and nutritionally important fish commonly found in Nigerian coastal waters. Genetic diversity of the fish was assessed to obtain information that may help in developing appropriate conservation and breeding programmes for improving the economic and nutritional quality of the fish. Twelve (12) Nigerian coastal populations and nine microsatellite loci were considered. All the loci were multi-allelic giving an average of 3.1 alleles per locus. The number of alleles (Na) ranged from two to four alleles per locus while the effective number of expected alleles (Ne) ranged from 1.087 to 2.612. Buguma, Badagry and Brass populations had the highest genetic diversity as was revealed by heterozygosity (observed and expected) and shannon index of the populations. The longest pairwise genetic distance of 0.30 was between Brass in Bayelsa State and River Ethiope in Delta State. Clustering using simple sequence repeat (SSR) data gave four major clusters which did not concur with geographical location clustering. We conclude that although genetic diversity is low in some populations of *T. guineensis* in Nigerian coastal waters, some populations (Buguma, Badagry and Brass) still retain some genetic variability which may be explored for fish improvement through appropriate breeding and conservation programmes.**

**Key words:** Molecular characterization, genetic diversity, microsatellite analysis, *Tilapia guineensis*, coastal rivers.

## INTRODUCTION

*Tilapia guineensis* is one of the most important Cichlid species, in view of its nutritional role in many tropical and sub-tropical countries (Saisithi, 1994). It is an important source of animal protein and income throughout the world especially in developing countries like Nigeria with many rural populations relying on subsistent farming (Sosa et al., 2005). It has continued to contribute immensely to the nutritional needs, economic growth and development of

*Corresponding author. E-mail: eekwelem@yahoo.com.

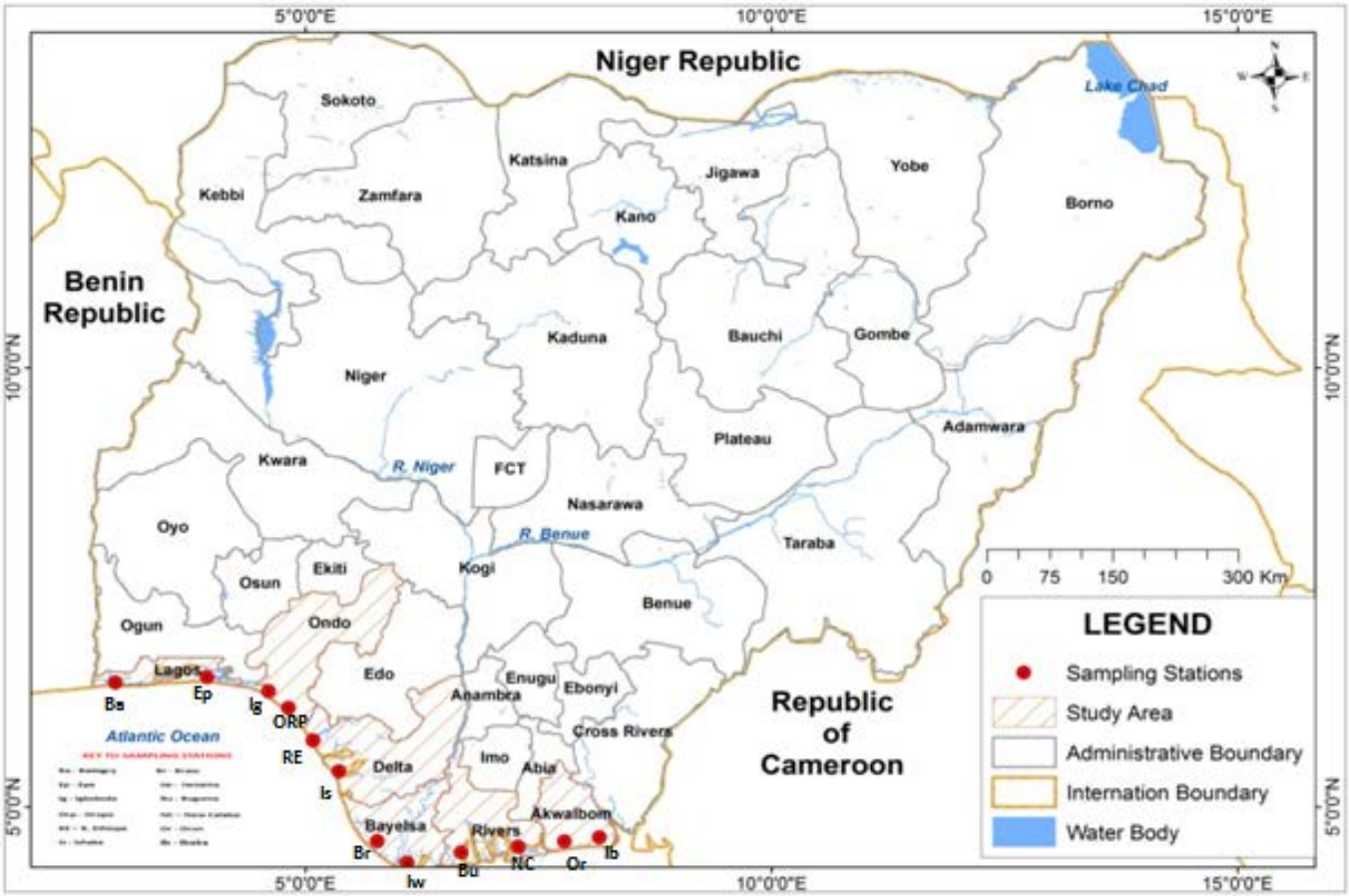


**Figure 1.** Map of Nigeria showing sampling stations.

many nations including Nigeria. The species is usually found in creeks, lagoons and other coastal waters of West Africa (Philippart and Ruwet, 1982). It shows good aquaculture potential and has been successfully raised in ponds, enclosures, cages and tanks. Despite the nutritional and economic importance of *T. guineensis* worldwide, our knowledge of the genetic status in terms of genetic diversity of its natural populations is still inadequate for sustainable aquaculture practices, conservation and improvement through selective breeding (Agenes et al., 1999).

The Nigerian coastal zone comprises of eight states (Figure 1), and the majority of these populations depend on catch from the wild. This includes *T. guineensis* as a source of animal protein. Thus, Tilapia has grown to represent the third most important finfish in the world (Sofia 2012). Therefore, efforts to determine the current level of diversity and genetic structure of *T. guineensis* populations in Nigeria and many other parts of the world are useful for fishery management, aquaculture, stock conservation, and fish improvement through breeding.

Information about the genetic diversity of the wild fish populations through microsatellite analysis is essential in breeding for heterosis and effective fish management (Bo-young et al., 2005; Lee and Kocher, 1998). An extensive search on the literature revealed that much work has not been done on genetic diversity of *T. guineensis* in Nigeria. In a recent study, Abd-el-kader et al. (2013) observed a relatively high level of genetic diversity within and between three Tilapia species namely *Tilapia zilli, Oreochromis aureus* and *Oreochromis niloticus* in Egypt. In Nigeria and many other West African countries, *T. guineensis* is one of the dominant Tilapia species; in view of the importance of genetic variability in selective breeding, it is necessary to assess genetic variability of *T. guineensis* for sustainable fish improvement in Nigeria and many other African countries (Bentsen and Olesen, 2002).

Apart from the usefulness of knowledge of genetic variability in fish improvement and management, assessment and conservation of genetic variability is important in assessing risk of reduction in fitness through inbreeding and loss of species through extinction (Frankham, 2003). Consequences of these are well known (Falconer, 1989; Keller and Waller, 2002). As with all *Tilapia* species, there is a great potential for enhancing

**Table 1.** Geographical location of the sampling stations.

| Location | Latitude | Longitude | State |
|---|---|---|---|
| Buguma | N04° 44.613′ | E006° 57.401′ | Rivers |
| New Calabar | N04° 448′ | E07° 010′ | Rivers |
| Ishaka | N05° 03.243′ | E005° 45.332′ | Delta |
| River Ethiope | N05° 53.397′ | E005° 33.671′ | Delta |
| Epe | N06° 35.832′ | E02° 59.096′ | Lagos |
| Igbokoda | N06° 21.028′ | E004° 48.319′ | Ondo |
| Oropo | N06° 25.238′ | E04° 75.228′ | Ondo |
| Iwoama | N04° 51.224′ | E06° 28.333′ | Bayelsa |
| Brass | N04° 31.500′ | E06° 24.167′ | Bayelsa |
| Badagry | N04° 25.012′ | E02° 52.98′ | Lagos |
| Oron | N04° 49.217′ | E008° 04.625′ | Akwa Ibom |
| Ibaka | N04° 27.200′ | E007° 19.618′ | Akwa Ibom |

growth and production through genetic selection. The need to assess the genetic diversity of *T. guineensis* populations using microsatellite markers for breeding and conservation purposes necessitated this study.

## MATERIALS AND METHODS

### Collection of fish samples

A total of 120 *T. guineensis* individuals with weight range of 20 to 35 g and length of 11.5 to 14.0 cm were identified and collected from 12 coastal rivers (ten from each river) in the Niger delta, Nigeria which includes; Epe lagoon, Badagry lagoon, Igbokoda, Oropo river, Ishaka Creek, River Ethiope, Buguma, New Calabar river, Iwoama river, Brass, Oron and Ibaka river (Figure 1). The geographical location in terms of longitudes and latitudes of the sampling stations are presented in Table 1. Experimental fish samples were identified to be *T. guineensis* by a fish taxonomist from Nigerian Institute for Oceanography and Marine Research Lagos, Nigeria, and obtained from the fishermen at the landing sites.

### Extraction of DNA and PCR amplification

Caudal fin tissue (1 $cm^2$) was collected from each individual and placed in 95% ethanol for preservation until analysis. Genomic DNA was extracted from the caudal fin tissue using phenol-chloroform protocol (Sambrook and Russell, 2001). The quality of extracted DNA was checked using a Nano-drop spectrophotometer (Shimadzu corporation Japan, MODEL UV-1800, 2000 series) at absorbance of 260/280 nm. Amplification was carried out using nine microsatellite primers (Table 2) originally developed for tilapia by Lee et al. (2005). A total volume of 20 µl of the PCR ingredients which consisted of 4 µl Solis Biodyne (SBD) 5x fire pol (master mix with 12.5 mM MgCl), 13.6 µl dd $H_2O$, 0.5 µl dNTP (0.2 mM; nucleotides), 0.2 µl forward primer, 0.2 µl reverse primer, and 2 µl of template DNA (10 ng) was run on a Thermocycler (Biorad, module 170 - 8731). The program for PCR amplification was: 2 min initial 96°C denaturation, 30 cycles of 94°C for 30 s, 30 s at the appropriate annealing temperature (Table 1), and 30 s at 72°C, followed by a 6 min final extension step at 72°C. The samples were stored at -20°C until separation on polyacrylamide gels (6% polyacrylamide gel, at 80 V for 2 h in a 1 × TBE buffer). The gel was stained with ethidium bromide and visualized in a UV transilluminator. Two researchers independently scored the gel bands to reduce or rule out error due to improper scoring.

We could not observe amplification at 65°C annealing temperature unlike Saad et al. (2013) who obtained amplicons at an annealing temperature of 65°C in tilapia. In the present study, we obtained PCR amplification at 55°C through optimization of PCR conditions.

### Data analysis

Population genetic data generated was analysed using PopGene v. 3.6 software to obtain the number of alleles per SSR locus, effective number of alleles, Shannon information index, observed heterozygosity, expected heterozygosity and Nei's Pairwise genetic distance (1972). Genetic relationship among populations was estimated by constructing a dendrogram using unweighted pair-group method of analysis (UPGMA). In an attempt to compare genetic relationship with geographical location, a dendrogram based on geographical location (longitude and latitude) was generated using clustering algorithm of SPSS version 21 software. Polymorphic information content (PIC), major allele frequency and gene diversity were determined using PowerMarker v. 3.6.

## RESULTS

### Genetic variability among microsatellite loci

All nine microsatellite loci were polymorphic in all populations (Figure 2) with polymorphic information content (PIC) values ranging from 0.07 at locus GM211 to 0.54 at locus UNH207 with an average of 0.31 (Table 3). A total of 28 alleles were found in the study. The mean number of alleles per locus was 3.1. Locus UNH207 and UNH 185 gave the highest number of alleles (four alleles, respectively) while UNH123 gave the least (two alleles). On average, 73% of the 120 individuals shared a common major allele at a giving locus ranging from 45% (UNH207) to 96% (GM211)

**Table 2.** SSR primer code, sequences, annealing temperature and band size.

| Primer code | Sequence | Annealing temperature (°C) | Molecular size (bp) |
|---|---|---|---|
| UNH995 | Forward 5' CCAGCCCTCTGCATAAAGAC 3'<br>Reverse 5' GCAGCACAACCACAGTGCTA 3' | 55 | 150-200 |
| GM538 | Forward 5' CAGCATGTTGTCTGGATCTTG 3'<br>Reverse 5' TTTGTTGCTGTGGTCTGTTCTT 3' | 55 | 150-200 |
| GM531 | Forward 5' AAAGCCAACGGTCTGAATTG 3'<br>Reverse 5' AGCAGAGGACACCCCTCAT 3' | 55 | 100-150 |
| GM211 | Forward 5' GCAAGTTGAGAGGCTACTGT 3'<br>Reverse 5' AAACAACCCACAACCTTAGTT 3' | 55 | 100-150 |
| UNH207 | Forward 5' ACACAACAAGCAGATGGAGAC3'<br>Reverse 5' CAGGTGTGCAAGCAGAAGC 3' | 55 | 100-150 |
| UNH185 | Forward 5' CAGACACACTAGACACATTCTA 3'<br>Reverse 5' GTGTTTCCATGTGTCTGTAC 3' | 55 | 120-150 |
| UNH146 | Forward 5' CCACTCTGCCTGCCCTCTAT 3'<br>Reverse 5' AGCTGCGTCAAACTCTCAAAAG 3' | 55 | 100-150 |
| UNH123 | Forward 5' CATCATCACAGACAGATTAGA 3'<br>Reverse 5' GATTGAGATTTCATTCAAG 3' | 55 | 100-150 |

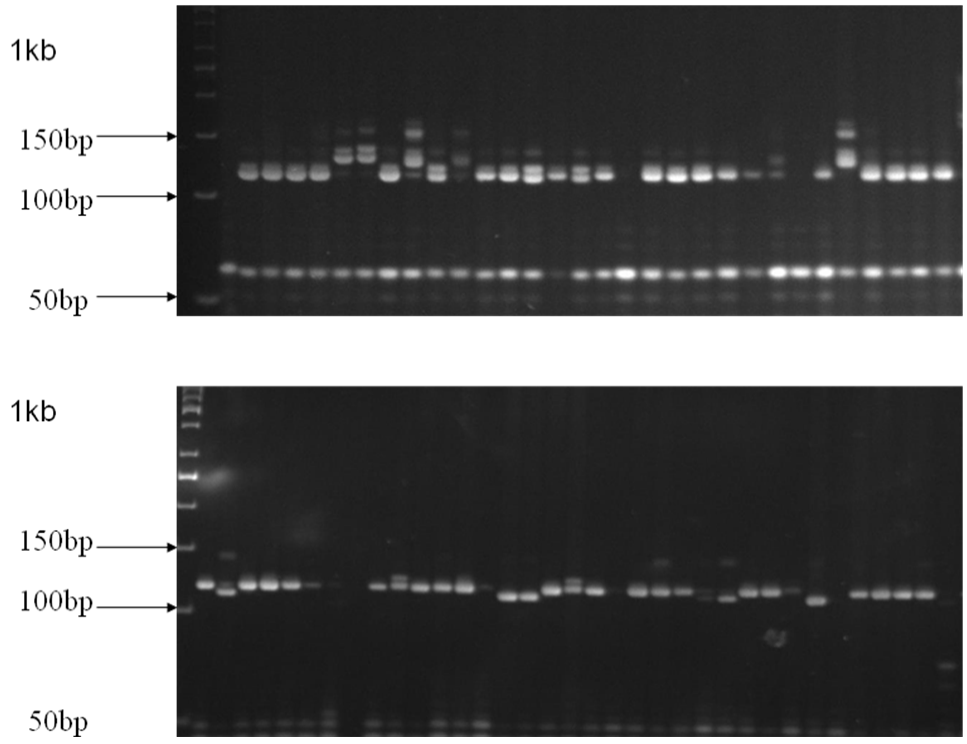


**Figure 2.** Electrophoresis of the amplified microsatellite loci using UNH995 primer. Upper panel: 1-60 individuals; Lower panel: 61-120 individuals.

**Table 3.** Characteristics of SSR Loci analyzed

| Marker | Freq. | Sample size | NA | Gene diversity | PIC |
|---|---|---|---|---|---|
| UNH995 | 0.83 | 120 | 3 | 0.29 | 0.27 |
| GM538 | 0.46 | 120 | 3 | 0.57 | 0.48 |
| GM531 | 0.47 | 120 | 3 | 0.58 | 0.48 |
| GM211 | 0.96 | 120 | 3 | 0.08 | 0.07 |
| UNH207 | 0.45 | 120 | 4 | 0.62 | 0.54 |
| UNH185 | 0.40 | 120 | 4 | 0.57 | 0.50 |
| UNH146 | 0.68 | 120 | 3 | 0.48 | 0.42 |
| UNH123 | 0.95 | 120 | 2 | 0.09 | 0.09 |
| UNH104 | 0.88 | 120 | 3 | 0.23 | 0.21 |
| Mean | 0.73 | 120 | 3.1 | 0.35 | 0.31 |

Freq., major allele frequency; NA, number of allele; PIC, polymorphic information content.

**Table 4.** Locus specific indices of genetic diversity in the combined population.

| Locus | No. of allele | Effective alleles | Observed heterozygosity | Expected heterozygosity | Fis | D |
|---|---|---|---|---|---|---|
| UNH995 | 3 | 1.410 | 0.100 | 0.292 | 0.656 | -0.658 |
| GM538 | 3 | 2.341 | 0.942 | 0.575 | -0.644 | 0.389 |
| GM531 | 3 | 2.358 | 0.817 | 0.578 | -0.418 | 0.414 |
| GM211 | 3 | 1.087 | 0.583 | 0.808 | 0.275 | -0.278 |
| UNH207 | 4 | 2.612 | 0.442 | 0.619 | 0.284 | -1.286 |
| UNH185 | 4 | 2.428 | 0.308 | 0.586 | 0.316 | -0.819 |
| UNH146 | 3 | 1.919 | 0.283 | 0.481 | 0.408 | -0.412 |
| UNH123 | 2 | 1.105 | 0.050 | 0.954 | 0.474 | -0.948 |
| UNH104 | 3 | 1.293 | 0.117 | 0.228 | 0.485 | -0.487 |
| Mean | 3.1 | 1.706 | 0.324 | 0.324 | 0.066 | -0.409 |

NA, number of alleles; NE, effective number of alleles; Ho, observed heterozygosity; He, expected heterozygosity; Fis, inbreeding coefficient; D, heterozygote deficiency calculated as D = (Ho-He)/He.

common allele per locus. The level of diversity revealed by the studied loci ranged from 0.08 to 0.6 with an average of 0.35 (Table 3). Locus UNH207 had the highest effective number of alleles (2.6) while GM211 gave the least (1.0) (Table 4). The highest observed heterozygosity was obtained by locus GM538 (0.94) while locus UNH123 had the lowest (0.050). The expected heterozygosity ranged from 0.186 to 0.954 with a mean of 0.349 (Table 4). The inbreeding coefficient (Fis) was positive across seven loci in all populations reflecting excess of homozygotes while two loci were negative indicating an excess of heterozygotes (Table 4).

## Genetic differences among populations

The Badagry population had the highest mean number of alleles (2.67), followed by Buguma (2.56) and Brass (2.44) while the lowest was found in Igbokoda (1.44). The mean effective alleles varied from 1.29 to 2.11. In all populations, the mean effective number of allele was lower than the mean number of alleles. Shannon information index was observed higher in Buguma population (0.77), Badagry (0.76) and Brass (0.64) reflecting high genetic diversity while other populations had low index. All populations showed low average observed heterozygosity. Badagry was the most variable (Ho = 0.467) followed by Buguma (Ho = 0.402) and Brass (Ho = 0.456) while Oron had the least observed heterozygosity (Ho = 0.211). The average expected heterozygosity was high in Buguma (0.503), Badagry (0.484) and Brass (0.411) and low in Oron (0.178) and Igbokoda (0.180) populations as shown in Table 5.

According to Table 6, Nei's genetic distance between the populations ranged from 0.01 to 0.30. The highest genetic dissimilarity was between River Ethiope and Brass with a genetic distance of 0.30. Based on geographical location, the highest distance was between Oron and Epe (29.0) (Table 7). Thus, genetic distance did not concur with geographical distance in this study.

The UPGMA dendrogram based on the genetic distances revealed four clusters: cluster 1 consists of

**Table 5.** Summary of the genetic diversity level in the twelve studied populations.

| Population | Na | Ne | I | Ho | He |
|---|---|---|---|---|---|
| Buguma | 2.7 | 2.11 | 0.77 | 0.402 | 0.503 |
| New Calabar | 1.7 | 1.53 | 0.36 | 0.400 | 0.247 |
| Ishaka | 2.2 | 1.49 | 0.44 | 0.333 | 0.273 |
| River Ethiope | 2.2 | 1.69 | 0.54 | 0.289 | 0.336 |
| Epe | 1.9 | 1.47 | 0.36 | 0.344 | 0.236 |
| Igbokoda | 1.4 | 1.35 | 0.25 | 0.300 | 0.180 |
| Oropo | 2.0 | 1.58 | 0.44 | 0.233 | 0.286 |
| Iwoama | 1.9 | 1.39 | 0.36 | 0.244 | 0.225 |
| Brass | 2.4 | 1.87 | 0.64 | 0.456 | 0.411 |
| Bdagry | 2.7 | 2.09 | 0.76 | 0.467 | 0.484 |
| Oron | 1.7 | 1.27 | 0.27 | 0.211 | 0.178 |
| Ibaka | 1.8 | 1.33 | 0.31 | 0.233 | 0.202 |

NA, number of alleles; NE, effective number of alleles; I, shannon information index; Ho, observed heterozygosity; He, expected heterozygosity.

New Calabar, Ishaka, Igbokoda, Epe, Oron and Ibaka, while cluster 3 consists of Oropo, Iwoama, Brass and Badagry and finally, cluster 4 consists of River Ethiope that formed an out-group (Figure 3). Three clusters were obtained from dendrogram based on geographical location (Figure 4). The tree topology based on genetic distance showed that Oron clustered with Ibaka while Igbokoda clustered with Ishaka. However, based on the geographical location, Oron clustered with Buguma while Ishaka clustered Brass.

## DISCUSSION

Nine microsatellite markers were utilized to characterize and investigate genetic variation in some coastal populations of *T. guineensis* in Nigeria with a view to stimulating interest and giving insights into possibilities of improving nutritional and economic qualities of *T. guineensis* through breeding and conservation programmes. We felt that nine SSR markers were sufficient to characterize the fish populations in view of the work of Abdul et al. (2012) who also used nine microsatellite markers to assess diversity in rice. Moreover, our results in which a total of 28 alleles were revealed is similar to that of Abdul et al. (2012) who got a total 27 alleles in rice. In contrast, the study of Hesham and Gilbey (2005) revealed 80 alleles in six loci from five populations of *O. niloticus.* An earlier report by Hesham and Gilbey (2005) is comparable to a more recent report of Gu et al., (2014) in which 10 microsatellite loci revealed 75 alleles in *O. niloticus* populations. These results suggest that *T. guineensis* had lower genetic diversity than *O. niloticus.* Urgent steps are therefore necessary to arrest further reduction in diversity of *T. guineensis* through various breeding and conservation programmes.

Electrophoresis of PCR–amplified DNA gave one or two bands. This is expected in microsatellite analysis where one band represents homozygosity and two bands represent heterozygosity. Nevertheless, null and multiple alleles were obtained for a few loci in this study. Occurrence of null alleles may indicate over-stringent PCR conditions and poor primer annealing due to nucleotide sequence divergence in one or both flanking regions resulting in non-amplification. Presence of multiple alleles might suggest aneploidy or/and existence of paralogy in the genome of *T. guineensis* in Nigerian coastal waters. Considering rarity of aneuploidies and its adverse effect on genome balance and survival in animals, paralogy seems to be a more attractive explanation for the existence of multiple bands in this study. Nevertheless, there is need for selective optimization for primer annealing coupled with cytogenetic analysis in future studies. The observed number of alleles (Na) and the effective number of alleles (Ne) varied among *T. guineensis* populations in the present study. The average number of alleles observed in Buguma, Badagry and Brass were higher than that of other populations indicating more allelic polymorphism in Buguma, Badary and Brass populations.

The polymorphic information content (PIC) of 0.31 obtained in our study suggests that the microsatellite loci considered were moderately informative with good discriminating power in accordance with the view of Bostein et al. (1980). Thus, these markers had good merits for detecting DNA identity and diversity in these populations and are therefore suitable for use in the characterization of natural populations and determination of genetic differentiation in *T. guineensis*. In all investigated populations, only Buguma population demonstrated polymorphism for all loci while others showed lower polymorphism. This result is contrary to the result obtained by Corujo et al. (2004) in nine populations

**Table 6**. Nei's genetic distance between twelve *T. guineensis* populations revealed by nine microsatellite loci.

| Location | Buguma | New calabar | Ishaka | River Ethiope | Epe | Igbokoda | Oropo | Iwoama | Brass | Badagry | Oron | Ibaka |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Buguma | 0.00 | | | | | | | | | | | |
| New Calabar | 0.09 | 0.00 | | | | | | | | | | |
| Ishaka | 0.08 | 0.01 | 0.00 | | | | | | | | | |
| River Ethiope | 0.12 | 0.15 | 0.14 | 0.00 | | | | | | | | |
| Epe | 0.09 | 0.03 | 0.01 | 0.11 | 0.00 | | | | | | | |
| Igbokoda | 0.1 | 0.02 | 0.01 | 0.12 | 0.01 | 0.00 | | | | | | |
| Oropo | 0.1 | 0.04 | 0.04 | 0.14 | 0.04 | 0.04 | 0.00 | | | | | |
| Iwoama | 0.18 | 0.1 | 0.11 | 0.26 | 0.13 | 0.14 | 0.05 | 0.00 | | | | |
| Brass | 0.18 | 0.17 | 0.17 | 0.30 | 0.18 | 0.21 | 0.14 | 0.11 | 0.00 | | | |
| Badagry | 0.08 | 0.08 | 0.07 | 0.18 | 0.08 | 0.09 | 0.06 | 0.09 | 0.05 | 0.00 | | |
| Oron | 0.12 | 0.04 | 0.03 | 0.16 | 0.03 | 0.02 | 0.07 | 0.21 | 0.27 | 0.11 | 0.00 | |
| Ibaka | 0.11 | 0.03 | 0.03 | 0.18 | 0.04 | 0.03 | 0.07 | 0.19 | 0.25 | 0.09 | 0.01 | 0.00 |

**Table 7**. Distance matrix based on geographical (longitude and latitude) location.

| Location | Badagry | Brass | Buguma | Epe | Ibaka | Igbokoda | Ishaka | Iwoama | N. Calabar | Oron | Oropo | R. Ethiope |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Badagry | 0.00 | | | | | | | | | | | |
| Brass | 10.00 | 0.00 | | | | | | | | | | |
| Buguma | 17.00 | 1.00 | 0.00 | | | | | | | | | |
| Epe | 9.00 | 13.00 | 20.00 | 0.00 | | | | | | | | |
| Ibaka | 16.00 | 2.00 | 1.00 | 25.00 | 0.00 | | | | | | | |
| Igbokoda | 8.00 | 2.00 | 5.00 | 5.00 | 8.00 | 0.00 | | | | | | |
| Ishaka | 13.00 | 1.00 | 2.00 | 10.00 | 5.00 | 1.00 | 0.00 | | | | | |
| Iwoama | 10.00 | 0.00 | 1.00 | 13.00 | 2.00 | 2.00 | 1.00 | 0.00 | | | | |
| N. Calabar | 17.00 | 1.00 | 0.00 | 20.00 | 1.00 | 5.00 | 2.00 | 1.00 | 0.00 | | | |
| Oron | 26.00 | 4.00 | 1.00 | 29.00 | 2.00 | 10.00 | 5.00 | 4.00 | 1.00 | 0.00 | | |
| Oropo | 8.00 | 2.00 | 5.00 | 5.00 | 8.00 | 0.00 | 1.00 | 2.00 | 5.00 | 10.00 | 0.00 | |
| R. Ethiope | 8.00 | 2.00 | 5.00 | 5.00 | 8.00 | 0.00 | 1.00 | 2.00 | 5.00 | 10.00 | 0.00 | 0.00 |

of brown trout in Spain with as many as seven populations having all loci polymorphic.

Three populations namely Buguma, Badagry and Brass were identified as having considerable biodiversity in this study. This was based on Shannon's information index and heterozygosity (observed and expected) which were higher in these populations when compared to others. Higher heterozygosity implies greater genetic variability according to Mu et al. (2011) who stated that heterozygosity is an important measure of population diversity at the genetic level. Thus, in order to embark on a meaningful breeding and conservation programme for *T. guineensis* in Nigerian coastal waters, the identified populations

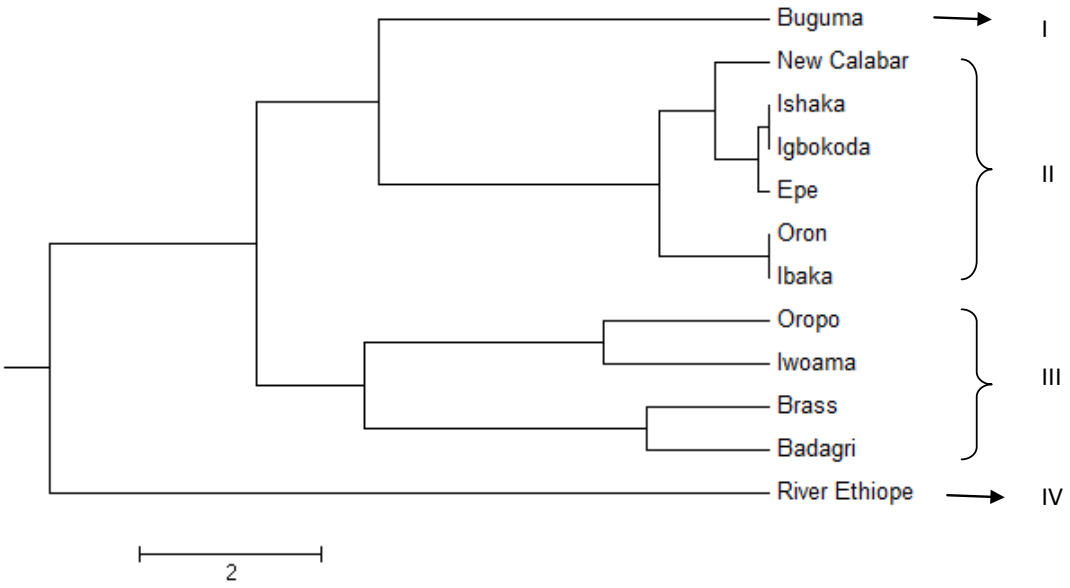


**Figure 3.** UPGMA dendrogram showing the genetic relationships among 12 populations based on Nei's genetic distance.

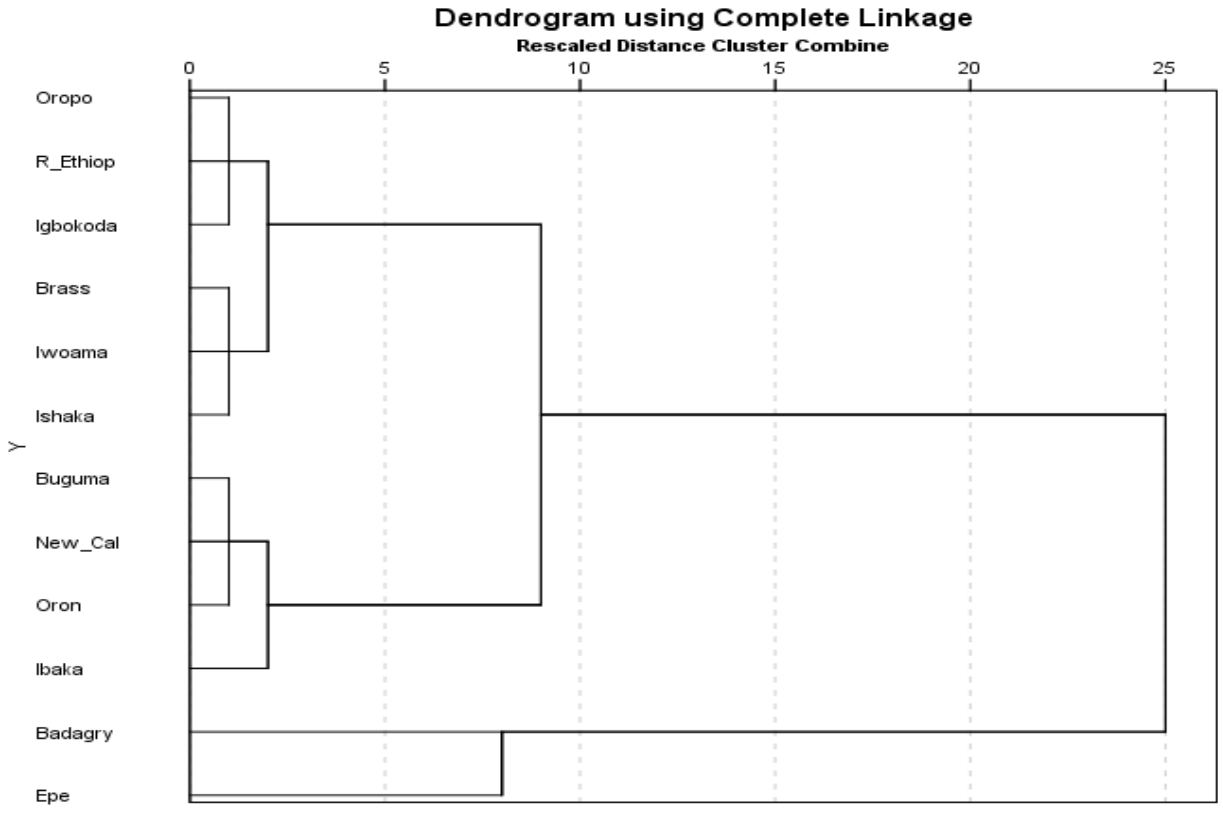


**Figure 4.** Dendrogram based on geographical location using longitudinal and latitudinal location of the populations.

(that is, Buguma, Badagry and Brass) should be considered as sources of fish for improvement programmes.

Inbreeding coefficient (Fis) is expressed as a deficiency in heterozygotes, the theoretical value ranges from -1 to +1, where by positive values indicate heterozygote deficiency possibly due to inbreeding (Boris Brinez et al., 2011). In the present study, all the loci showed positive

inbreeding coefficient (Fis) except two loci suggesting that inbreeding leading to reduced number of heterozygotes exists in *T. guineensis*. Deleterious consequences of inbreeding which include loss of diversity, fitness amd extinction had been discussed in other studies (Antunes et al., 2006).

Clustering based on the genetic distance gave four major clusters indicating some level of genetic variability between the studied populations. However, genetic clustering due to microsatellite analysis did not agree with clustering based on geographical location. Therefore, proximity may not be a significant factor favoring gene flow between these populations. It is therefore likely that *T. guineensis* migrate through long distance during breeding season.

## Conclusion

Despite some evidence of inbreeding and low biodiversity among *T. guineensis* populations, there is still some genetic variability in some of the studied populations. Buguma, Badagry and Brass still contain sufficient genetic diversity that can be exploited for breeding and conservation programmes to improve economic and nutritional qualities of *T. guineensis* in Nigeria.

## Conflict of interests

The authors have not declare any conflict of interest.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance and support of Genomic Division of Biotechnology, Department of Nigerian Institute for Oceanography and Marine Research (NIOMR), Lagos, Nigeria. The cooperation of the Bioscience Laboratory staff of International Institute of Tropical Agriculture (IITA) Ibadan, Nigeria, in conducting part of this research is highly appreciated.

### REFERENCES

Abd el-kader Z, Abdel-hamid Karima G, Mahrous F (2013). Genetic Diversity among Three Species of Tilapia in Egypt Detected by Random Amplified Polymorphic DNA Marker. J. Appl. Biol. Sci. 7:57-64.

Abdul MS, Musharaf H, Mosnaz ATMJ, Hosneara H, Monirul I, Shamsher A, Shamsul HP (2012). SSR marker-based molecular characterization and genetic diversity analysis of aromatic landreces of rice (*Oryza sativa* L.). J. BioSci. Biotech. 1(2):107-116.

Agnese J, Gourene B, Owino J, Pouyand I, Aman R (1999). Genetic characterization of a pure relict population of *Oreochromis esculentus*, an endangered tilapia. J. Fish Biol. 12: 1119-1123.

Antunes A, Faria R, Johnson WE, Guyomard R, Alexandrino P (2006). Life on the edge: the long-term persistence and contrasting spatial genetic structure of distinct brown trout life histories at their ecological limits. J. Hered. 97:193-205

Boris Briñez R, Xenia Caraballo O, Marcela Salazar V MD (2011). Genetic diversity of six populations of red hybrid tilapia, using microsatellites genetic markers. *Revista.MVZ Cordoba* 16: 129-132.

Bentsen HB, Olesen I (2002). Designing aquaculture mass selection programs to avoid high inbreeding rates. Aquaculture 204:349-359.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Bo-young L, Woo-Jai L, Todd Streelman J, Karen LC, Aimee EH, Gideo H, Audun Slettan, Justin ES, Terai Y, Kocher TD (2005). A second-generation genetic linkage map of tilapia (Oreochromis niloticus). J. Genet. 17:237-244.

Corujo M., Blanco G., Vasquez e, Sanchez JA (2004). Genetic structure of northwestern Spanish brown trout (*Salmo trutta* L.) populations, differences between microsatellite and allozyme loci. Hereditas 141:258-271.

Gu DE, Mu XD, Song HM, Luo D, Xu M, Luo JR, Hu YC (2014). Genetic diversity of invasive Oreochromis spp.(tilapia) populations in Guangdong province of China using microsatellite markers. Biochem. Syst. Ecol. 55:198-204.

Falconer DS (1989). Introduction to quantitative genetics. 3rd Edition. New York.

Frankham R (2003). Genetics and conservation biology. C R Biol. 326:22-29.

Hesham AH, John Gilbey (2005). Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. J. Aquacult. Res. 36:1450-1457.

Keller LF, Waller DM (2002). Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230-241.

Lee WJ, Kocher TD (1998). Microsattelite mapping of the prolactin locus in the tilapia genome. J. Anim. Genet. 29(1):68-69.

Lee B-Y, Lee W-J, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD (2005). A second-generation genetic linkage map of tilapia (Oreochromis spp.). Genetics 170(1):237-244.

Mu XD, Hu, YC, Wang XJ, Song HM, Yang YX, Luo JR (2011). Genetic variability in cultured stocks of *Scleropages formosus* in Mainland China revealed by microsatellite markers. J. Anim. Vet. Adv. 10:555-561.

Nei M (1972). Genetic distance between populations. Am. Nat. 106:283-292.

Philippart J-Cl, Ruwet J-Cl (1982). Ecology and distribution of tilapias. Conference Proceedings, International 1 Center for Living Aquatic Resources Management, Manila, Philippines. 7:432.

Saad YM, Rashed MA, Atta AH, Ahmed NE (2013). The efficiency of microsatellite DNA markers for estimating genetic polymorphism in some Tilapia species. Life Sci. J. 10(3):2230-2234.

Saisithi P (1994). Traditional fermented fish source production. In: Fisheries Processing Biotechnological. pp. 1389-1400.

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

Sofia, (2012). State of world fisheries and aquaculture. Food and Agriculture Organization of the United Nations, Rome.

Sosa I, Adillo M, Ibanez AL, Figueroa J (2005). Variability of tilapia (Oreochromis spp) introduced in Mexico: Morphometric, Meristic and genetic characters. J. Appl. Ichthyol. 20:7-10.

# African Journal of Biotechnology

*Related Journals Published by Academic Journals*

*Biotechnology and Molecular Biology Reviews*
*African Journal of Microbiology Research African Journal of Biochemistry Research African Journal of Environmental Science and Technology*
*African Journal of Food Science*
*African Journal of Plant Science*
*Journal of Bioinformatics and Sequence Analysis*
*International Journal of Biodiversity and Conservation*

academicJournals