African Journal of Biotechnology Vol. 10 (4), pp. 507-518, 24 January, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.482
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Isolation and analysis of microsatellites in the genome of turbot (*Scophthalmus maximus* L.)

Xiaohong Ruan[1,2], Weiji Wang[2], Jie Kong[2]* and Jingjie Hu[1]

[1]Ocean University of China, Qingdao 266003, People's Republic of China.
[2]Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, People's Republic of China

Accepted 8 December, 2010

A microsatellite enrichment libraries construction colony hybridization strategy was used for large-scale isolation of microsatellites from the whole genome of the turbot (*Scophthalmus maximus* L.). Using 14 different probes, paired-end sequencing of 952 clones resulted in 753 positive and individual clones. Using Vector NTI Suite 8.0 and tandem repeats finder software, 1004 tandem repeat sequences were found, including 831 microsatellites and 173 minisatellites, which accounted for 82.77 and 17.23% of the total repeat sequences, respectively. From the microsatellites, 158 polymorphic microsatellite markers were developed. A detailed analysis of the different nucleotide units across the whole genome of turbot was also presented, the dinucleotide unit was the dominant repeat unit among the six repeat unit types (as represented by number of repeats and the repeat unit cumulative length), while the pentanucleotide type was the most frequently found. The AGG (126.11) motif was the most variable repeat unit type in this research with a coefficient of variation value of 126.11. In this study, a negative correlation between the length of repeat units and the average copy number was detected (r = -0.5, P = 0.072).

Key words: Microsatellite, turbot, enrichment library, primer.

INTRODUCTION

Microsatellites, also called simple sequence repeats (SSRs), are tandemly repeated sequences of 1 to 6 bases in length. They are ubiquitous in prokaryotes and eukaryotes (Tóth et al., 2000). Analysis of SSRs has suggested that such repeats are important in genomic function and might be associated with some diseases (Subramanian et al., 2003; International Human Genome Sequencing Consortium, 2001). Simple sequence repeats reveal high levels of polymorphism and their convenient amplification by polymerase chain reaction (PCR) assays make them excellent molecular markers (Powell et al., 1996). Among DNA markers, microsatellites perform better than other PCR-based methods, such as random amplified polymorphic DNA (RAPD) (Williams et al., 1990) and amplified fragment length polymorphism (AFLP) (Vos et al., 1995), because they show Mendelian inheritance and co-dominance. Presently, in fishery and aquaculture, SSRs are widely used in the construction of molecular maps, in evolutionary studies (Danzmann and Gharbi, 2001; Kenchington et al., 2006), genetic diversity analysis (Coughlan et al., 1996), the identification of parentage (Blouin, 2003; Norris et al., 2000; Selvamani et al., 2001), the analysis of disease-linkage and in marker-assisted breeding programs (Chistiakov et al., 2006). SSRs have become one of the most commonly used molecular markers. Moreover, the relative conserved flanking sequences of SSRs widened the opportunity of universal application of SSRs among different researchers and laboratories.

The turbot (*Scophthalmus maximus* L.) is a marine flatfish of great commercial value. This flatfish is found along virtually the entire European coast, from Norway to Morocco and across the northern coasts of the Mediterranean, up to the Black Sea (Blanquer et al., 1992). Due to its easy domestication, commercial value and appealing taste, turbot represents one of the most promising marine species in European aquaculture and was

*Corresponding author. E-mail: kongjie@ysfri.ac.cn. Tel: 86 532 85823291.

introduced into China in 1992. Accompanied by certain key technological breakthroughs, especially artificial breeding, the turbot farming industry developed rapidly and turbot has become one of the most flourishing marine species in the coast of Northern China in the last 10 years (Lei and Liu, 1995, Lei and Zhan, 2001, Lei, 2002 and Lei et al., 2003). In 2006, according to official statistical data, its domestic production was 50000 tons and its export value was more than three billion yuan. At the same time, genetic breeding and improvement did not receive enough attention, which could be observed in the loss of genetic diversity, increased inbreeding and the appearance of albino turbot in culture stocks (Lei et al., 2005; Shen et al., 2004). Thus, some measures must be taken to avoid inbreeding depression and effective breeding programs should be carried out to increase the turbot yield in China. Although traditional pedigree tracing based on physical targeting could avoid inbreeding depression, it is currently not practical for large-scale fish farming. However, highly informative SSR markers could facilitate pedigree tracing and parentage determination. Moreover, marker assisted selection (MAS) technologies effectively complement traditional pedigree tracing. Highly informative SSRs could facilitate MAS progress of turbot through linkage mapping and quantitative loci (QTL) analysis. Nevertheless, it is generally accepted that difficulties in SSR development limit its widespread adoption. Although a few SSR sequences of turbot have been reported, it was only recently that an efficient and labor-saving method was reported (Iyengar et al., 2000; Bouza et al., 2002; Pardo et al., 2005, 2006; Liu et al., 2006). Prior to 2006, less than 40 microsatellite sequences were reported and were available in GenBank for the turbot (DQ365892–DQ365896, AY641450–AY641453, AY697865, AF182080–182100, and Z78097–Z78101). Recently, 12 polymorphic microsatellite sequences were developed from a cDNA library from the turbot (Chen et al., 2007). Two hundred and forty-eight novel microsatellite markers were isolated from enriched genomic libraries in turbot, constructed in accordance with the FIASCO (Fast Isolation by AFLP of Sequences containing repeats) protocol (Pardo et al., 2007).

Presently, five methods and strategies have been used to isolate microsatellite sequences: PCR screening; colony hybridization; microsatellite enrichment libraries construction; FIASCO; EST database mining. Among these, EST database mining has proven to be an efficient and low-cost approach; however, the small number of ESTs (3656 sequences) in turbot and low ratio (2.2%) of microsatellite-containing ESTs hampers the isolation of a large numbers of markers (Chen et al., 2007). Recently, Zhan et al. (2007) reported that, the method of modified screening SSR enriched libraries was also efficient and suitable for isolating a large amount of microsatellite markers for the target species of interest. This method has been used in many species, such as bay scallop (Li et al., 2007), sand lance (Ren et al., 2009) and sea cucumber (Hu et al., 2007). In this work, we isolated and characterized microsatellite markers of turbot according to the methods of Zhan et al. and investigated the base composition and distribution of tandem repeats.

MATERIALS AND METHODS

A total of turbot individuals sampled from Haiyang City (Shandong 30Province) in China, were employed for polymorphism assessment. DNA was extracted using the method of Zou et al. (2001). DNA quantity was determined with UV spectrophotometry. Microsatellite DNA enrichment was performed essentially according to the methods of Karagyozov et al. (1993), Edwards et al. (1996) and Zhan et al. (2007), with some modifications. The genomic DNA was digested with *Hae*III and the fragments (400 and 1200 bp) were recovered from agarose gels. The products were ligated to oligonucleotide adaptors (21-mer: 5'-CTCTTGCTTGAATTCGGA CTA-3′ and phosphorylated 25-mer: 5′-pTAGTCCGAATTCAA GCAAGAGCACA-3′) with T4 DNA ligase at 16°C for 12 h. Ligation products were amplified by PCR using 21-mer adaptor, then microsatellite-containing fragments were captured on nylon membranes (Hybond N^+) bound with different probes. After hybridization at 37°C for 24 h, the captured DNA was eluted by washing the nylon membrane and the DNA fragments were re-amplified. The PCR products were ligated into pMD18-T Easy-vector (Takara) and the ligation product was transformed into competent *Escherichia coli* DH5α cells. The clones were screened with probes and labeled by the ECL system (Amersham). Hybridization, stringency of washing, membrane blocking, antibody incubation, signal generation and detection were all performed according to the manufacturer's protocol. Considering nucleobase mutual partnerships and the order difference of recording the first base of repeat units and similar repeat types could be sorted into a single type. In this study, the libraries were screened with 14 probes, they were $(AC)_{15}$, $(AG)_{15}$, $(AAG)_8$, $(ACT)_8$, $(AAC)_8$, $(ATC)_8$, $(AGG)_8$, $(ACC)_8$, $(AAT)_8$, $(AGC)_8$, $(AAAT)_4$, $(AACC)_4$, $(ACGC)_4$ and $(AGAC)_4$ (Table 1).

After library screening, all clones were sequenced using BigDye chemistry with M13 forward and reverse primers on an Applied Biosystems ABI PRISM 3730 genetic analyzer. All developed clones in this study were compared with 1080 previously published sequences in GenBank using Vector NTI Suite 8.0, and all the duplicate sequences were deleted.

The SSR clones that had enough suitable flanking sequence were selected to design primers. The forward and reverse primers were designed with primer premier 5.0 software (Premier Biosoft International). The parameters used in the primer design were as follows: (1) primer length between 18 and 30 bp; (2) percentage of GC between 40 and 60%; (3) T_m values of the primers between 55 and 75°C; (4) PCR products from 100 to 450 bp.

The polymerase chain reaction mixture contained 90 ng genomic DNA, 1 U *Taq* polymerase (TaKaRa), 1× PCR buffer, 250 µM of each dNTP, 3.0 mM of Mg^{2+} and 0.2 µM each forward and reverse primer. Thermal cycling was performed in a PCR system (Eppendorf) and consisted of an initial denaturing at 95°C for 5 min, followed by 30 cycles each of 30 s denaturing at 95°C, 30 s annealing at locus-specific temperatures, 45 s extension at 72°C and concluding with a final extension at 72°C for 5 min (Zhan et al., 2007). PCR products were separated on 8% denaturing polyacrylamide gels (acrylamide: bisacrylamide, 29:1) using 1× Tris borate EDTA (TBE) buffer. The bands were visualized using silver staining. The lengths of the PCR products were analyzed using UVP BioImaging Systems and VisionWorkLS software.

The tandem repeats were initially identified using the tandem repeat finder (TRF) software (Version: 2.02) (Benson, 1999), with the following parameters: alignment parameters (match, mismatch, indels), (2, 7, 7); minimum alignment score to report repeat, 50;

Table 1. Types of 2, 3, and 4 bp probes and their corresponding targeting repeat units.

Types of probe		Corresponding targeting units
Di-nucleotide [a]	$(AG)_{15}$	AG, GA, TC, CT
	$(AC)_{15}$	AC, CA, TG, GT
Tri-nucleotide [b]	$(AAT)_8$	AAT, ATA, TAA, TTA, TAT, ATT
	$(AAC)_8$	AAC, ACA, CAA, TTG, TGT, GTT
	$(AAG)_8$	AAG, AGA, GAA, TTC, TCT, CTT
	$(ATC)_8$	ATC, TCA, CAT, TAG, AGT, GTA
	$(ACT)_8$	ACT, CTA, TAC, TGA, GAT, ATG
	$(AGC)_8$	AGC, GCA, CAG, TCG, CGT, GTC
	$(AGG)_8$	AGG, GGA, GAG, TCC, CCT, CTC
	$(ACC)_8$	ACC, CCA, CAC, TGG, GGT, GTG
Tetra-nucleotide [c]	$(AAAT)_4$	AAAT, AATA, ATAA, TAAA, et al.
	$(AACC)_4$	AACC, ACCA, CCAA, CAAC, et al.
	$(ACGC)_4$	ACGC, CGCA, GCAC, CACG, et al.
	$(AGAC)_4$	AGAC, GACA, ACAG, CAGA, et al.

[a], Probes of $(AC)_{15}$ and $(AG)_{15}$ occupied 50% of the four types of di-nucleotide repeats; [b], the type of tri-nucleotide probe used in this research occupied 80% of the ten types of tri-nucleotide repeats; [c], The type of tetra-nucleotide probe used in this research occupied 12.1% of the 33 types of tetra -nucleotide repeats.

Table 2. Fate of 952 clones obtained from the enrichment library.

The number of clones	952				
	↓				
	Negative clones	66 (6.93%)			
	Positive clones	886 (93.07%)			
		↓			
		Duplicate	133 (15.01%)		
		Non-redundant	753 (84.99%)		
			↓		
			repeat sequences	1004	
				↓	
				Microsatellite sequences	831 (82.77%)
				Minisatellite sequences	173 (17.23%)

maximum period size, 500. Subsequently, all independent tandem repeats in this study were analyzed using the excel macro software designed by our group, which could calculate the length, percentage and so on. Considering nucleobase mutual partnerships and the order difference of recording the first base of repeat units, similar repeat types could be sorted into a single type (Gao and Kong, 2005), such as AG (AG/ GA/ TC/ CT), AC (AC/ CA/ TG/ GT) (Table 1). According to the analysis principle, the number of types of 2 and 6 bp repeat sequences was 4, 10, 33, 102 and 350, respectively.

For statistical analysis, SAS was used for windows and coefficient of variation (*CV*) as a measure of SSR variation (see equation) (Luan et al, 2007).

$$CV = \frac{SD}{\overline{x}} \times 100 \quad \%$$

SD, Standard deviation; $\overline{x}$, mean.

RESULTS AND DISCUSSION

As shown in Table 2, 952 clones were obtained through microsatellite enrichment libraries construction-colony hybridization. The number of positive clones was 886, accounting for 93.07% of the total clones. All these tandem sequences were analyzed using Vector NTI Suite 8.0 with the following parameter: Minimum length of any overlap (50). As a result, 753 independent clones were obtained that were different from the repeats reported in GenBank. The tandem repeats were initially identified using the TRF software. Each of the independent clones

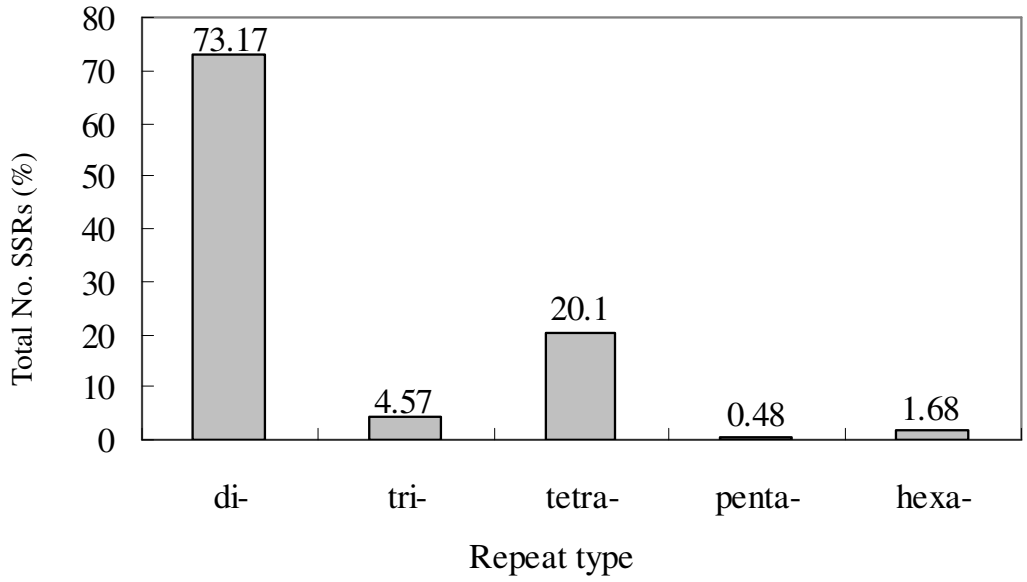

Figure 1. Relative abundance of the various sized repeat sequences in this study.

contained one or more repeat sequences. A total of 1004 sequences were obtained from 753 individual clones, including 831 microsatellites and 173 minisatellites, accounting for 82.77 and 17.23% of all tandem repeats, respectively.

Primer design from the SSR libraries

A total of 30 turbot individuals were employed for polymorphism assessment. In total, 360 SSR markers were developed in this study and 158 SSR markers were polymorphic, the percent of polymorphic microsatellite markers was 44%. These sequences can be retrieved with the following accession numbers: FJ965547-FJ965554; GQ121160-GQ121250; GQ132059-GQ132118 (Ruan et al., 2010). The number of alleles (*A*) observed heterozygosity (*Ho*), expected heterozygosity (*He*), *P*-value and conformance to Hardy-Weinberg (*HW*) expectations were estimated using the program POPGENE version 1.31. Null allele frequencies were obtained with the algorithm implemented in CERVUS 3.0 program (Marshall et al., 1998). Parts of the data were showed in Appendix 1.

SSR markers, in addition to their co-dominant character and abundant genetic variety, are selectively neutral and are widely distributed throughout genomes (Chistiakov et al., 2006); therefore, except for the difficulty in developing them, these molecular markers are suitable for genetic linkage mapping (Shimoda et al., 1999; Wang et al., 2007) for the turbot. These SSR markers can also serve as bridges for comparative mapping in other species and are more easily transferable across populations or even across species (Thomas and Scott, 1993).

Repeat sequence types, numbers and distributions

To construct microsatellite enriched libraries with higher efficiency, we characterized the published repeat sequences of turbot before this study, following which 14 different probes were designed and used in present study. The results obtained validated the effectiveness of our probe design protocol. The sequence data of the microsatellites detected in GenBank for the turbot are shown in Appendix 2. According to the rule of base mutual partnership, the two probes, $(AC)_{15}$ and $(AG)_{15}$, accounted for 50% of the four types of dinucleotide repeat. $(AAG)_8$, $(ACT)_8$, $(AAC)_8$, $(ATC)_8$, $(AGG)_8$, $(ACC)_8$, $(AAT)_8$ and $(AGC)_8$, accounted for 80% of the 10 types of trinucleotide repeat. $(AAAT)_4$, $(AACC)_4$, $(ACGC)_4$ and $(AGAC)_4$ accounted for 12% of the 33 tetra-nucleotide repeats (Table 1). In this study, 831 microsatellites of turbot were isolated from several enrichment libraries constructed using 14 probes. These present findings provided important information about the genomic organization of 2 to 4 repeat sequences in turbot. As shown in Figure 1, the proportions of the various types of SSRs (mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats) were not evenly distributed. The estimates of SSR-type frequency calculated here were similar to the published database (Figure 2), which suggested that, dinucleotide repeats were the most common followed by tetra-, tri-, hexa- and pentanucleotide repeats, respectively.

Characteristics and abundance of SSR types in the present research was showed in Table 1. Mono-nucleotide repeats were not identified in this study. And it is evident that dinucleotides were the most common type of repeat, accounting for 73.17% of the total number of

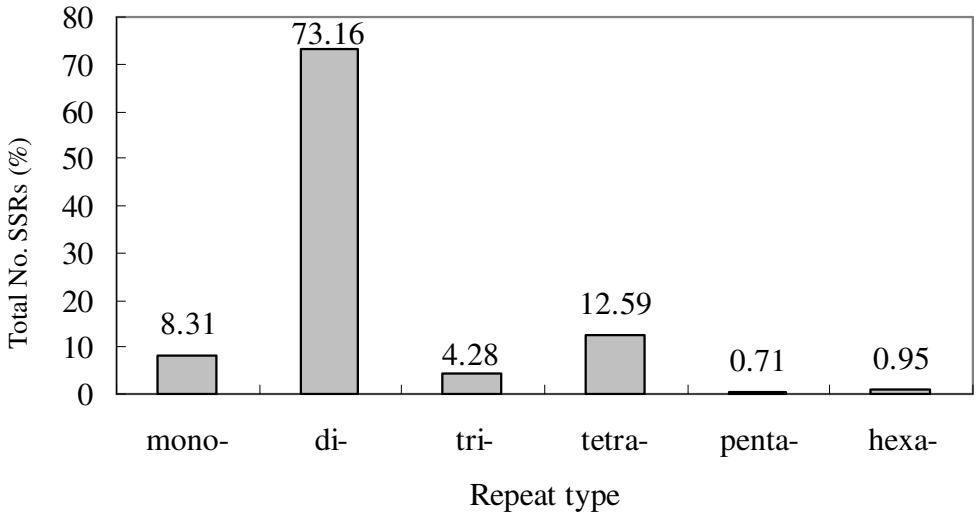

Figure 2. Relative abundance of the various sized repeat sequences registered in GenBank.

tandem repeats (831) and 53.50% of the total length of SSRs (75882 bp) in this work. Within these dinucleotide repeats, the unit AC (390) was the most abundant, followed by AG repeats (218). This result was similar with the analysis of the public turbot database in GenBank (Appendix 1) and the units of AT and GC were comparatively rare. This tendency could also be observed in some other species, such as the *Takifugu rubripes* genome (Cui et al., 2006). In contrast, other species (*Marsupenaeus japonicus* and *Fenneropenaeus chinensis*) contained more AT repeats, followed by AC, AG and GC (Gao and Kong, 2005). GenBank data did not include any GC dinucleotide microsatellite sequences for the turbot, so no GC probe was used in this research. According to previous reports, the GC repeat unit is rare in genomes of many organisms (Tóth et al., 2000). One explanation of this GC suppression might be that methylation of CpG represents a mutational hot spot through deamination of methylcytidine to thymidine (Schorderet and Gartler, 1992). In this research, eight types of trinucleotide repeat probes were used, accounting for 80% of the 10 types of trinucleotide repeats. The result showed that only 38 repeats were found (Table 1). SSR sequences constitute a large fraction of non-coding DNA and are relatively rare in protein coding regions. However, the disease-associated trinucleotide repeats are mostly found in coding regions in certain species. For example, abnormal expansions (≥ 39) of CAG repeats, which translated into polyglutamine stretches, cause Machado-Joseph disease (Rubinsztein et al., 1995). Metzgar et al. (2000) suggested that, such excesses of trinucleotide coding repeats were controlled primarily by mutation pressure. Most trinucleotides are not extensively conserved, even when they form parts of protein coding sequences, because some mutations may be deleterious for species and long stretches of trinucleotide repeats can be unstable during meiosis or gametogenesis. Each of the twenty nucleic acids were encoded by three bases, such as Ala (GCA), Ser (AGC) and Met (AUG), out of the four bases (A, U, C and G). The composition of codons is similar to the SSR repeat unit (Table 3), thus some trinucleotide repeats might disturb the function of codons. For instance, nucleic acid sequences around start codons contain fewer AUG trinucleotides due to negative selection against disruptive trinucleotides that could disturb the accurate detection of true start codons (Saito and Tomita, 1999). With further understanding of trinucleotide repeat mutation and evolution, the genetic pathogenesis of some diseases that have appeared recently in turbot farming, including albinism, could be clarified, which will no doubt speed the MAS progress in turbot farming.

Among the 167 tetranucleotide classes, all our enrichment libraries were separately screened with four oligonucleotide probes. The results showed that, aside from AACC (15), ACGC (51) and AGAC (78), other SSR unit types (ATCT (5), AATC (2) and AGTG (3)) were occasionally found. Little data about tetranucleotide repeat sequences are available for turbot, thus further studies will be required to analyze the frequency of each tetrameric repeat type. However, four classes of pentanucleotide repeats and six classes of hexanucleotide repeats were occasionally found in this study.

The correlation between the copy number of repeat units and the length of repeat sequence

In this research, the copy number of repeat units ranged from 5 to 245. The average copy number of dinucleotide repeats was 33.46 and the copy number ranged from 13 to 242. For trinucleotide repeats, these parameters were 36.35 and 8 to 245, respectively. For tetranucleotide repeats, they were 42.36 and 6 to 141. For pentanucleotide

Table 3. Characteristics and abundance of SSR types in this study.

Types of repeats	No. of SSRs	No. SSRs[a] (%)	Length	Length of SSRs[b] (%)	Total length of tandem repeats[c] (%)	No. of repeats times in range	Average No. of repeat times	*CV*/%
AC	390	46.93	26187	34.51	23.95	13–166	33.65	71.58
AG	218	26.23	14409	18.99	13.18	13–242	33.26	89.60
Subtotal	608	73.17	40596	53.50	37.13	13–242	33.46	
AGC	14	1.68	1007	1.33	0.92	8–52	23.87	57.17
ATC	4	0.48	222	0.29	0.20	9–32	19.32	55.70
AGG	8	0.96	1410	1.86	1.29	12–245	61.30	126.11
AAG	9	1.08	1293	1.7	1.18	9–142	48.22	87.86
ACC	1	0.12	45	0.06	0.04	15–15	15.30	0
AAC	1	0.12	57	0.08	0.05	19–19	19	0
CGG	1	0.12	33	0.04	0.03	11–11	11	0
Subtotal	38	4.57	4067	5.36	3.71	8–245	36.35	
ACAG	78	9.39	17494	23.05	16.00	6–141	56.28	67.1
AATC	2	0.24	161	0.21	0.15	14–27	20.15	46.67
ATCT	5	0.60	739	0.9	0.68	6–114	36.98	120.46
ACGC	51	6.14	3210	4.23	2.93	8–34	15.81	42.40
AACC	15	1.81	3415	4.50	3.12	8–124	57.33	65.44
AGTG	3	0.36	270	0.36	0.25	17–26	22.53	20.32
Subtotal	167	20.10	28187	37.07	25.78	6–141	42.36	
ACACA	1	0.12	25	0.03	0.02	5–5	5	0
GAGGA	1	0.12	207	0.27	0.19	42–42	42.40	0
CAACG	1	0.12	224	0.30	0.20	45–45	44.80	0
ACAGT	1	0.12	39	0.05	0.04	8–8	7.60	0
Subtotal	4	0.48	495	0.65	0.70	5–45	24.95	
CAGAGA	1	0.12	146	0.19	0.13	24–24	24.30	0
GAAGAA	1	0.12	299	0.39	0.27	50–50	49.80	0
AGGAAG	1	0.12	322	0.42	0.29	54–54	53.70	0
GAGGAG	1	0.12	278	0.37	0.25	46–46	46.30	0
GCTGTT	1	0.12	160	0.21	0.15	26–26	25.70	0
TGCTGC	2	0.24	313	0.41	0.29	25–26	26.05	1.36
TGCTGA	2	0.24	460	0.61	0.42	17–60	38.30	79.39
GTGTGC	2	0.24	123	0.16	0.11	10–11	10.10	8.40
GACGAG	1	0.12	214	0.28	0.20	36–36	35.80	0
TGTGTG	1	0.12	141	0.19	0.13	24–24	23.5	0
AGAGAG	1	0.12	81	0.11	0.07	14–14	13.5	0
Subtotal	14	1.68	2537	3.34	2.31	10–60	30.11	
Total	831		75882		69.63			

[a], Total Number of SSR was 831; [b], total length of the SSR sequences was 75880 bp; [c], total length of repeat sequences was 109336 bp; No., number.

repeats, they were 24.95 and 5 to 45. Moreover, for hexanucleotide repeats, they were 30.11 and 10 to 60. The distribution of copy number of SSR sequences is summarized in Figure 3. The frequency distributions of dinucleotides were parabola-like in shape. In addition, the di- and tetranucleotides repeats were centralized in a middle-leaning-left position. The frequency distribution of tri-nucleotides was centralized in the domain composed of low copy number. Due to the relatively lower copy number of classes and frequencies, the figures for penta- and hexanucleotide repeats were not clear. Among dinucleotide repeat sequences, repeats with 15 copies were the most common. The average copy number of dinucleotide repeats was 33.46. Among trinucleotide repeats, repeats with nine copies were the most common and the average copy number was 36.35. The average copy number was 24.95 for tetranucleotide repeats and 30.11 for hexanucleotide repeats.

The length of each repeat sequence was determined by the copy number of its repeat unit. The cumulative

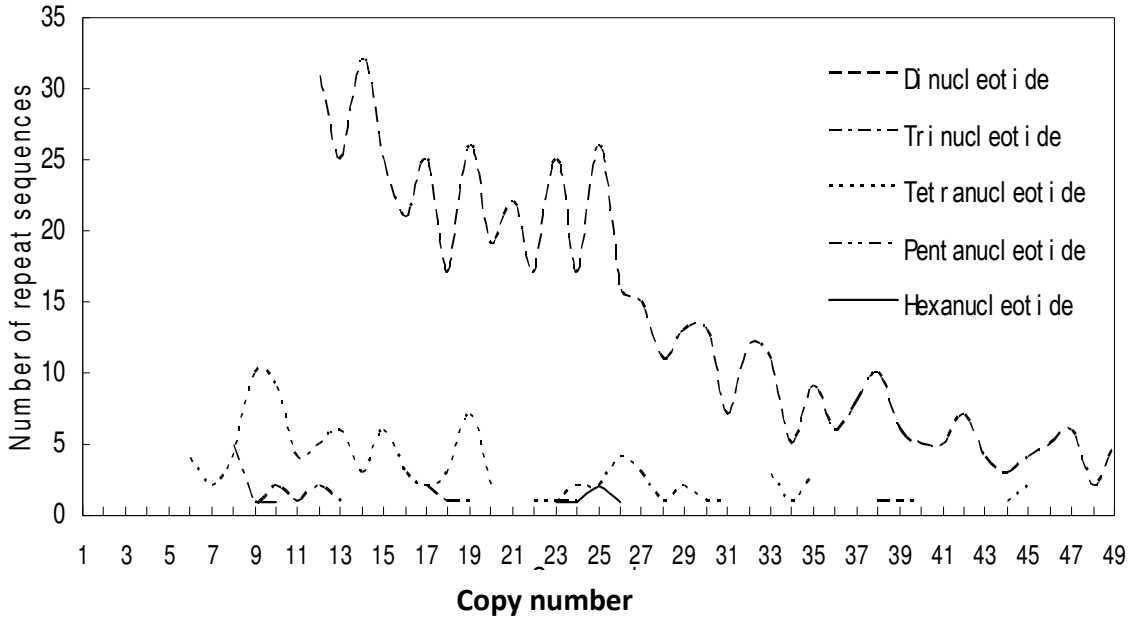

Figure 3. Distribution of the number of SSR sequences.

length of dinucleotide repeats was the greatest, at 40596 bp, which accounted for 53.50% of the total length of SSR repeats. In decreasing order of cumulative length, the others were tetranucleotide, trinucleotide, hexanucleotide and pentanucleotide repeats in turn. The cumulative length of pentanucleotide repeats was the shortest, only occupying 0.70% of the total length of repeat sequences (Table 1). Thus, we concluded that, the length of repeat sequences was determined by the copy number and the dinucleotide repeats were absolutely predominant, not only in the number of repeat sequences, but also in terms of cumulative length. In this study, a negative correlation between the length of repeat units and the average copy number was detected ($r = –0.5$, $P = 0.072$). The result showed that, the frequency of microsatellites decreases with increasing repeat unit length. That is, the smaller the unit length, the higher repeat number, possibly due to the greater selective pressure against longer unit lengths (Samadi et al., 1998).

In this study, the coefficient of variance (*CV*) was introduced as a measure of SSR variation ability, where the higher the repeated variation coefficient, the higher the repeated variation ability of microsatellite. Within dinucleotide repeats, the variance of AG was the highest, with a *CV* value of 89.60. For trinucleotide repeats, AGG had the highest *CV* at 126.11. For tetranucleotide repeats, ATCT had the highest *CV* at 120.46 and for hexanucleotide repeats; they were TGCTGA with a *CV* of 79.39. The average *CVs* of dinucleotides, tetranucleotide, trinucleotide, hexanucleotide and pentanucleotide repeats were, in decreasing proportions, 80.59, 57.43, 46.69, 8.10 and 0, respectively (Table 3). Hence, dinucleotides repeats were the most variable among the different types of repeat.

It is evident that dinucleotides were the most common type of repeat, accounting for 73.17% of the total number of tandem repeats (831) and 53.50% of the total length of SSRs (75882 bp) in this work. The average *CV* of dinucleotides was also the highest (Table 3). The number, length and average *CV* of dinucleotides were also the most prevalent in some other species, such as the genomes of Chinese shrimp and Drosophila (Katti et al., 2001). Only dinucleotides were found in the genome of the prokaryote methanogenic archaeon (Samadi et al., 1998). It could be concluded that the possibility of dinucleotide repeats may be the basis of origin and evolution of repeat sequences. In this study, the coefficients of variance (*CVs*) of trinucleotide repeats ranged from 0 to 126.11. The variance of AGG was the highest, with a *CV* value of 126.11. Therefore, some types of trinucleotides may be more variable than others.

Conclusion

A microsatellite enrichment libraries construction-colony hybridization strategy was used in isolating the turbot (*S. maximus* L.). Using 14 different probes, 831 microsatellites were detected and 158 polymorphic microsatellite markers were developed. And so far, about 316 SSR markers have been reported by other groups (Coughlan et al., 1996; Iyengar et al., 2000; Bouza et al. 2002, Pardo et al. 2005, 2006, 2007). All these highly polymorphic microsatellite markers will be useful tools for constructing genetic linkage maps of turbot and for identifying quantitative trait loci (QTLs) in this species. To

date, the full genome of turbot has not been determined, due to limits on research outlays, therefore, analyzing the different nucleotide repeat units across the whole genome of turbot and provide information regarding the repeat sequence characteristics of the turbot genome.

ACKNOWLEDGEMENTS

This work was supported by the National High Technology Research and Development Program of China (Grant No. 2006AA10A406) and Research on Genotypic Milieu and Mechanism of the important Characters in Turbot (Grant N0. 2010-cb-04).

REFERENCES

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

Blanquer A, Alayse JP, Berrada-Pkhami O, Berrebi P (1992) Allozyme variation in turbot (*Psetta maxima*) and brill (*Scophthalmus rhombus*) (*Osteichthyes, Pleuronectiformes, Scophthalmidae*) throughout their range in Europe. J Fish Biol. 41 (5): 725-736.

Blouin MS (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends Ecol. Evol. 18: 503-511.

Bouza C, Presa P, Castro J, Sánchez L, Martínez P (2002). Allozyme and microsatellite diversity in natural and domestic populations of turbot (*Scophthalmus maximus*) in comparison with other Pleuronectiformes. Can. J. Fish. Aquat. Sci. 59: 1460-1473.

Chen SL, Ma HY, Jiang Y, Liao XL, Meng L (2007). Isolation and characterization of polymorphic microsatellite sequences from an EST library of turbot (*Scophthalmus maximus*) and cross-species amplification. Mol. Ecol. Notes. 7: 848–850.

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.

Coughlan J, Mccarthy R, McGregor D (1996). Four polymorphic microsatellites in turbot (*Scophthalmus maximus*). Anim. Genet. 27(6): 441-447.

Cui JZ, Shen XY, Yang GP, Gong QL, Gu QQ (2006). The analysis of simple sequence repeats in Takifugu rubripes genome. Periodical of Ocean University of China 36(2): 249-254.

Danzmann RG, Gharbi K (2001). Gene mapping in fishes: a means to an end. Genetica, 111: 3-23.

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

Gao H, Kong J (2005). Analysis of tandem repeats in the genome of Chinese shrimp Fenneropenaeus chinensis. Chin. Sci. Bull. 50(14): 1462-1469.

Hu JJ, Zhan AB, Lu W, Hu XL, Bao ZM (2007). An Efficient Procedure for Isolating Microsatellite DNAs from Sea Cucumber (*Apostichopus japonicus*). J. Ocean Univ. China (Oceanic and Coastal Sea Research) 6(1): 12-15.

International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature, 409: 860-921.

Iyengar A, Piyapattanakorn S, Heipel DA, Stone DM, Howell BR (2000). A suite of highly polymorphic microsatellite markers in turbot (*Scophthalmus maximus L.*) with potential for use across several flatfish species. Mol. Ecol. 9: 365-378.

Karagyozov L, Kalcheva ID, Chapman VM (1993). Construction of random small-insert genomic libraries highly enriched for simple sequences repeats. Nucleic Acids Res. 21: 3911-3912.

Katti MV, Ranjekar PK, Gupta VS (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. 18(7): 1161-1167.

Kenchington EL, Patwary MU, Zouros E, Bird CJ (2006). Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placopecten magellanicus*). Mol Ecol. 15: 1781-1796.

Lei JL, Liu XF (1995). An Primary Study on Culture of Turbot, Scophthalmus Maeoticus L. Modern Fisheries information (in Chinese) 10(11): 1-3.

Lei JL, Zhan XL (2001). Experiment of Rearing Scophthalmus maximus Linnaeus by Using Deep Well Sea Water in the Farming Plant. Modern Fisheries information (in Chinese) 16(3): 10-12.

Lei JL (2002). Some Problems and Suggestion on Industrialized Fish Farming along the North Coast of Our Country. Modern Fisheries information (in Chinese) 17(4): 5-8.

Lei JL, Ma AJ, Liu XF, Meng Q (2003). Study on the Development of Embryo, Larval and Juvenile of Turbot (*Scophthalmus maximus* L.). Oceanologia et limnologia sinica (in Chinese) 34: 9-18.

Lei JL, Ma AJ, Chen C, Zhuang ZM (2005). The Present Status and Sustainable Development of Turbot (*Scophthalmus maximus* L.) Culture in China. Eng. Sci. (in Chinese) 7(5): 12-16.

Liu YG, Liu LX, Lei ZW, Gao AY, Li BF (2006). Identification of polymorphic microsatellite markers from RAPD product in turbot (*Scophthalmus maximus*) and a test of cross-species amplification. Mol. Ecol. Notes, 6: 867-869.

Li HJ, Hu JJ, Bao ZM, Zhang GF (2007). A set of polymorphic microsatellite loci for the bay scallop, *Argopecten irradians*. Mol. Ecol. Notes, 7: 422-424.

Luan S, Kong J, Wang QY, Gao H, Wang WJ, Zhang QW (2007). Analysis of microsatellite repeats in the genome of Marsupenaeus japonicas. Progress Nat. Sci. 17(6): 731-740.

Marshall TC, Slate J, Kruuk LEB and Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639-655.

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

Norris AT, Bradley DG, Cunningham EP (2000). Parentage and relatedness determination in farmed Atlantic salmon (*Salmosalar*) using microsatellite markers. Aquaculture, 182: 73-83.

Pardo BG, Casas L, Fortes GG, Bouza C, Martínez P, Clark MS, Sánchez L (2005). New microsatellite markers in turbot (*Scophthalmus maximus*) derived from an enriched genomic library and sequence databases. Mol. Ecol. Notes, 5(1): 62-64.

Pardo BG, Hermida M, Fernández C, Bouza C, Pérez M, Llavona A, Sánchez L, Martínez P (2006). A set of highly polymorphic microsatellites useful for kinship and population analysis in turbot (*Scophthalmus maximus* L.). Aquat. Res. 37(15): 1578-1582.

Pardo BG, Fernández C, Hermida M, Vázquez-López A, Pérez M, Presa P, Calaza M, Alvarez-Dios JA, Comesaña AS, Raposo-Guillán J, Bouza C, Martínez P (2007). Development and characterization of 248 novel microsatellite markers in turbot (*Scophthalmus maximus*). Genome 50(3): 329-322.

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 7: 215-222.

Ren GJ, Hu JJ, Bao ZM, Jiang X, Gao TX (2009). Isolation and characterization of eleven polymorphic microsatellite markers of sand lance (Ammodytes personatus). Consrey. Genet. 10: 1837-1839.

Ruan XH, Wang WJ, Kong J, Yu F, Huang XQ (2010). Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellite markers and its application in QTL analysis. Aquaculture doi: 10.1016/j.aquaculture.2010.08.010.

Rubinsztein DC, Leggo J, Coetzee GA (1995), Sequence variation and size ranges of CAG repeats in the Machado-Joseph disease, spinocerebellar ataxia type 1 and androgen receptor genes. Hum. Mol. Genet. 4: 1585-1590.

Samadi S, Artiguebielle E, Estoup A, Pointier JP, Silvain JF, Heller J, Cariou ML, Jame P (1998). Density and variability of dinucleotide microsatellites in the parthenogenetic polyploid snail *Melanoides tuberculata*. Mol. Ecol. 7(9): 1233-1236.

Saito R, Tomita M (1999). On negative selection against ATG triplets near start codons in eukaryotic and prokaryotic genomes. J. Mol. Evol. 48: 213-217.

Schorderet DF, Gartler SM (1992). Analysis of CpG suppression in methylated and nonmethylated species. Proc. Natl. Acad. Sci. USA, 89: 957-961.
Selvamani MJ, Degnan SM, Degnan BM (2001). Microsatellite genotyping of individual abalone larvae: Parentage assignment in aquaculture. Mar. Biot. 3(5): 478-485.
Shen XY, Gong QL, Lei JL, Kong J, Zhai JM, Li B (2004). Population genetic structure analysis of the imported turbot seedlings *Scophthalmus maximus* L. using RAPD and Microsatellite technique. Oceanologia et limnologia sinica, 35(4): 332-341.
Shimoda N, Knapik EW, Zinitj J, Sim C, Yamada E, Kaplan S, Jackson D, Sauvage FD, Jacob H, Fishman MC (1999). Zebrafish genetic map with 2000 microsatellite markers. Genomics, 58: 219-232.
Subramanian S, Mishra RK, Singh L (2003). Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol. 4 R13.
Thomas MR, Scott NS (1993). Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). Theor. Appl. Genet. 86: 985-990.
Tóth G, Gáspári Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. Genome Res. 10(7): 967-981.
Vos P, Hogers R, Bleeker M, Reijans M, Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.
Wang CM, Zhu ZY, Lo LC, Feng F, Lin G, Yang WT, Li J, Yue GH (2007). A microsatellite linkage map of Barramundi, *Lates calcarifer*. Genetics, 175: 907-915.
Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18: 6531-6535.
Zhan AB, Bao ZM, Lu W, Hu XL, Peng W, Wang ML, Hu JJ (2007). Development and characterization of 45 novel microsatellite markers for sea cucumber (*Apostichopus japonicus*). Mol. Ecol. Notes. 7: 1345-1348.
Zou SM, Li SF, Cai WQ (2001). Molecular genetic markers and variations of cultured *Paralichthys olivaceus* and *Scophthalmus maximus*. J. Fish. Sci. China, 7(4): 25-28.

Appendix 1. Characteristics of the 21 microsatellite loci of turbot (*S. maximus* L.).

Locus	no.	Primer sequence (5'–3')	Repeat motif	Size (bp)	*A*	*Ho*	*He*	*P-value*	Null frequency
YSKr01	GQ121160	F: CGAGTCAACAGCCATCAAGC R: AGGTGGTGGACCGTTCAAGT	$(GT)_7GC(GT)_{12}$	187-202	3	0.1250	0.1215	0.9952	-0.0348
YSKr18	GQ121165	F: GCATCACACCACTCACAACAC R: TCAATCATCCCACAGGCACAC	$(AC)_{20}$	153-170	5	1.0000	0.6950	0.0024**	-0.2176
YSKr24	GQ121168	F: GGTGAGCGACAATCCATCTG R: GGTCACGCACAAGTGTCCTG	$(AG)_{14}$	118-166	9	0.9000	0.7734	0.1059	-0.0878
YSKr40	GQ121175	F: GCCACTTACACGAGACCTG R: AAGAGCAAACCAAGCACG	$(GCA)_4GCG(GCA)_{10}$	76-103	5	0.5000	0.5922	0.0020**	+0.1740
YSKr41	GQ121176	F: CCTCTCATTGGAAGACGG R: TGCCTGTCAGTGGAAGTAG	$(CT)_{19}$	140-208	6	0.8333	0.7261	0.0067**	-0.0722
YSKr43	GQ121177	F: TATGCCCTGGTTACGCTTC R: TCTGGTGATGTCTTCTGGAG	$(CA)_{11}$	125-147	4	1.0000	0.7216	0.0000***	-0.1798
YSKr51	GQ121182	F: GAACGACCGACGGAAACTG R: TTCTAAACACCACCGACCCT	$(GA)_7T(AT)_{12}$	241-284	5	0.9167	0.7961	0.5057	-0.0793
YSKr81	GQ121200	F: TGCTTCTCTGTAAAAGTCTG R: ATGTCCCATCCTGCCAC	$(TTGG)_4…(TTGG)_{13}…(TTGG)_6$	184-241	9	0.9333	0.8678	0.0313*	-0.0459
YSKr85	GQ121201	F: TACTTACACTGTGTATGTGC R: GAGAACCGAAGAAATGAGA	$(GTGC)_9$	252-304	9	0.7667	0.8198	0.2053	+0.0424
YSKr09	FJ965547	F:GTCTGGTGTCGTCTTGCTT R:TGGAAACAGGTTGTAACTC	$(GT)_5A(TG)_{28}…(GA)_{18}$	203-256	5	0.6667	0.7463	0.2660	+0.0458
YSKr11	FJ965548	F: ATGTTTGCCACTATGCGT R: AACTGTTTATTGTGCTCGTC	$(GT)_{12}$	211-226	3	0.2759	0.3503	0.0003***	+0.0962
YSKr12	FJ965549	F: GGACAACAGACCGAAACC R: ATGTACCTGCCTGGAATG	$(GT)_{13}TT(GA)_6…(GA)_{11}$	128-187	6	0.7667	0.7910	0.2120	+0.0070
YSKr17	FJ965550	F: GTGGGAATGAATCGGACAGG R: AACGCCTCCCCCTCATCTCT	$(ACAGAGAG)_4(AG)_8$	196-234	5	0.8276	0.7447	0.0064**	-0.0355
YSKr21	FJ965551	F: TGCCTCAGTTCAGAGTCAGC R: GGGTGTCAGAGTAGGGTTGT	$(CA)_{17}$	230-245	4	0.8667	0.6627	0.3320	-0.1487
YSKr26	FJ965552	F: GCTCTGTAAGTCGGTGAAGGT R: TAAACTGCCGCTGGATGAAG	$(TC)_{14}$	228-242	5	0.7667	0.6723	0.9214	-0.0823
YSKr27	FJ965553	F: AAGTTGGAGGGTGAAGAGTT R: TATGTAAAGTGGCTTGCTGT	$(TG)_{28}$	277-340	9	0.7000	0.8232	0.0145*	+0.0863
YSKr37	GU997109	F: GCAACATCACATAGGAGC R: AGGTGGCGGAGACAAAG	$(AC)_{23}$	74-101	6	1.0000	0.7828	0.0004***	-0.1377

Appendix 1. Contd.

YSKr57	GU997110	F: TGTTTGAAGTCGCACGCCT R: ACACACTGGAGCCTTTGAGC	$(AG)_{19}$	174-180	2	0.4348	0.4329	0.9824	-0.0511
YSKr285	GU997111	F: AGGAGGACGAGGTCAAACG R: GAGACACAAACCAACAACAGACA	$(AC)_{20}$	122-137	4	0.2917	0.2642	0.9966	-0.0672
YSKr286	GU997112	F: AGCATCAGCAACGGAGGAC R: GGCAGCACATTGTTCTGT	$(TC)_{9}$	201-284	7	0.8333	0.6667	0.9532	-0.1736
YSKr287	GU997113	F: TATCCTCATCCCTTTCATCT R: CAAACTGGAGCTTAGTAACA	$(CTGT)_{6}$	276-300	4	0.4583	0.5895	0.1146	+0.2284

no, GenBank Accession number; *T*a, optimal annealing temperature; *A*, number of alleles; *H*o, observed heterozygosity; *H*e, expected heterozygosity; *P*-value, exact *P* value for Hardy–Weinberg equilibrium test; * significant at the 5% level; ** significant at the 1% level; and *** significant at the 0.1% level.

Appendix 2. Characteristics and abundance of SSR types registered in GenBank.

Types of repeats	No. of SSRs	No. SSRs[a] (%)	Length	Length of SSRs[b] (%)	Total length of tandem repeats[c] (%)	No. of repeats times in range	Average no. of repeat times	*CV*/%
A	35	8.31	1255	5.63	4.99	25–77	35.86	34.32
Subtotal	35	8.31	1255	5.63	4.99	25–77		
AC	276	65.56	14383	64.52	57.23	13–101	26.09	55.16
AG	26	6.18	1157	5.19	4.60	13–45	22.29	33.56
AT	6	1.42	232	1.04	0.92	14–24	19.33	19.40
Subtotal	308	73.16	15772	70.75	62.75	13–101	25.64	
ATT	2	0.48	84	0.38	0.33	11–17	14	27.27
ATC	7	1.66	391	1.75	1.56	9–61	18.66	101.52
AGC	4	0.95	144	0.65	0.57	8–16	11.9	28.10
AAG	2	0.48	67	0.30	0.27	11–11	11.15	1.90
AAC	1	0.24	26	0.12	0.10	9–9	8.70	0
AGG	2	0.48	54	0.24	0.21	8–10	9	10.10
Subtotal	18	4.28	766	3.44	3.04	8–61	14.18	
ACGC	7	1.66	250	1.12	0.99	8–13	9.04	20.03
ACCT	4	0.95	558	2.50	2.22	12–64	35.08	67.69
ACGG	1	0.24	27	0.12	0.11	7–7	6.80	0
ATCT	24	5.70	2161	9.69	8.60	8–68	22.59	72.29
AGAC	3	0.71	231	1.04	0.92	12–23	19.43	31.79
ATGT	10	2.38	650	2.92	2.59	7–50	16.31	76.87
CTTT	1	0.24	55	0.25	0.22	14–14	14	0
ACTC	2	0.48	224	1.00	0.89	18–40	28.80	54.02

Appendix 2. Contd.

GTTT	1	0.24	30	0.13	0.12	8–8	7.5	0
Subtotal	53	12.59	4186	18.77	16.66	7–7	19.87	
ACACA	1	0.24	57	0.26	0.23	12–12	12.2	0
CCGGA	1	0.24	39	0.17	0.16	8–8	7.80	0
TACAC	1	0.24	55	0.25	0.22	11–11	11	0
Subtotal	3	0.71	151	0.68	0.61	8–12	10.33	
CAGATC	2	0.48	104	0.47	0.41	9–9	8.670	0
TGACCG	1	0.24	30	0.13	0.12	5–5	5	0
AATTCG	1	0.24	29	0.13	0.12	5–5	5	0
Subtotal	4	0.95	163	0.73	0.65	5–9	6.84	
Total	421		22326		88.73			

[a], Total Number of SSRs was 421; [b], total length of the SSR sequences was 22291 bp; [c], the total length of repeat sequences was 25,150 bp.