Full Length Research Paper

Isolation and identification of microsatellite repeat motifs from the *Epinephelus fuscoguttatus* genome

Mohd Azinuddin Ahmad Mokhtar[1,2], Syarul Nataqain Baharum[1]*, Normah Mohd Noor[1,2] and Vijay Kumar[3]

[1]Institute of Systems Biology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.
[2]Schools of Bioscience and Biotechnology, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia.
[3]Biotechnology Research Institute, Universiti Malaysia Sabah, Jln UMS, 88400 Kota Kinabalu, Sabah, Malaysia.

Epinephelus fuscoguttatus belongs to one of the largest serranidae fish family. Genetic information regarding existing fish populations in the wild is crucial for the conservation, particularly since the species is listed under the IUCD Red List due to intense fishing. Microsatellites of *E. fuscoguttatus* were isolated using streptavidin-biotin enrichment method. In total, 378 microsatellites were identified and characterized. Of these 378 total microsatellites, 46 (12.2%) were mononucleotides, 175 (46.3%) were dinucleotides, 109 (28.8%) were trinucleotides, 36 (9.5%) were tetranucloetides, 7 (1.9%) were pentanucleotides, 4 (1.1%) were hexanucleotides and 1 (0.3%) was a heptanucleotide. The most abundant microsatellite present in *E. fuscoguttatus* was the dinucleotide motif, $(AC)_n$.

Key words: Enrichment, *Epinephelus fuscoguttatus*, microsatellite, diversity.

INTRODUCTION

Groupers belong to the order of Perciformes, which the family contains about 500 species of serranides with approximately 159 species of Epinephelinae (Heemstra and Randall, 1993). Groupers have an oblong body with small scales, saw-toothed edges, a very large mouth, and coarse, spiny fins (Braise, 1987) that inhabit a stony environment in tropical areas with coral reefs. Groupers vary in color according to their habitat, water depth, and the extent of distressing conditions (Heemstra and Randall, 1993). Among Epinephelinae species that are commercially important is *Epinephelus fuscoguttatus* and cultured intensely in Kuwait, Indonesia, Malaysia, Thailand, the Philippines, Hong Kong, Taiwan, The Republic of China, Japan, and Mexico (INFOFISH, 1989; Tookwinas, 1989; Bombeo-Tuburan, et al., 2001; Lin and Shiau, 2003). In 2007, world grouper production increased by 12,358 tonnes from the previous year, with 22.9% alone comes from Taiwan (Froese and Pauly, 2010). International Union for Conservation of Nature

and Natural Resources (IUCN) has status *E. fuscoguttatus* as 'being near threatened' due to extensive collection from the wild since every stages or size has its own value (Cornish, 2004).

Over the years, different types of genetic markers have been introduced in studying genetic structures of species or populations and among it is microsatellite. Microsatellites represent powerful markers in population genetics primarily because they evolve rapidly, are found throughout the nuclear genome, generally have several alleles per locus, and are typically inherited in a co-dominant fashion (Liu and Cordes, 2004). In fisheries and aquaculture industry, microsatellites have been used in the characterization of genetic stocks, broodstock selection, the construction of dense linkage maps, the mapping economically important quantitative traits, for identifying genes responsible for these traits, and for marker-assisted breeding programs (Chistiakov et al., 2006). The major disadvantage of microsatellites is the high cost and extensive time required to develop the marker as microsatellites must be developed for individual species, or at least for each group of closely related species (Parker et al., 1998). Although, the development of microsatellite markers may be a deterrent, many

---

*Corresponding author. E-mail: nataqain@ukm.my.Tel: +603-89214550. Fax: +603-89213398.

researchers realized the benefits of microsatellites. Numerous reports detail the processes of marker isolation and characterization from various fish species, such as *Epinephelus quernus* (Rivera et al., 2003), *Plectropomus maculates* (Zhu et al., 2005), *Morone saxatilis* (Rexroad et al., 2006), *Prochilodus lineatus* (Yazbeck and Kalapothakis, 2007), *Leporinus macrocephalus* (Morelli et al., 2007), *Varicorhinus alticorpus* (Chiang et al., 2008), *Johnius belengerii* (Xu et al., 2009), *Trachinotus blochii* (Gong et al., 2008), and *Epinephelus awoara* (Zhao et al., 2009).

This study reports on the isolation of a large number of microsatellite markers from *E. fuscoguttatus* using an enrichment method, followed by the identification and characterization of the markers according to type and repeat unit.

MATERIALS AND METHODS

Sample collection

A fin sample was taken from Fisheries Research Institute (FRI), Besut, Terengganu, Malaysia and was identified as *E. fuscoguttatus* based on the FAO catalogue by Heemstra and Randall (1993). The fin from the fish was preserved in 100% ethanol while transporting to the laboratory.

Library development

Development of the genomic DNA library was done following the Glenn and Schable (2005) enrichment method, with some modifications. Genomic DNA was extracted using MasterPure™ Complete DNA and RNA Purification Kit (Epicentre, USA), based on manufacturer's instructions. Electrophoresis was done on 1% agarose at 80 V for 40 min. The concentration and purity of the extracted genomic DNA was analyzed by NanoDrop™ Spectrophotometer ND-1000 (NanoDrop Technologies, USA). Approximately 10 µg of genomic DNA was digested using *Rsa*I. Oligonucleotide linkers designated as SuperSNX24 (5'-GTTTAAGGCCTAGCTAGCAGAATC-3') and SuperSNX24+4p, with a 5' modification by the addition of phosphate (5'-pGATTCTG CTAGCTAGGCCTTAAACAAAA-3'), were ligated to the digested DNA using Fast-Link™ DNA Ligation Kit (Epicentre, USA). PCR amplification was done using SuperSNX24 as the primer with the following reaction condition: initial denaturation at 95°C for 2 min, 30 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 20 s and elongation at 72°C for 1 min 30 s, and a final extension at 72°C for 5 min. PCR products were visualized by separation on 1% agarose at 80 V for 40 min. The microsatellite enrichment method was done using a 5' terminal biotinylated microsatellite oligo probe according to the protocol of Glenn and Schable (2005): i) $(AG)_{12}$, $(TG)_{12}$, $(AAC)_6$, $(AAG)_8$, $(AAT)_{12}$, $(ACT)_{12}$, $(ATC)_8$, ii) $(AAAC)_6$, $(AAAG)_6$, $(AATC)_6$, $(AATG)_6$, $(ACAG)_6$, $(ACCT)_6$, $(ACTC)_6$, $(ACTG)_6$, and iii) $(AAAT)_8$, $(AACT)_8$, $(AAGT)_8$, $(ACAT)_8$, $(AGAT)_8$. PCR was employed to anneal probes to the digested genomic DNA. The PCR program used included an initial denaturation at 95°C for 5 min, followed by ramp to 70°C, and further steps down of 0.2°C every 5 s for 99 cycles. A final hold at 50°C for 10 min was used. These conditions were followed by ramping down 0.5°C every 5 s for 20 cycles with a final hold at 4°C. Isolation of microsatellite DNA fragments was done using Dynabeads M-280 Streptavidin (Invitrogen, USA), following the manufacturer's instructions. The repeat enriched DNA was eluted from the beads and ethanol precipitated (Sambrook et al., 1989). The product purified after ethanol precipitation was amplified with PCR again, using only SuperSNX24 as the primer. Amplicons were purified using the Wizard® Gel and PCR Clean Up System (Promega, USA), based on the manufacturer's instructions. PCR products were visualized on 1% agarose gels. The enriched, fragmented DNA was cloned into pGEM®-T Easy Vector (Promega, USA), transformed into JM109 competent cells (Promega, USA), and spread onto LB plates containing Ampicillin (100 µg/ml), bromo-chloro-indolyl-galactopyranoside (X-Gal) (80 µg/ml), and Isopropyl β-D-1-thiogalactopyranoside (IPTG) (0.5 mM). White colonies (331) representing possible recombinant plasmids in transformed cells were picked, and 36 random colonies were amplified with PCR using M13 universal primer. Colonies were grown overnight at 37°C and proceeded to plasmid extraction. Plasmids were extracted using the Wizard® Plus SV Miniprep DNA Purification System Kit (Promega, USA) and were sequenced in both directions using M13F and M13R via an ABI PRISM® 3100 DNA Sequencer.

Sequence analysis

Sequences were analyzed using Biology Workbench 3.2 and Microsatellite Repeat Finder 9 http://biophp.org/minitools/ microsatellite_repeats_finder.php). Four sequence analyses were carried out: (i) Ability for forward and reverse sequences to be aligned,(ii) Presence of SuperSNX24 linker at the 5' and 3' terminal, (iii) Identification of the type of microsatellite as mono-, di-, tri-, tetra-, penta- or hexanucleotide, and, (iv) Identification of the repeat successfully isolated as either perfect, imperfect or a compound repeat. Duplicate sequences were eliminated for microsatellite identification. Sequences were analyzed through BLASTn to find similarities within the GenBank Database. Novel sequences were subsequently deposited into the GenBank database (Accession no: GU799128 - GU799322, HM149347).

RESULTS AND DISCUSSION

Library development

The enrichment method was employed for its ability to isolate different types of microsatellite repeats, compared to degenerate primer. Even though the use of degenerate primers is a fast and robust technique, it is limited by its use of only specific repeated sequences (Fisher et al., 1996). Different biotinylated oligo probes were used to target various repeated sequences present in the *E. fuscoguttatus* genome. Glenn and Schable (2005) suggested three mixtures of oligo repeats to be used that would cover different microsatellite repeats present in the genome. The incorporation of the 'GTTT' sequence at the 5'- end of the SuperSNX linker has made cloning more simple and has eliminated a purification step by reducing small PCR products (Glenn and Schable, 2005).

Approximately, 500 colonies were grown on LB plates. Overnight incubation resulted in 331 white colonies. PCR of 13 randomly selected colonies showed differently sized PCR products for each colony, indicating that differently sized fragments were cloned. These size differences occur as a result of random cuts by the restriction enzyme *Rsa*I.

Extracted plasmids were sequenced in both directions using the M13 universal primer. From the 331 sequence

Table 1. Microsatellite distribution in *E. fuscoguttatus* genome.

Category	Repeat Motif	(%)	(%)
Perfect	Mono	9.1	
	Di	40.2	
	Tri	35.8	
	Tetra	10.9	75.2
	Penta	2.2	
	Hexa	1.4	
	Hepta	0.4	
	Total	100	
Imperfect	Mono	11.4	
	Di	72.7	
	Tri	13.6	11.9
	Tetra	2.3	
	Total	100	
Compound	Mono	23.4	
	Di	21.3	
	Mix	55.3	12.9
	Total	100	
Total			100

pairs, 45 sequence pairs were not complimentary and did not carry either of the SuperSNX24 or SuperSNX24+4p linkers. This is often as a result of poor sequencing quality that is likely caused by low plasmid purity after ethanol precipitation. Besides, size fractionation was suggested to be done after digestion in order to reduce the number of low read length sequences. Alignments also demonstrated that 91 sequences were duplicated. Sequences that were unable to be aligned, as well as duplicated sequences, were eliminated during micro-satellite identification.

Microsatellite identification

Microsatellite finder is a useful tool for microsatellite sequence searches. It has the ability to determine whether the sequence was a perfect or imperfect repeat as well as identify the type of microsatellite sequence. After the elimination of unusable sequences, 195 sequences were considered valid because they were able to align between forward and reverse sequences and they carried the linker sequence.

From the 195 sequences, there were 378 microsatellites present. Of these 378, 46 (12.2%) microsatellites were mono-, 175 (46.3%) were di-, 109 (28.8%) were tri-, 36 (9.5%) were tetra-, 7 (1.9%) were penta-, 4 (1.1%) were hexa- and 1 (0.3%) was a heptanucleotide repeat. Table 1 shows the microsatellite distribution in the *E. fuscoguttatus* genome. In brief, identified microsatellites

can also be categorized under type of microsatellite. In this case, 276 sequences (75.2%) were perfect, 44 (11.9%) were imperfect, and 49 sequences (12.9%) were compound repeats. According to repeat motifs, common perfect microsatellites are, for mono, $(T)_n/(A)_n$, at 7.2% and for di-, $(AC)_n/(TG)_n$, at 20.6% of the total perfect dinucleotide microsatellites. These were followed by 12.3% of $(CA)_n/(GT)_n$ and 3.9% of $(GA)_n/(CT)_n$ repeats. Whereas, for tri-nucleotide, repeat motifs, 3.9% for two motifs, $(CTC)_n/(GAG)_n$ and $(CTG)_n/(GAC)_n$, and, for tetra-nucleotide repeat motifs, 1.4% of $(GGGA)_n/(CCCT)_n$ and 0.7% for both $(CAAA)_n/(GTTT)_n$ and $(AAAG)_n/(TTTC)_n$, respectively. The longer the repeat motif, the less chance it can occur in the genome of *E. fuscoguttatus*. From 378 microsatellites identified, only 162 were suitable to be used in designing primer for future study.

Based on microsatellite data accessed through the GenBank Database (until 6 April 2011), there are three other research groups also focusing on *E. fuscoguttatus*. The first group (unpublished data) (accession no: GQ912319 - GQ912328) was able to isolate more trinucleotide repeats (81.8%) than we did because the used of a 5'-anchored degenerate primer. A second group (Lo and Yue, 2008) (accession no: EU016533 - EU016545) was able to isolate more dinucleotide repeats (55.2%) than we did because they employed an enrichment method. The third group (Koedprang et al., 2007) (accession no: AY736039) only deposited one sequence containing a dinucleotide repeat. A total of 63 microsatellites were deposited in GenBank; of these, 2

were mono-, 21 were di-, 35 were tri- and 5 were tetranucleotides. These can also be differentiated according to type: 32 were perfect, 9 were imperfect, and 11 were compound microsatellites.

Our study supported the findings of Lo and Yue (2008), who successfully isolated the most abundant microsatellite in the *E. fuscoguttatus* genome, the dinucleotide repeat, via the same method. Even though this method of study can only isolate microsatellites that are randomly selected from the genome, it has proven that the *E. fuscoguttatus* genome is rich with dinucleotide repeats, $(AC)_n$, in comparison with other vertebrate organisms.

Supported by Bhassu et al. (2008), it was delineated that 30 - 67% of microsatellites found in the fish genome are dinucleotide. The $(AC)_n$ dinucleotide motif is the most common in all vertebrate genomes, generally 2.3-fold more frequent than $(AT)_n$, which is the second most common type of microsatellite (Toth et al., 2000). From this finding, we concluded that the *E. fuscoguttatus* genome is also abundant in dinucleotide repeats, with the most frequent motif being $(AC)_n$. In general, the higher order of SSR classes that are tri-, tetra-, and hexanucleotides are about 1.5-fold less common in genomic DNA of vertebrates than dinucleotides (Toth et al., 2000). In comparison, in the Japanease pufferfish or in the fugu *Takifugu rubripes*, the most frequent microsatellites are dinucleotide repeats (34%), followed by tetra- (21%), tri- (19%), mono- (16.5%), hexa- (6%) and pentanucleotides (3%) (Edwards et al., 1998). Indeed, 1.29% of the Japanese pufferfish genome consists of microsatellites (Bhassu et al., 2008). Whereas, in the closely related spotted green pufferfish *Tetraodon nigroviridis*, SSRs covered 3.21% of the genome (Crollius et al., 2000). Chiang et al. (2008) managed to isolate 8 dinucleotide microsatellites: 4 of the $(CA)_n$ motif, 2 of the $(TG)_n$ motif, 1 of the $(TC)_n$ motif, and 1 of the $(TA)_n$ motif from *Varicorhinus alticorpus*. Further, Zhao et al. (2009) isolated 12 dinucleotide microsatellites, of which 5 were of the $(CA)_n$ motif, 5 were of the $(GT)_n$ motif, and 2 were of the $(TG)_n$ motif from *Epinephelus awoara*. Rivera et al. (2003) isolated microsatellites from *Epinephelus quernus* and presented 20 clones that contained microsatellite repeat motifs, 15 of which were dinucleotide repeats and mainly $(CA)_n$; the remainder were trinucleotides. Generally, in all vertebrates, (G+C) rich motifs are the most common among trinucleotide repeats and are usually prominent in exon and less common in intron regions (Toth et al., 2000).

Conclusion

The *E. fuscoguttatus* genome was similar to other vertebrate genomes in that it contained many dinucleotide repeats; the most common dinucleotide repeat found was $(AC)_n$. The identification of microsatellites is one of the most important steps in developing a microsatellite marker that can be used in many applications, such as genetic population studies, broodstock selection programs, and for the future construction of linkage maps. However, characterization was not done since this is only a preliminary study of *E. fuscoguttatus* genome.

REFERENCES

Bhassu S, Bakar Y, Rashi ZA (2008). Isolation of Dinucleotide Microsatellite Markers in Commercially Threatened Malaysian Mahseer *Tor tambroides*. Russ. J. Genet., 44: 1000-1002.

Bombeo-Tuburan I, Coniza EB, Rodriguez EM, Agbayani RF (2001). Culture and economics of wild grouper (*Epinephelus coioides*) using three feed types in ponds. Aquaculture, 201: 229-240.

Braise (1987). Grouper abstracts, SEAFDEC Aquaculture Dept. Tigbauan, Iloilo, Philippines, p. 95.

Chiang TZ, Lee TW, Lin FJ, Huang KH, Lin HD (2008). Isolation and characterization of microsatellite loci in the endangered freshwater fish *Varicorhinus alticorpus* (Cyprinidae). Conserv. Genet., 9: 1399-1401.

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.

Cornish A (2004). *Epinephelus fuscoguttatus*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.2. www.iucnredlist.org (July 14, 2010).

Crollius HR, Jaillon O, Dasilva C, Ozouf-Costaz C, Fizames C, Fischer C, Bouneau L, Billault A, Quetier F, Saurin W, Bernot A, Weissenbach J (2000). Characterization and Repeat Analysis of the Compact Genome of the Freshwater Pufferfish *Tetraodon nigroviridis*. Genome Res., 10: 939-949.

Edwards YJ, Elgar G, Clark MS, Bishop MJ (1998). The identification and characterization of microsatellite in the compact genome of the Japanese pufferfish, *Fugu rubripes*: Perspective in functional and comparative genomic analyses. J. Mol. Biol., 278: 843-854.

Fisher PJ, Gardner RC, Richardson TE (1996). Single locus microsatellites isolated using 5' anchored PCR. Nucleic Acids Res., 24: 4369-4371.

Froese R, Pauly D (2010). FishBase. World Wide Web electronic publication. www.fishbase.org, version (03/2010).

Glenn TC, Schable NA (2005). Isolating microsatellite DNA loci. Method Enzymol., 395: 202-222.

Gong P, Li J, Yue GH (2008). Twelve novel microsatellite loci from an endangered marine fish species golden pompano *Trachinotus blochii*. Conserv. Genet., 10: 1365-1367.

Heemstra PC, Randall JE (1993). FAO Species Catalogue. Groupers of the world (Family Serranidae, Subfamily Epinephelinae). An annotated and illustrated catalogue of the grouper, rockcod, hind, coral grouper and detail species known to date. Food and Agriculture Organization of the United Nations, Fisheries Synopsis, 16(125): p. 382.

INFOFISH (1989). Markets for reef fish in Singapore and Hongkong. INFOFISH Int. 1: 17-19.

Koedprang W, Na-Nakorn U, Nakajima M, Taniguchi N (2007). Evaluation of genetic diversity of eight grouper species Epinephelus spp. based on microsatellite variations. Fish. Sci., 73: 227-236.

Lin YH, Shiau SY (2003). Dietary lipid requirement of grouper, *Epinephelus malabaricus*, and effects on immune responses. Aquaculture, 225: 243-250.

Liu ZJ, Cordes JF (2004). Review: DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238: 1-37.

Lo LC, Yue GH (2008). Microsatellites for broodstock management of the Tiger grouper, *Epinephelus fuscoguttatus*. Anim. Genet., 39: 90-91.

Morelli K, Revaldaves E, Oliveira C, Forest F (2007). Isolation and characterization of eight microsatellite loci in *Leporinus macrocephalus* (Characiformes: Anostomidae) and cross-species amplification. Mol. Ecol. Notes, 7: 32-34.

Parker PG, Snow AA, Schug MD, Booton GC, Fuerst PA (1998). What molecules can tell us about populations: choosing and using a molecular marker. Ecology, 79: 361-382.

Rexroad C, Vallejo R, Coulibaly I, Couch C, Garber A, Westerman M, Sullivan C (2006). Identification and characterization of microsatellites for striped bass from repeat-enriched libraries. Conserv. Genet., 7: 971-982.

Rivera MA, Graham GC, Roderick KG (2003). Isolation and Characterization of Nine Microsatellite Loci from the Hawaiian Grouper *Epinephelus Quernus* (Serranidae) for Population Genetic Analyses. Mar. Biotechnol., 5: 126-129.

Sambrook J, Fritsch EF, Maniatis T (1989). Molecular cloning: a laboratory manual. Second edition. Cold Spring Harbor Press, New York, USA.

Tookwinas S (1989). Review of knowledge on Grouper aquaculture in Southeast Asia. Advances in Tropical Aquaculture , 20 Feb-4 Mac, Tahiti.

Toth G, Gaspari Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: Survey Anal. Genome Res., 10: 967-981.

Xu G, Shao C, Liao X, Tian Y, Chen S (2009). Isolation and characterization of 12 dinucleotide microsatellite loci from Belenger's jewfish (*Johnius belengerii*, Cuvier 1830). Conserv. Genet., 10: 1009-1011.

Yazbeck GM, Kalapothakis E (2007). Isolation and characterization of microsatellite DNA in the piracema fish *Prochilodus lineatus* (Characiformes). Genet. Mol. Res., 6: 1026-1034.

Zhao L, Shao C, Liao X, Ma H, Zhu X, Chen S (2009). Twelve novel polymorphic microsatellite loci for the Yellow grouper (*Epinephelus awoara*) and cross-species amplifications. Conserv. Genet., 10: 743-745.

Zhu ZY, Lo LC, Xu YX, Yue GH (2005). Isolation and characterization of polymorphic microsatellites from red coral grouper (*Plectropomus maculatus*). Mol. Ecol. Notes, 5: 579-581.