The Open Access Israeli Journal of Aquaculture – Bamidgeh

Analysis and Characterization of Expressed Sequence Tag (EST) and Simple Sequence Repeats (SSR) from Red Sea Bream, *Chrysophrys major*, and Gilthead Sea Bream, *Sparus aurata*

Yun-Guo Liu[1, 2*], Song-Lin Chen[3], Chun-Ying Liu[4], and Fang-Zheng Li[4]

[1] *Fishery College, Ocean University of China, Qingdao 266003, China*

[2] *Agricultural Product Inspection Center, Shandong Entry-Exit Inspection and Quarantine Bureau, Qingdao 266002, China*

[3] *Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, China*

[4] *Qingdao Agricultural University, Qingdao 266109, China*

Abstract

A total of 101 and 119 simple sequence repeats (SSR) were obtained from expressed sequence tags (EST) of the red sea bream (*Chrysophrys major)* and gilthead sea bream (*Sparus aurata*), respectively. Dimeric repeats were the most abundant class of SSR in both red (70.30%) and gilthead (51.26%) sea bream. Trimeric, tetrameric, and pentameric repeats occurred in decreasing proportions, i.e., 21.36%, 6.8%, and 0.97% in red sea bream and 45.38%, 3.36%, and 0 in gilthead. There were no hexameric repeats in EST of either species. The frequency of every class of SSR (dimeric, trimeric, tetrameric, and pentameric) decreased with increasing repeat length. In cross-species amplification of 20 existing microsatellite loci from gilthead sea bream in 20 red sea bream specimens, twelve loci showed positive amplification and four loci showed polymorphic amplification with two or three alleles.

Introduction

Microsatellites or simple sequence repeats (SSR) are stretches of DNA consisting of tandemly-repeated units of 2-6 base pairs in length. They are abundant, codominant, ubiquitous in eukaryotic genomes, and highly polymorphic. Microsatellites have significantly influenced genome mapping and behavioral ecology; they have become the markers of choice for a wide spectrum of molecular genetic, population genetics, and evolutionary biology studies (Jarne and Lagoda, 1996; Powell et al., 1996; Bulle et al., 2002). Microsatellites also provide molecular tools to understand spatial relationships between

---

* Corresponding author. Mailing address: Fishery College, Ocean University of China, Qingdao 266003, China, tel.: +86-86770612, e-mail: liuyg@ysfri.ac.cn

chromosome segments, which aid in analyzing temporal relationships between species and genera (Kashi et al., 1997).

Studies on various organisms provide evidence that repeat sequences in a genome can drastically vary among taxa in SSR length, base composition, mutation rate, and chromosomal distribution (Subramanian et al., 2003). Because of current emphasis on functional genomics, expressed sequence tag (EST) sequences of a large number of species are being accumulated in databases at a very high rate. SSR can be identified from information in the EST databases, significantly reducing the costs of obtaining SSR markers. Study of the distribution pattern of SSR in the genome will likely be helpful in understanding their significance.

There is much evidence to suggest that SSR regulate gene expression (Kunzler et al., 1995; Moxon and Wills, 1999). The reason why microsatellites are special becomes obvious with a brief look at a sequencing gel. While other sequences generally produce a scrambled pattern on such a gel, microsatellites are quickly identified by their simple structure. Microsatellites have a simple internal repeat structure, with repeat units ranging 1-6 bp. The prototype of a microsatellite has a single type of repeat, such as (AT)n. Microsatellite variability results from a gain or loss of repeat units. The difference in repeat number can rapidly be identified by PCR, using sequences that flank the microsatellite as primers.

The red sea bream (*Chrysophrys major*), widely cultured in northern China and highly valued for its good flavor, and the gilthead sea bream (*Sparus aurata* L.), important in aquaculture in southern Europe, are interesting prospective experimental models for genome studies that will provide ready access to genomes of other commercially important Sparidae (Sarropoulou et al., 2007). Therefore, they are important subjects for molecular genetics and genome research. Microsatellites may serve as an invaluable tool for this purpose. However, as far as we know, there are no reports concerning red and gilthead sea bream SSR analysis and a microsatellite conservation study between them. Thus, we investigated the repeat type and frequency of red and gilthead sea bream on EST-SSR and cross-species microsatellite amplification.

Materials and Methods

EST-SSR screening. The 2010 red sea bream EST used in the present paper were taken from Chen et al. (2004, 2005). The 2207 gilthead sea bream EST used in the present paper were taken from the GenBank database at the National Center for Biotechnology Information (NCBI), accession nos. CB176660-177159, CB184056-184950 (Sarropoulou et al., 2005), CX244458-244544, CX734847-735033, DN048387-048410, and CV133223-133736. SSR were screened from these EST using Tandem Repeats Finder, version 2.02 (Benson, 1999), according to the following criteria: seven repeats for a dinucleotide repeat, five for a trinucleotide repeat, four for a tetranucleotide repeat, three for a pentanucleotide repeat, and two for a hexanucleotide repeat. The rationale for choosing the small cutoff value is that SSR are often disrupted by single base substitutions (Subramanian et al., 2003).

Genomic DNA extraction. DNA was extracted according to standard protocol described by Liu et al. (2005) with some modifications. Each blood sample (100 µl) was collected with a 1-ml syringe and immediately expelled into a tube containing 500 µl DNA extraction buffer (100 mM NaCl, 10 mM Tris, pH 8.0, 25 mM EDTA, 0.5% SDS, and freshly added 0.1 mg/ml proteinase K). Blood was quickly expelled into lysis buffer to disperse the blood cells. The lysates were incubated at 55°C overnight. DNA was extracted twice with phenol and once with chloroform. DNA was precipitated by adding half the original blood volume of 7.5 M ammonium acetate and two volumes of ethanol. DNA was collected by brief centrifugation and washed twice with 70% ethanol, air-dried, and dissolved in TE buffer. The concentration was measured with a GeneQuant *pro* (Pharmacia Biotech Ltd.) RNA/DNA spectrophotometer for absorption at 260 nm.

Cross-species microsatellite amplification. Primer sets for 20 gilthead sea bream

microsatellite loci were designed to amplify 20 red sea bream specimens. The sequence of primers, microsatellite core sequence, optimum PCR amplification conditions, and specific annealing temperature of each primer set are showed in Results. PCR protocols followed those of Liu et al. (2005) with some modifications. PCR amplification was carried out in a 25 μl reaction mixture that included 10 pmol of each primer set, 100 μM of dNTPs, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl$_2$, 1 unit of Taq polymerase, and approximately 100 ng of template DNA. Amplifications were performed on a Peltier Thermal Cycler (PTC-200). PCR cycles were as follows: 5 min preamplification denaturation at 94°C, 35 cycles of 45 s at 94°C, 40 s at a primer-specific annealing temperature, and 1 min at 72°C. As a final step, products were extended for 5 min at 72°C.

After PCR, an equal volume of formamide dye was added to each reaction. The samples were heated to 95°C to denature for 5 min and immediately placed on ice. The final PCR products were run on a 5% denaturing polyacrylamide gel in TBE buffer. For separation of microsatellite markers, a gel with uniform 0.4-mm thickness using a comb with "shark teeth wells" was used. Analysis was carried out by silver staining according to Liu et al. (2004) with modifications and overnight drying before the final products were photographed.

Results

A total of 101 and 119 SSR were obtained from the red sea bream and gilthead sea bream EST, respectively. The dimeric repeats were the most abundant class of SSR in both species with trimeric, tetrameric, and pentameric repeats represented in decreasing proportions (Table 1). No hexameric repeats were observed in EST from either bream. When classified into two categories (≤10 and >10), the ≤10 repeat unit constituted a large majority of the total SSR (Table 2). When grouped in four classes, the dimeric repeat combinations AG and AC were most abundant (Fig.1). When grouped into ten classes, the AAT, AAG, and AAC trimeric repeat combinations were predominant in both species.

Table 1. Frequency (%) of different repeat types from dimer to pentamer in red sea bream and gilthead sea bream expressed sequence tags (EST).

	Red sea bream	Gilthead sea bream
Di	70.30	51.26
Tri	21.36	45.38
Tetra	6.80	3.36
Penta	0.97	0

Table 2. Distribution of simple sequence repeats (SSR) of different repeat units in red sea bream and gilthead sea bream expressed sequence tags (EST).

	Red sea bream	Gilthead sea bream
≤10 di	53	49
>10 di	18	12
≤10 tri	20	50
>10 tri	2	4
≤10 tetra	7	4
>10 tetra	0	0
≤10 penta	1	0
>10 penta	0	0
Total ≤10	81(80.2%)	103(86.6%)
Total >10	20(19.8%)	16(13.4%)

In cross-species amplification of 20 existing microsatellite loci from gilthead sea bream, twelve loci were positive in 20 red sea bream individuals (Table 3). The other eight produced no amplification or only a smear while four loci showed polymorphic amplification with two or three alleles.

Discussion

While classifying the SSR into repeat types or categories, the complementary sequence was also considered. For example, the repeat motifs AG and GA were put in the same class,
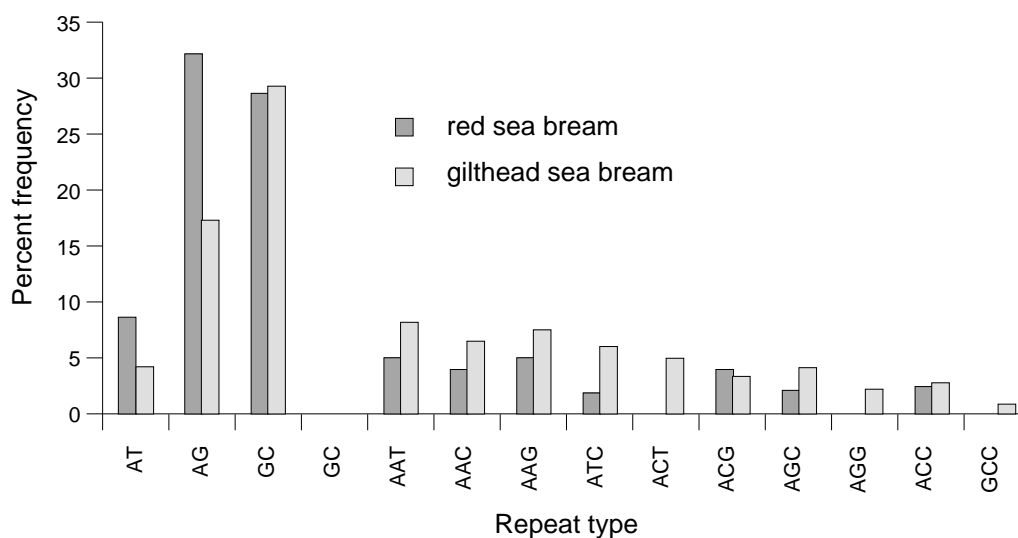
Fig. 1. Frequency (%) of different simple sequence repeats (SSR) repeat types in red sea bream and gilthead sea bream expressed sequence tags (EST).

as were TC and CT. The amount of different SSR unit sizes was not the same in red and gilthead sea bream EST. Dimeric, trimeric, tetrameric, and pentameric repeats appear in decreasing proportions. Most dimeric repeats were distributed in 3' and 5' un-translation regions (UTR).

If translation regions are considered, trimeric repeats were the highest in both red and gilthead sea bream EST. Gilthead sea bream had a higher proportion of trimeric repeats than red sea bream because more translation regions were sequenced in the gilt-head sea bream EST, in agreement with ear-lier findings about abundance of SSR unit size classes (Temnykh et al., 2001). This domi-nance of trimeric SSR may be explained by the suppression of non-trimeric-times SSR in coding regions due to the risk of frame shift mutations that may occur when SSR alternate in size of one unit (Metzgar et al., 2000).

We confirmed earlier observations that the frequency of every class of SSR decreases with increasing repeat length (Temnykh et al., 2001). Among the four types of dimeric repeats, AT, AG, and AC were far more abun-dant than GC in EST libraries of both species,

findings that agree with previous observations about differences for monomer repeats (Katti et al., 2001; Subramanian et al., 2003). Interestingly, GC dimeric repeats are extremely rare in the cDNA library. The lower frequency of CpG dinucleotides in vertebrate genomes has been attributed to methylation of cytosine, which increases its chances of mutation to thymine by deamination (Schorderet et al., 1992). However, CpG sup-pression by this mechanism can not explain the rarity of (GC)n dinucleotide repeats in invertebrates, since they do not show cytosine methylation (Katti et al., 2001).

The proportion of motifs differed among the trimeric repeats. AAT, AAG, and AAC were the most common in both species, GCC repeats were extremely rare among trimers. The abundance of trinucleotide repeats in the coding regions can be partially limited by selection at the protein level (Chakraborty et al., 1997; Katti et al., 2001). However, differ-ent abundances of trimeric repeats have been reported for different species (Katti et al., 2001; Temnykh et al., 2001; Subramanian et al., 2003) suggesting that, in addition to an alternate DNA structure formed by repeat

Table 3. Cross-species amplification of 20 microsatellite loci from gilthead sea bream (*Sparus aurata*) in 20 red sea bream (*Chrysophrys major*) individuals.

Accession no.	Repeat sequence	Primer sequences (5'-3')	Ta (°C)	Amplification (no. observed alleles)
CB184473	$(AT)_{10}$	F: TGAGACCACCGTAGACATTT	50	1
		R: GGTCAAGAGCGTCTGATAAA		
CB184424	$(AT)_3GAC(AT)_3GG(AT)_3AC(AT)_8$	F: GCTACCACTAGCTCCCAAAG	52	*
		R: GAACCAGATGCCAGGAATAA		
CB177073	$(AT)_8$	F: AGTTGTTGATTATGATGTAT	41	1
		R: GTGTCTTTAATACTTACAGATA		
CB176854	$(AT)_{10}$	F: TGCTTTCATCTAACCAATGGGACT	58	1
		R: ACATGGCCGCCACCAACATC		
CV133485	$(AC)_{11}(GC)_2(AC)_2$	F: CCCCTTAATGAAGTTCTAAT	48	1
		R: ATGTACGAAATAATGCCTGT		
CV133313	$(AC)_{15}$	F: AAGAAATCTAAATCAGCAGGAC	51	3
		R: CTGTTTGTAAAGTCGCCATC		
CX734972	$(AC)_7GC(AC)_7$	F: GATAGTTGAATGCAAACGTC	48	*
		R: AGTGCTGAATTAAAGGTCTC		
CX734877	$(AC)_7CCCC(AC)_3$	F: GAGTTTCTGACTGAGCGTGGAG	55	*
		R: GGCGTAAGGAAGGGACATCA		
CX734870	$(AC)_{22}$	F: AACGCCTTCTAGCGGCACAG	57	3
		R: CACCCTTGACCACAGGTTACAACA		
DN048405	$(AC)_{28}$	F: GTGACGGCTGCAGGACAAGA	56	2
		R: TTAGCGACCCATTTCTGACG		
DN048405	$(TG)_{11}GG(TG)_{12}$	F: TTCTTCGTCCTCGGCTGTAA	52	1
		R: CTCTGCTCGTGTCTCCTGAA		
DN048405	$(TG)_{12}(AT)_7(TG)_3$	F: TTCAGGAGACACGAGCAGAG	53	*
		R: TTGCACGAGTCCAGATGTAAAG		
CB184152	$(AC)_8GC(AC)_5$	F: CCCATAATGAGCAGATAAAAG	51	*
		R: GAATGATGCGAGGTCCACAA		
CB184076	$(AC)_7$	F: CCTCACATGCAGCCTTCACC	56	*
		R: CAACGCAGTCTTCCTCCACA		
CB177098	$(AC)_7$	F: ATCAAAGCTGAATGTAGGAG	49	1
		R: AGAATGACTCAAGAGTGCTG		
CB177002	$(AC)_7$	F: AAAGCCAAAGAACAGGAGCG	55	*
		R: GAGGCAACTTGGTCAACAGC		
CB176779	$(AC)_7$	F: TGGAAGGACAGATAGGACAT	50	*
		R: CAAATCAACAAATCGTAGGG		
CB176713	$(AC)_{11}$	F: AACATGGACTGGAGGAACTT	51	1
		R: GAGGCATCAGGTGGTCATAG		
CB184888	$(AG)_7$	F: TGTAAGCAGTCGGCTGTGGT	53	1
		R: TGGAGCCTTTCACTGACATA		
CB184149	$(AG)_{12}$	F: CCGTGTCCTCCAGCCAAGTT	55	2
		R: GTGCTCCACATTCACAGTCC		

* no amplification, or smear only

motifs, species-species cellular factors interacting with them likely play an important role in the genesis of repeats (Toth et al., 2000).

Predominant microsatellite variability occurs by a process known as DNA slippage (Schlotterer and Tautz, 1992; Weber and Wong, 1993; Wierdl et al., 1997). The current model of DNA slippage assumes that, during DNA synthesis, the elongating strand misaligns in the microsatellite region; when DNA synthesis continues along the misaligned strand, the gain or loss of microsatellite repeats on the synthesized strand leads to a loop structure (Henderson and Petes, 1992). Surprisingly, within one class of repeats, there may be many differences in abundance of a particular sequence repeat in the red and gilthead sea bream EST libraries. The density of poly (A) or poly (T) is far more than that of poly (G) or poly (C) in all repeats. Such a situation also exists in the human genome (Bowcock et al., 1994; Garza et al., 1995; Kunzler et al., 1995). It is possible that during SSR evolution, the poly (A) stretches present in the genome mutate to produce A-rich repeats. It is also possible that the abundance of repeats is influenced by their secondary structures and the effect on DNA replication. If a repeat sequence is selected during evolution for transcriptional regulation or is a target of a binding protein for one or more nuclear processes (such as chromatin organization, DNA replication, transcription, and recombination), its abundance and distribution are expected to be controlled (Subramanian et al., 2003).

Microsatellite loci generally show considerable evolutionary conservation, which suggests that microsatellite primers developed for any one locus may be useful across a wide range of taxa. Cross-species amplification is effective only if primer sequences are conserved between species. Generally the number of loci amplifying tends to decrease with increasing divergence between species (Moore et al., 1991; Peakall et al., 1998). Red sea bream and gilthead sea bream are both in the Sparidae family. Of the twenty loci tested in this study, twelve fulfilled the criteria of amplifying well, and four exhibited levels of polymorphism. These results show evolutionary conservation between

red sea bream and gilthead sea bream for some microsatellite loci. They also indicate cross-species application of known microsatellite loci in closely related species is a highly promising source of microsatellite markers.

References

Benson G., 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.*, 27:573-580.

Bowcock A.M., Ruiz-Lineares A., Tonfohrde J., Minch E., Kidd J.R. and L.L. Cavalli-Sforza, 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature,* 368:455-457.

Bulle B., Millon L., Bart J.M., Gallego M., Gambarelli F., Portus M., Schnur L., Jaffe C.L., Fernandez-Barredo S., Alunda J.M. and R. Piarroux, 2002. Practical approach for typing strains of *Leishmania infantum* by microsatellite analysis. *J. Clin. Microbiol.*, 40:3391-3397.

Chakraborty R., Kimmel M., Stivers D.N., Davison L.J. and R. Deka, 1997. Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA*, 94:1041-1046.

Chen S.L., Xu M.Y., Hu S.N. and L. Li, 2004. Analysis of immune-relevant genes expressed in red sea bream (*Chrysophrys major*) spleen. *Aquaculture*, 240:115-130.

Chen S.L., Liu Y.G., Xu M.Y. and J. Li, 2005. Isolation and characterization of polymorphic microsatellite loci from an EST-library of red sea bream (*Chrysophrys major*) and cross-species amplification. *Mol. Ecol. Notes*, 5:215-217.

Garza J.C., Slatkin M. and N.B. Freimer, 1995. Microsatellite allele frequencies in humans and chimpanzees with implications for constraints on allele size. *Mol. Biol. Evol.*, 12:594-603.

Henderson S.T. and T.D. Petes, 1992. Instability of simple sequence DNA in *Saccharomyces cerevisiae. Mol. Cell Biol.,* 12:2749-2757.

Jarne P. and P.J.L. Lagoda, 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.,* 11:424-429.

Kashi Y., King D. and M. Soller, 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.,* 13:74-78.

Katti M.V., Ranjekar P.K. and V. S. Gupta, 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.,* 18:1161-1167.

Kunzler P., Matsuo K. and W. Schaffner, 1995. Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. *Biol. Chem. Hoppe-Seyler*, 4:201-211.

Liu Y., Wang X. and L. Liu, 2004. Analysis of genetic variation in surviving apple shoots following cryopreservation by vitrification. *Plant Sci.*, 166:677-685.

Liu Y., Chen S. and B. Li, 2005. Assessing the genetic structure of three Japanese flounder (*Paralichthys olivaceus*) stocks by microsatellite markers. *Aquaculture*, 243:103-111.

Metzgar D, Bytof J. and C. Wills, 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.*, 10:72-80.

Moore S.S., Sargeant L.L., King T.J., Mattick J.S., Georges M. and D.J.S. Hetzel, 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics,* 10:654-660.

Moxon E.R. and C. Wills, 1999. DNA microsatellites: Agents of evolution? *Sci. Am.,* 280:94-99.

Peakall R., Gilmore S., Keys W., Morgante M. and A. Rafalski, 1998. Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. *Mol. Biol. Evol.*, 15:1275-1287.

Powell W., Machray G.C. and J. Provan, 1996. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.*, 1:215-222.

Sarropoulou E., Power D.M., Magoulas A., Geisler R. and G. Kotoulas, 2005. Comparative analysis and characterization of expressed sequence tags in gilthead sea bream (*Sparus aurata*) liver and embryos. *Aquaculture*, 243:69-81.

Sarropoulou E., Franch R., Louro B., Power D.M., Bargelloni L., Magoulas A., Senger F., Tsalavouta M., Patarnello T., Galibert F., Kotoulas G. and R. Geisler, 2007. A gene-based radiation hybrid map of the gilthead sea bream *Sparus aurata* refines and exploits conserved synteny with *Tetraodon nigroviridis. BMC Genomics*, 8:44.

Schlotterer C. and D. Tautz, 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.,* 20:211-215.

Schorderet D.F. and S.M. Gartler, 1992. Analysis of CpG suppression in methylated and nonmethylated species. *PNAS*, 89:957-961.

Subramanian S., Mishra R.K. and L. Singh, 2003. Genome-wide analysis of microsatellite repeats in human: Abundance and their density in specific genomic regions. *Genome Biol.*, 4:R13.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S. and S. McCouch, 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.*, 11:1441-1452.

Toth G., Gaspari Z. and J. Jurka, 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.*, 10:967-981.

Weber J.L. and C. Wong, 1993. Mutation of human short tandem repeats. *Human Mol. Genet.*, 2:1123-1128.

Wierdl M., Dominska M. and T.D. Petes, 1997. Microsatellite instability in yeast: Dependence on the length of the microsatellite. *Genetics*, 146:769-779.