

Available online at www.sciencedirect.com**Aquaculture**

Aquaculture 271 (2007) 558–574

www.elsevier.com/locate/aqua-online

Development of novel EST-SSR markers in common carp by data mining from public EST sequences

Dan Wang^{a,b}, Xiaolin Liao^{a,b}, Lei Cheng^{a,b}, Xiaomu Yu^a, Jingou Tong^{a,*}^a State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, PR China^b Graduate School of Chinese Academy of Sciences, Beijing 100039, PR China

Received 9 November 2006; received in revised form 30 May 2007; accepted 1 June 2007

Abstract

Expressed sequence tags (ESTs) are a source for microsatellite development. In the present study, EST-derived microsatellites (EST-SSRs) were generated and characterized in the common carp (*Cyprinus carpio*) by data mining from updated public EST databases and by subsequent testing for polymorphism. About 5.5% (555) of 10,088 ESTs contain repeat motifs of various types and lengths with CA being the most abundant dinucleotide one. Out of the 60 EST-SSRs for which PCR primers were designed, 25 loci showed polymorphism in a common carp population with the alleles per locus ranging from 3 to 17 (mean 7). The observed (H_O) and expected (H_E) heterozygosities of these EST-SSRs were 0.13–1.00 and 0.12–0.91, respectively. Six EST-SSR loci significantly deviated from the Hardy–Weinberg equilibrium (HWE) expectation, and the remaining 19 loci were in HWE. Of the 60 primer sets, the rates of polymorphic EST-SSRs were 42% in common carp, 17% in crucian carp (*Carassius auratus*), and 5% in silver carp (*Hypophthalmichthys molitrix*), respectively. These new EST-SSR markers would provide sufficient polymorphism for population genetic studies and genome mapping of the common carp and its closely related fishes.

© 2007 Published by Elsevier B.V.

Keywords: Expressed sequence tag (EST); Data mining; EST-SSRs; Polymorphism; Cross-species amplification; *Cyprinus carpio*

1. Introduction

The common carp (*Cyprinus carpio* L.) is one of the most widely distributed fish species and has been extensively cultured in the world (Wohlfarth, 1984).

Simple sequence repeats (SSRs) are powerful for a variety of applications in aquaculture because they are hypervariable, nuclear-encoded genetic markers (Liu and Cordes, 2004). Traditionally, the isolation of SSRs has relied on the screening of genomic libraries using repetitive probes and sequencing of positive clones in order to develop locus-specific primers. These processes have been practiced for many organisms but are normally time-consuming and labor-intensive.

As a byproduct of EST projects in many organisms in which a vast amount of sequence data were

* Corresponding author. Tel.: +86 27 68780751; fax: +86 27 68780123.

E-mail address: jgtong@ihb.ac.cn (J. Tong).

generated, microsatellite-mining from SSR-containing ESTs is inexpensive and time-saving, and has proved to be an effective approach to develop microsatellites for genetic mapping and population genetics in animals (Serapion et al., 2004; Yue et al., 2004; Wang et al., 2005; Chen et al., 2005; Pérez et al., 2005; Maneeruttanarungroj et al., 2006) and plants (Cordeiro et al., 2001; Kantety et al., 2002; Chen et al., 2006). In aquaculture animals, Serapion et al. (2004) reported a pioneer study on the development of EST-SSRs by bioinformatic mining from the channel catfish (*Ictalurus punctatus*) databases. In the common carp, some EST-SSRs generated by data mining from partial EST databases have been previously reported (Yue et al., 2004). Since then publicly accessible EST sequences have increasingly accumulated in the common carp.

Here we report a new batch of EST-SSRs in the common carp by data mining from updated public EST resources and by laboratory testing for polymorphism of selected EST-SSRs. The cross-species amplifications of these markers are also explored in two other cyprinid species, the crucian carp (*Carassius auratus*) and the silver carp (*Hypophthalmichthys molitrix*).

2. Materials and methods

Thirty-two individuals of common carp and eight individuals of silver carp were collected from Dongting Lake in Hunan Province, the second largest lake of China. Eight individuals of crucian carp were sampled from Niushan Lake in Hubei Province, China. Genomic DNA was extracted from alcohol-preserved fin tissues of these specimens by using a salt-extraction method with slight modifications (Aljanabi and Martinez, 1997).

EST sequences of common carp were downloaded from GenBank, DDBJ and EMBL databases between January 1, 2002 and October 18, 2005. Online EST sequences of the common carp prior to 2002 were analyzed by Yue et al. (2004), and those ESTs were not included in the present study. All matched sequences were displayed in FASTA format and saved as a text file. EST sequences were analyzed by cluster analysis using the ContigExpress module in VectorNTI package (available at <http://download.invitrogen.com>) and linear assembly algorithm was applied. The criteria for clustering were set at a minimum overlap of 30 bases (default is 20 bases). Each cluster was visually inspected to ensure the fidelity of alignment to avoid pseudo-clusters caused by repetitive elements or long

strings of microsatellite repeats. ESTs belonging to contigs and singletons were recorded. The unique ESTs were then subjected to BLASTx search against the GenBank (protein database) for putative identification of gene function.

All the ESTs were screened for potential microsatellites by using the software Tandem Repeat Finder (Benson, 1999) with the following parameters: match: 2; mismatch 7; indel: 7; PM: mini-score; 30; and max period size 500. Strings of oligo sequences were used to search for microsatellites: 6 repeats for dinucleotides; 4 repeats for trinucleotides, and 3 repeats for tetranucleotides and pentanucleotides as described by Stallings et al. (1991).

A web-based software ‘Primer 3’ (<http://www.genome.wi.mit.edu/cgi-bin/primer/>) was used to design primers for the amplification of repeat regions of interest across the flanking regions of unique ESTs or genes. During the primer design, the range of annealing temperature was set up to be between 45 and 55 °C, and that of expected size of PCR products 150–250 bp. A single pair of “best” primers was designed and synthesized for each unique EST or gene that contains SSR, and no repeated designs and syntheses were carried out.

PCR amplifications of microsatellites were carried out on a thermocycler (PTC-100, MJ Research) by using the following program: 94 °C for 5 min, followed by 34 cycles of 94 °C for 35 s, appropriate annealing temperature (see Table 1) for 35 s, and 72 °C for 50 s, and a final extension of 72 °C for 10 min. The PCR reactions were performed in a 25 µl-reaction mixture, which contained 2.5 µl 10× reaction buffer, 2 µl Mg²⁺ (1.5 mmol L⁻¹), 1 µl dNTP (10 mmol L⁻¹), 0.5 U *Taq* polymerase (2 U/µl), 2 µl template DNA and 17 µl sterile water. PCR products were separated in 6% denaturing polyacrylamide gel and visualized by silver staining. Allele sizes were determined by comparison with pBR322 DNA/*Msp* I markers (Sino-American, Luoyang, China) combined with image analysis as described previously (Tong et al., 2005).

The number of alleles (N_a) and the range of alleles were calculated for each locus using the software PopGen (Yeh et al., 1999). Expected (H_E) and observed (H_O) heterozygosity, and the fitness of genotypic frequency to the Hardy–Weinberg equilibrium (HWE), were analyzed using ARLEQUIN (Schneider et al., 2000). Unbiased estimates for the exact P -values for tests of conformation to HWE were calculated using the Markov chain randomization method (Guo and Thompson, 1992). All these tests were adjusted for

Table 1
Characterization of 25 polymorphic EST-SSR loci in common carp

Locus	$T_a(^{\circ}\text{C})$	N_a	H_O	H_E	P
CCE13	55	9	0.97	0.80	0.1326
CCE14	50	3	0.13	0.12	1.0000
CCE15	55	4	1.00	0.75	0.0000*
CCE18	48	4	0.59	0.47	0.4607
CCE21	50	8	0.59	0.55	0.0103
CCE23	50	17	0.97	0.91	0.3076
CCE24	48	14	0.79	0.81	0.1117
CCE25	48	6	0.78	0.57	0.1491
CCE26	46	5	0.38	0.53	0.0354
CCE27	48	10	0.94	0.82	0.8992
CCE28	55	4	0.47	0.59	0.0106
CCE29	50	6	0.84	0.74	0.0000*
CCE31	50	5	0.81	0.64	0.0394
CCE35	50	6	0.59	0.52	0.9779
CCE37	48	8	0.56	0.77	0.0020
CCE41	52	9	0.94	0.79	0.2255
CCE43	55	3	0.72	0.51	0.0386
CCE45	50	5	0.66	0.70	0.0000*
CCE46	55	7	0.78	0.84	0.1371
CCE48	46	14	1.00	0.90	0.0000*
CCE49	55	4	0.47	0.60	0.1555
CCE51	50	3	0.72	0.48	0.0038
CCE53	50	11	0.72	0.89	0.0000*
CCE57	55	3	0.97	0.52	0.0000*
CCE60	48	6	0.53	0.74	0.0023

T_a : annealing temperature; N_a : number of alleles; H_O : observed heterozygosity; H_E : expected heterozygosity; P : probability in Hardy–Weinberg equilibrium. * denoted significant departure from HWE after Bonferroni correction ($P < 0.002$).

multiple simultaneous comparisons using a sequential Bonferroni correction.

3. Results

3.1. Identification of EST-derived SSRs in common carp

A total of 10,088 ESTs of common carp with an average length of 531 bp were downloaded from public databases and subject to bioinformatic analyses. A total of 555 (about 5.5%) of these ESTs contained SSRs inside. After clustering and assembly, 465 unique ESTs were identified (Appendix B), including 400 singletons and 65 contigs. The BLASTx results revealed that about 165 of these ESTs showed similarity to genes or proteins of known function (Appendix A and Appendix B).

Most of these common carp EST-SSRs were composed of dinucleotide and trinucleotide repeats.

Specifically, the abundance of di-, tri-, tetra-, and penta-nucleotide motifs among these ESTs is 37.2%, 30.8%, 20.4%, and 11.7%, respectively. For dinucleotides, AC/TG is the most abundant, with AG/TC the second, and CG/GC the least (Fig. 1a). The proportion of the trinucleotide repeats was also not evenly distributed, with the two most frequent types (AAT and ATC) accounting for 24.0% and 23.4% of the total motifs, respectively (Fig. 1b). In addition, AT-rich types were predominant, in contrast to low occurrence of CG-rich types among the trinucleotide EST-SSRs.

3.2. Primer screening

Among the 465 unique microsatellite-containing ESTs or genes (see Appendix B), 60 were randomly chosen for pilot tests for primer design, locus amplification and polymorphism. Appendix A summarizes the detailed information for the amplification

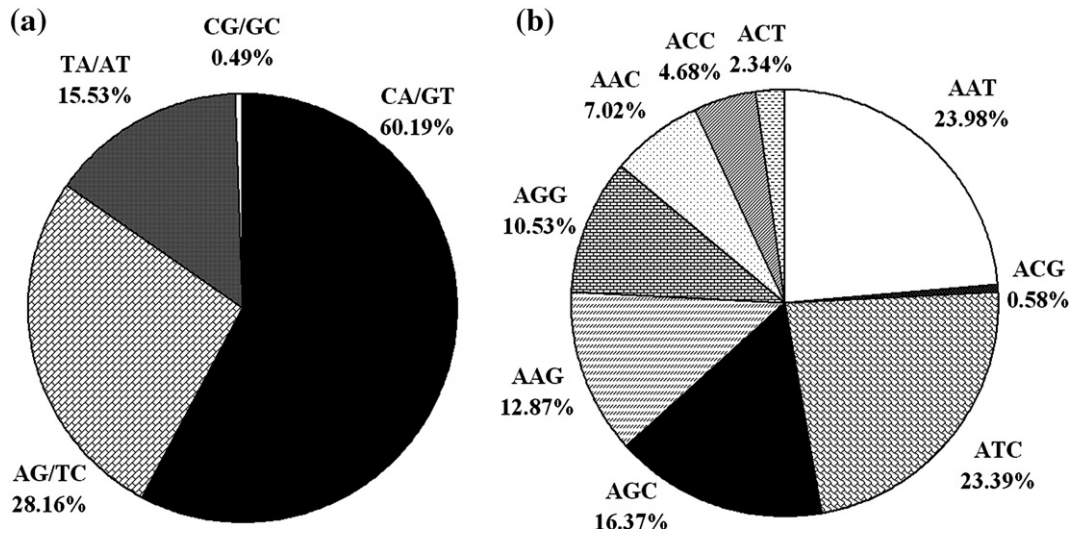


Fig. 1. Distribution of the repeat types of dinucleotides (a) and trinucleotides (b) in common carp EST-SSRs identified by mining public expressed sequence tags databases.

of EST-SSRs in three carp species. Out of the 60 common carp EST-SSRs for which primers were designed, 54 primers worked (25 polymorphic, 11 monomorphic, 18 with multiple bands) and 6 failed in the common carp; 37 primers worked (10 polymorphic, 14 monomorphic, 13 with multiple bands) and 23 failed in crucian carp; and 36 primers worked (3 polymorphic, 17 monomorphic, 16 with multiple bands) and 24 failed in silver carp.

3.3. Polymorphism of the EST-SSRs

Twenty-five of the 60 EST-SSRs were found to be polymorphic in a common carp population. The observed heterozygosity of these polymorphic loci ranged from 0.13 to 1.00, and expected heterozygosity ranged from 0.12 to 0.91. The number of alleles of the polymorphic EST-SSRs in common carp ranged from 3 to 17 (mean 7) (Table 1). When the frequencies and distributions of the alleles and genotypes were compared under the Hardy–Weinberg equilibrium expectation, 6 of the 25 loci (CCE15, CCE29, CCE45, CCE48, CCE53 and CCE57) showed significant departure from HWE after Bonferroni correction ($P < 0.002$), and the remaining 19 EST-SSRs were in HWE (Table 1).

3.4. Cross-species amplifications of EST-SSRs

Of the 60 common carp EST-SSRs, 10 (17%) of them showed polymorphism in a pilot panel of 8

individuals in crucian carp. In silver carp, only 3 (5%) of these loci were found to be polymorphic.

The average level of polymorphism of 8 common carp individuals randomly picked from the test population (Dongting Lake) was comparable with that of the same number of individuals from crucian carp and silver carp. The results showed that in common carp it had 2–11 alleles (mean 4.9), with the observed heterozygosity ranging from 0.25 to 1. The ranges of the number of alleles were 2–6 (mean 4.1) in crucian carp and 2–4 (mean 3.0) in silver carp, with the observed heterozygosity being 1 in crucian carp and ranging from 0.50 to 1 in the silver carp. In general, these loci are less polymorphic in crucian carp and silver carp than in their source species (common carp).

4. Discussion

Since Serapion et al (2004) first demonstrated the development of Type I marker (EST-SSRs) in an aquaculture species by using bioinformatic analysis, mining EST-SSRs from EST databases have been reported in several other aquaculture animals including fish (Yue et al., 2004; Rexroad et al., 2005; Chen et al., 2005), shrimp (Wang et al., 2005; Pérez et al., 2005; Maneeruttanarungroj et al., 2006) and scallop (Zhan et al., 2005). Online resources of EST sequences have been increasingly accumulated in fish. In this study, we analyzed updated ESTs of common carp from different public databases, and found that the percentage of SSR-

containing ESTs in all screened ESTs is 5.5% in common carp, which is lower than values reported in some aquaculture animals e.g. black tiger shrimp (*Penaeus monodon*) (13.7%, Maneeruttanarungroj et al., 2006), Japanese pufferfish (*Fugu rubripes*) (11.5%, Edwards et al., 1998) and channel catfish (11.2%, Serapion et al., 2004), but higher than those in some other species e.g. Chinese shrimp (*Fenneropenaeus chinensis*) (2.2%, Wang et al., 2005), bay scallop (*Argopecten irradians*) (3.9%, Zhan et al., 2005), and red sea bream (*Chrysophrys major*) (4%, Chen et al., 2005). The abundance of EST-derived microsatellites seems to be highly species-specific in aquacultured animals studied.

Dinucleotides are the dominant type of microsatellite repeats in most aquaculture species characterized so far, although trinucleotide repeats are most abundant in plants (Chen et al., 2006; Kantety et al., 2002). Of the dinucleotides, CA repeat is the most abundant in common carp (Fig. 1a), which is consistent with previous findings for both Type I and Type II microsatellites in fish (Edwards et al., 1998; David et al., 2001; Serapion et al., 2004), various plant species (Gupta and Varshney, 2000), and vertebrates as a whole (Neff and Gross, 2001). Based on analyses of 28 microsatellite-containing EST sequences, Yue et al. (2004) indicated that AT motif was the most abundant dinucleotide repeats in common carp. Because much more EST-SSRs were characterized in the present study, therefore, the resultant estimation for the types of microsatellite motifs in common carp should be more accurate.

Because of the evolutionary conservation, mutation rates within gene-coding sequences are lower than those in non-coding genomic sequences. In addition, polymorphism was not entirely dependent on repeat length, since repeats of five dinucleotides were polymorphic while some of six or seven ones were not (Nonneman and Waldbieser, 2005). Therefore, when compared with the Type II microsatellites in the same species, the level of polymorphism of EST-SSRs is normally slightly lower (Cho et al., 2000; Pérez et al., 2005). This is also the case in common carp, when previous data for Type II microsatellites (David et al., 2001; Liao et al., 2006) were compared with the results of the present study. The level of polymorphism of EST-SSRs in the present study may also be underestimated because of primer failure. Primers of the six loci (CCE4, CCE11, CCE12, CCE40, CCE50 and CCE55) failed to amplify in the common carp, and primers of some other loci could not amplify specific products (Appendix A). This could be due to one or

both primers being designed across the junction of the spliced ends of exons in the EST sequence, which in genomic DNA is interrupted by an intron (Cordeiro et al., 2001), or due to the inaccuracy of some EST sequences.

For those 60 EST-SSRs of the common carp, the rates of successful cross-species amplifications were 62% (37/60) in crucian carp and 60% (36/60) in silver carp, respectively. These data are higher than those for Type II microsatellites (David et al., 2001; Tong et al., 2002; Sun and Liang, 2004), and has confirmed that EST-SSRs of common carp have higher probability of success than that of Type II microsatellites in the cross-species amplifications among closely-related species (Yue et al., 2004). High amplification transferability of EST-SSRs has recently been reported across salmonid fishes (Rexroad et al., 2005), red sea breams (Chen et al., 2005) and shrimps (Pérez et al., 2005). The rate of success for EST-SSR primers in other species is also dependent on the source of the cDNA (Rexroad et al., 2005; Pérez et al., 2005). The common carp and the crucian carp belong to different genera of the same subfamily, while the common carp and the silver carp are distantly-related species of the different subfamilies. Using the same set of designed primers in the present study, the percentages of polymorphic EST-SSRs among common carp, crucian carp and silver carp were 42% (25/60), 17% (10/60), and 5% (3/60), respectively (Appendix A). These results again confirmed that the transferability of EST-SSR polymorphism is in accordance with the phylogenetic relatedness of the fish species studied (Yue et al., 2004; Rexroad et al., 2005).

In summary, the results of the present study confirm previous findings that development of Type I microsatellite markers by data mining is a relatively easy and efficient way for aquaculture species with adequate EST sequences. These novel EST-SSRs have shown sufficient level of polymorphism for such future studies as population genetics and genomics in the common carp and its closely related species.

Acknowledgments

The authors thank Dr. S. Ye and Ms C. Wang for their assistance in the collection of fish samples and in laboratory analysis, Prof. K.H. Chu and three anonymous reviewers for their critical comments on early versions of the manuscript. This work was partially supported by “973” (2004CB117405), “948” (2006G55-D), and NSFC (30370225, 30490234) projects.

Appendix A. EST-SSRs developed from common carp and cross-species amplifications in crucian carp and silver carp

Locus	Accession no.	Gene name	Repeat motif	Primer sequences(5'–3')	Common carp	Crucian carp	Silver carp
CCE1	AU279337	Zgc:55870	(GGA) ₄ +(GGA) ₇	F: CGAAACTCCATACGACTGAA R: TGAGAGTGTCCCTGCCTCTT	MB		
CCE2	AU301034	Unknown	(TG) ₁₆	F: GCCTGAGGGCTCATAAAGA R: AGCTTCTCACAACGGTTTC	MB		
CCE3	CA967329	MutS homolog 6	(GAT) ₈	F: TGACTTCATCCCAATGACA R: GAAGACTCGGTCCACAGGAG	242		
CCE4	CA967494	Nucleophosmin 1	(GAT) ₃ +(GAT) ₄	F: CAGCGTTGGTGACTCAGAAG R: GTATTGGCCTTGGGTCTTTT			
CCE5	CA965262	Unknown	(AATC) ₆	F: GCGCACACATTCAGAGAAAA R: ACTCCCAAAGATTCCCATCA	MB		
CCE6	CA965579	Unknown	(CA) ₁₃	F: TTTGTCGAGCAGTTCTCTGTC R: GGTTC AAGGTGGCACATTCT	MB		
CCE7	CA965974	Unknown	(TC) ₁₄	F: TCCATAGAGAGAGAGGCACCA R: TGCTTTGTTGAAGGGAGTTG	MB		
CCE8	CA966207	Unknown	(GT) ₁₅	F: AGCCTGCTGCAATGACTTTT R: CCCACCCAAATGCATATTATC	MB		
CCE9	CA966850	Mitochondrial carrier	(AG) ₇ +(AG) ₇	F: AGTGCGCAGCTCAGGTTT R: CCTCATCTTCACCTCCTGCT	198		
CCE10	CA966871	Unknown	(TGA) ₆ +(TGA) ₃ , (GAT) ₆ +(GAT) ₂	F: GATGAGGAGGATGATGAGGTG R: CCCAGTGGACGAGTTAAGGA	242		460
CCE11	CA966887	Zgc:55741	(CAG) ₅ , (AGGTGC) ₆	F: CAACCCACAAGTGAACAAC R: CTGCTGACCAGCCTGTATGA			
CCE12	CA967010	Unknown	(GTTT) ₆ , (GCTT) ₂	F: CGGTTCCGTGATGCTTTAGT R: GCCTTTCATGTAGCCCTTG			
CCE13	CF662230	Warm temperature acclimation related 65 kDa protein	(TTCAA) ₇	F: CTGTGGGCAAGATCAAACCT R: CCTTGTATTGCCCTAATGG	167–242	MB	
CCE14	CF662233	Binding protein	(CA) ₁₇	F: CCGCAGATCCTTGAAGAAGA R: GAGTCGAAGCTAGCGTGAGG	179–185	180–200	160
CCE15	CF662378	Unknown	(TTTC) ₆	F: GCATCGTTTTGGATGCTTTT R: CCTGTCCCAATTGCTATGCT	190–218	210	220
CCE16	CF662378	Unknown	(AT) ₁₁	F: CCGTTACCAAAGCGAACAGT R: TGAATTGGGGACTGCACATA	MB		
CCE17	CF662383	Chromosome 20 open reading frame 149	(TA) ₁₃	F: GAGCGAAAGTAGCGATGGAG R: GCAAGTTTCAGCTCTTCCATT	MB	214	MB
CCE18	CF662449	Pyruvate dehydrogenase (lipoamide) alpha 1	(AC) ₁₃	F: GCTTGAATTATCGGGACTTT R: AGCCAAAACACACTCCTGCT	180–216	180	
CCE19	CF662477	Unknown	(CA) ₂₁ , (TA) ₅	F: TGAGAGGAAAAACTGCACAAGA R: TGAGAAGGAAAACTGCACGA	MB	190	215
CCE20	CF662729	Unknown	(TG) ₂₄	F: GCAGCATATCTAGCCATGACC R: GTGCATTTTTGGAGCCAGAT	MB	MB	MB
CCE21	CF662748	Unknown	(TAA) ₁₀	F: ACGCTGTGCATTTTGTTCAG R: AAGTTGGCACTGACCCTACA	202–309	MB	200
CCE22	CF662761	Nuclear factor I/A	(CA) ₄ +(CA) ₃ + (CA) ₅ +(CA) ₅	F: TGTGAAAGCAAGAAGCAAGG R: AACTCCTGAATTGGCGATGT	169	MB	190
CCE23	CF662764	Unknown	(TATC) ₁₇	F: ATGGTTTGGACTTTGGAGCA R: CGTGAATCCACAGCGATCTA	180–310	MB	MB
CCE24	CF662811	Unknown	(TAA) ₁₄	F: TGCAAACGAGCAAATTGAGT R: ATTTTGCTTGTAGCCGTTG	201–2406		MB
CCE25	CF662846	GNAS complex locus	(AC) ₃ +(AC) ₁₃	F: CGCACCAATATCAAACCACA R: CAGCCACGCTTTCATCAGTA	215–246	180–210	240

(continued on next page)

Appendix A (continued)

Locus	Accession no.	Gene name	Repeat motif	Primer sequences(5'–3')	Common carp	Crucian carp	Silver carp
CCE26	CF662856	Unknown	(TC) ₁₄ , (AC) ₉ , (AT) ₅ , (AC) ₃	F: TGTGAGAAGCAGAGCGATATT R: TCAGTATTATGTGTGTTTTCCA	238–280		
CCE27	CF662926	Unknown	(CA) ₃ +(CA) ₁₅	F: CACGCCTCTCTTCCTTCA R: GGATACCAACGCAGGCTGT	234–252	MB	242–246
CCE28	CF662942	Myosin IXB	(TTA) ₉	F: CCATGTTGTCCTGTTGTGTTT R: ACACAAAAAGTTGGGTTAGACA	238–280		
CCE29	CF662952	SMT3B protein	(TC) ₁₆	F: CAGCAACAGACAGGAGGACA R: CCGCAATTAACAATCCCAAC	190–216	234	MB
CCE30	CF662964	Unknown	(GT) ₃ +(GT) ₂₆	F: CAGAATCAGCAAAGCCAACA R: CACAGAAAACATCCCACACAGA	MB	170	120
CCE31	CF662969	Unknown	(CA) ₁₄	F: TGGTTGTGCACCCAGATGTTT R: AGATGCTTTCGTCCTGCATT	196–238	200–242	180–210
CCE32	CF662995	H3 histone, family 3B	(CT) ₁₄ , (GT) ₁₁ +(GT) ₇	F: ATCTGGTGGGTCTGTTCGAG R: ATACAAAACCCCGTCTGATT	MB	MB	MB
CCE33	CF663003	Ring finger protein 128	(TC) ₈ , (TG) ₁₃	F: GACGTTTACAGCCAGAGTGA R: CAGTGTGATCACCCACAGAGA	247	220–242	MB
CCE34	CF663085	Unknown	(ATT) ₂ +(ATT) ₄ , (GTT) ₂ , (ATT) ₅	F: GCGGCAAAAAAGGACTAAGG R: TGA AACCGTCATTGTTTTCC	MB	MB	MB
CCE35	CF660468	c-Myc	(CT) ₂₇	F: AGGGACCCACAGCCTAATTC R: ACCGAACTTGTCCAAAATGC	170–242		MB
CCE36	CF660503	Unknown	(AT) ₇ +(AT) ₃	F: AACATATCCAGGCTCCTCTATCC R: GGGAACGTATGTAAACCGTGA	MB		
CCE37	CF660529	SUMO conjugating enzyme activity	(TC) ₁₀	F: CAGCAACAGACAGGAGGACA R: CCGCAATTAACAATCCCAAC	185–230	236	220
CCE38	CA966616	Unknown	(AAT) ₁₂	F: GGCAGTCAGTCACAGCAT R: CCGATACTGGGAGATCATGG	242	MB	MB
CCE39	CA968433	Unknown	(TAT) ₅	F: TCCAGGCAACCACAAACATA R: GCCCAAATAAAAAGCCAACAC	195	MB	215
CCE40	CF660577	Unknown	(TC) ₃ +(TC) ₇	F: GACTAATTCCTCCGGATGGAT R: TTCGAGGGCGATAAAAACT			
CCE41	CF663110	unknown	(TTTC) ₅ +, (TTTC) ₃	F: CAGAGGAAACAGCGGGATAC R: TGGCTTGCAGCTTTACAGAA	195–290	215–244	MB
CCE42	CF660696	Unknown	(CCA) ₅ +(CCA) ₂	F: ATCGTGGTCTCATCCGAAAAG R: CGTGGCGTATTGTTTGATTG	230	242	217
CCE43	CF660711	Human dynein heavy chain	(AAATT) ₆	F: CTCCCATGTGAGTGTGGAGA R: TGCACTGTAAGTCGCTTTGG	170–180	MB	MB
CCE44	CF663113	Reverse transcriptase	(CA) ₁₁	F: CCCCAAAAGCTTGCCAAT R: ACAAGAAGTTGGGTGGCTTG	MB	MB	MB
CCE45	CF660760	E1A binding protein p400	(AAAG) ₃ , (AATG) ₅	F: GAGGCAGAGGCTTGTCTGAG R: AGCCCTGAAGTGATCTCCA	180–216	270	110
CCE46	CF661014	Unknown	(TG) ₁₁	F: CAAAGATCCCATCCGTTTCAG R: ACCCTATCTTGCTCCGATT	231–308	220–250	240–246
CCE47	CF661036	Unknown	(AC) ₁₂ +, (AC) ₅ , (AC) ₁₁	F: AACGCAAAAAGTAACGTAATGC R: TGCATGTGAGTGTGTGAGC	249	160	270
CCE48	CF661039	Unknown	(TG) ₁₅	F: CAGCGAAAAGAAAAGGAGCAT R: TTCCAGCCATTCATGTTTGA	201–270	170–210	
CCE49	CF661049	Unknown	(CAG) ₆ , (AC) ₁₆	F: CTGAAGACCCGCAAAGAGTC R: CAAACGGGCTCGTCATTAGT	230–260		MB
CCE50	CF661118	Unknown	(TCA) ₇	F: ACCTGTTCCTGTATCAGCA R: TTGCTGATGTGATTGTTGAGC			
CCE51	CF661207	Unknown	(ACC) ₄ , (ATC) ₃ +(ACC) ₃	F: CCGGAGAAGATGGGAAAAT R: GGGATGAGGATGACGAAGAA	217–238	230	238
CCE52	CF661239	Unknown	(TC) ₁₇	F: TGTGTGTTCTTCATGAGTG R: AGCACATTGATTTTCTGATGACA	MB	MB	
CCE53	CF661254	Unknown	(AC) ₈	F: TGAACGCAATATGCCAATA R: CCAGTGGTGAAGCAGAGTT	147–238	150–154	
CCE54	CF661341	Transducer of ErbB-2.1	(TC) ₁₈	F: TGTTCCCAAGGGTGTGTTGTT R: CGTGTGGGAGAAAACACAAGC	MB	120	

Appendix A (continued)

Locus	Accession no.	Gene name	Repeat motif	Primer sequences(5'–3')	Common carp	Crucian carp	Silver carp
CCE55	CF661412	Polyposis locus-encoded protein	(CAT) ₈ , (CA) ₇	F: GAACGCCACCAATAACTCGT R: TATTGTTCTCCAGGGTTGG			
CCE56	CF661452	Unknown	(CAC) ₃ +(CAC) ₃	F: ATGTTTCAGCAGATGCACGTC R: AGAATGGGGTGGCACAGTAA	260	210–242	238
CCE57	CF661501	Zgc:66100	(AGC) ₅	F: AGTAGCCGGGACTGGACATT R: GAGTTGGGGAGAGACATTG	160–175	141–147	178
CCE58	CF661540	Unknown	(AT) ₂ +(AT) ₇	F: CGAGGTGTTAGCCATCGTTT R: TGGATGAACTCACCGTTCAA	MB		
CCE59	CF661664	Unknown	(TG) ₃ +(TG) ₆	F: TAACATCACCTGGCCATTCA R: TGCAAAGCACATGACGTCTC	186	320	MB
CCE60	CA966927	Unknown	(CT) ₈	F: AGAAGGAGATCGAGCGTCTG R: ATGGAGACATGCGAATAACG	160–201	270	MB

Blank — no product amplified; MB — multiple bands.

Appendix B. Microsatellite-containing genes identified by bioinformatic mining from common carp expressed sequence tags databases

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CF661605	(GT) ₁₃	NP_998111	Zgc:85718; Hypothetical protein LOC405882
CF660468	(CT) ₂₇	D37887	c-Myc
CF661341	(TC) ₈₆	XP_523795	Transducer of ErbB-2.1
CF661049	(CAG) ₆ ,(AC) ₁₆	XP_523795	Zgc:100812
CF663113	(CA) ₁₁ ,(CA) ₇₁	NP_001032317	Reverse transcriptase
CF663003	(TC) ₈ ,(TG) ₁₃	NP_997780	Ring finger protein 128
CF662995	(CT) ₁₄ ,(TG) ₂₀	XP_900402	H3 histone, family 3B
CF662846	(TA) ₁₂ ,(AC) ₁₇	XP_417485	GNAS complex locus
CF662761	(CA) ₂₀	NP_990604	Nuclear factor I/A
CF662449	(AC) ₁₃	NP_998558	Pyruvate dehydrogenase (lipoamide) alpha 1
CF662383	(TA) ₁₃	NP_956302	Chromosome 20 open reading frame 149
CF662233	(CA) ₁₇	CAE17591	Si:busm1–146n9.1;protein coding
CA966850	(AG) ₁₅	NP_957466	Solute carrier family 25 (mitochondrial carrier; dicarboxylate transporter), member 10
CA964662	(CT) ₁₂	XP_692569	HSPC323
CA964381	(AC) ₁₂	NP_001038915	Zgc:153425
CA966967	(ATATAC) ₅ ,(AC) ₉	XP_689457	C13orf18 protein
CF661838	(AC) ₈	NP_001017583	Zgc:110726
CF661568	(TG) ₁₀	XP_515168	Transcription factor 20
CF660853	(AT) ₁₂	BAE93468	Heat shock protein 27
CF660529	(TC) ₁₀	NP_001003422	Zgc:92241
CF660415	(AT) ₈	XP_699400	Chromodomain helicase DNA binding protein 6
CF663029	(TC) ₂₃	NP_990768	Capping protein (actin filament) muscle Z-line, beta
CF662774	(AC) ₁₅	XP_694746	Cold autoinflammatory syndrome 1 homolog; protein coding
CF662735	(GT) ₁₁	XP_695669	Alpha-2,8-sialyltransferase ST8Sia I/V/VI-r2
CF662593	(CA) ₇	XP_417567	Solute carrier family 35, member E2
CF662584	(TA) ₁₂	BAA36619	Complement C3-H1
CF662487	(AC) ₁₀	XP_693445	Cryopyrin isoform b
CF662248	(TA) ₁₁	AAP80678	Fibronectin 1b
CA966862	(GT) ₉	XP_689202	ELM1 protein
CA966346	(TG) ₁₂	NP_938181	Translocase of inner mitochondrial membrane 17 homolog A (yeast)
CA966302	(TG) ₁₃	CAK03625	Microtubule-actin crosslinking factor 1 (Actin cross-linking family 7)

(continued on next page)

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA966108	(AG) ₁₃	XP_419772	ST3A1
CA966040	(AC) ₈	NP_999922	Zgc:76977
CA965987	(AT) ₈	NP_001004667	F11 receptor
CA964678	(TG) ₈	NP_956632	Adaptor-related protein complex 4, beta 1 subunit-like
CA964323	(TC) ₁₁	NP_956229	Capping protein (actin filament) muscle Z-line
CA970379	(TC) ₁₂	XP_696232	Hypothetical protein LOC567835
CA969687	(GA) ₁₁	XP_426614	Sodium and chloride-activated ATP-sensitive potassium channel
CA969380	(GT) ₁₀	NP_999855	Zgc:56077
CA969131	(TA) ₁₂	XP_686493	Atp6v1c11 protein
CA968310	(CA) ₁₀	NP_998558	Pyruvate dehydrogenase (lipoamide) alpha 1
CA968209	(AG) ₁₇	XP_695682	Liprin-alpha 3
CA968074	(AC) ₉	XP_912147	Itch itchy
CA967654	(CT) ₁₈	NP_998581	Zgc:56567
CA967628	(AC) ₁₁	NP_001038311	Hypothetical protein LOC557946
CA967558	(TA) ₁₀	XP_687119	Leucine-rich alpha-2-glycoprotein 1
CA967498	(AC) ₈	NP_998139	Transducer of ERBB2, 1b
AU301113	(TG) ₈	NP_998259	Zgc:76908
CF662942	(TTA) ₁₀	XP_690454	Myosin IXB
CA964345	(CAT) ₁₀	NP_001007395	Zgc:103414
CA964287	(GAT) ₁₁	AAH68344	Nuclear autoantigenic sperm protein (histone-binding)
CA969551	(GGA) ₉	NP_001002697	Zgc:92629
CA968726	(AGC) ₂₀	AAD23878	Vitellogenin precursor (Vtg)
CA967494	(GAT) ₂₃	NP_955460	Nucleophosmin 1
CA967329	(GAT) ₈	AAL04170	MutS homolog 6
AU279337	(GGA) ₁₃	NP_956437	Zgc:55870
CA968776	(CAG) ₈ ,(GA) ₁₅	NP_001014057	Myocyte enhancer factor 2C
CA967202	(TAT) ₅	XP_709029	Aldolase a, fructose-bisphosphate
CF661915	(AGA) ₅	NP_957031	Finkel-Biskis-Reilly murine sarcoma virus (FBR-MuSV) ubiquitously expressed (fox derived); ribosomal protein S30
CF661726	(AGC) ₆	NP_001018376	Zgc:109973
CF661501	(AGC) ₁₀	AAH67667	Zgc:66100
CF661452	(CAC) ₇	XP_695381	Im:7150932
CF661348	(CAT) ₇	NP_999890	Zgc:77223
CF661279	(GCT) ₈	CAK11492	Myocyte enhancer factor 2a
CF661207	(AGC) ₉ ,(ACC) ₁₃	AAH44435	Wu:fi75b02
CF661154	(TTA) ₁₀	XP_684469	Connectin/titin
CF661046	(GCT) ₅	NP_957421	Surfeit gene 4
CF660958	(CCT) ₈	CAD61268	Isoleucine-tRNA synthetase 2, mitochondrial
CF660760	(AAATAA) ₅ ,(GAAT) ₆	XP_691425	E1A binding protein p400
CF660747	(TTTG) ₄	XP_696963	Dysferlin interacting protein 1
CF660519	(GCA) ₇	NP_001019593	Zgc:112335
CF660508	(ATT) ₆	NP_001002360	Zgc:92520
CF660452	(AGA) ₅	AAI08082	Troponin T2, cardiac
CF660447	(GAA) ₁₀	XP_707007	High mobility group protein
CF660421	(CAG) ₆	NP_001002050	Zgc:86607
CF662974	(GAT) ₇	NP_001013502	Zgc:113084
CF662973	(ATT) ₇	XP_695568	Chimerin (chimaerin) 2
CF662868	(CAG) ₆	XP_690532	Microtubule-actin crosslinking factor 1
CF662616	(TGC) ₅	NP_001007306	Zgc:92257
CA965454	(AGG) ₅	NP_998388	Zgc:85717
CF662469	(CAT) ₁₁	NP_840083	Selenoprotein P, plasma, 1b
CF662354	(CAG) ₈	NP_991292	Nuclear receptor subfamily 1, group d, member 1
CF662316	(CAG) ₈	XP_691434	Glucagon receptor precursor
CF662311	(GAA) ₆	NP_001004676	Zgc:101844
CA967240	(TAT) ₇	XP_699484	Secretory granule proteoglycan core protein precursor
CA966887	(AGC) ₆ ,(AGGTGC) ₇	NP_997929	Zgc:55741
CA966747	(CA) ₅	XP_693776	RNA binding motif protein 10

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA966300	(TCA) ₈	XP_528082	Phytanoyl-CoA hydroxylase interacting protein
CA966248	(GAGCCA) ₁₈ , (AAG) ₆	NP_857636	Troponin T3b
CA965984	(AGA) ₆	NP_001017690	Zgc:112000
CA965899	(GAT) ₉	XP_001086511	Glutamate receptor 6 isoform 1 precursor
CA965639	(GAT) ₆	XP_701006	Zinc finger protein 658
CA965557	(CCT) ₅	CAK04811	Zgc:65774
CA965343	(TTA) ₆	NP_001002332	Zgc:92414
CA965288	(TCC) ₁₀	NP_001002039	Fibrinogen alpha chain
CA965027	(TCA) ₇	XP_706128	LOC562542
CA964909	(TGA) ₅	NP_571318	Complement C3-H1
CA964596	(TCA) ₅	NP_998644	CCR4-NOT transcription complex, subunit 8
CA964457	(CTG) ₅	NP_001004551	Zgc:92099
CA964438	(GAT) ₁₁	AAH45917	High-mobility group box 1
CA964304	(TAC) ₆	NP_001003486	Zgc:92039
CA964127	(AAT) ₅	Q90YQ2	40S ribosomal protein S21
CA969997	(AAT) ₇	XP_702631	LOC555065
CA969923	(AAG) ₇ , (TGTC) ₄	XP_691442	Epidermal growth factor receptor pathway substrate 8-like protein 3
CA969764	(GAG) ₇	XP_689251	G protein-coupled receptor kinase 4 isoform alpha
CA969760	(GGA) ₈	NP_958484	Signal sequence receptor, alpha
CA969679	(GTTTC) ₃ , (GCA) ₆	XP_001171667	Hepatic leukemia factor
CA969639	(GCT) ₅	NP_080682	Zinc finger protein 364
CA969399	(CTG) ₈	XP_418646	Claudin-12
CA969088	(GAG) ₅	NP_001017895	Zgc:110609
CA969023	(TCC) ₆	AAS89344	Similar to C9 orf16
CA968990	(AAC) ₅	NP_958901	Nucleobindin 2a
CA968805	(AGG) ₇	XP_692494	Quiescin/sulfhydryl oxidase
CA968552	(TGG) ₅	CAK11306	Zgc:92533
CA968285	(GGT) ₇	AAH97003	Zgc:113907
CA968195	(GTT) ₅	NP_956378	Phosphomannomutase 2
CA967932	(AGG) ₇	XP_698867	LOC570305
CA967632	(GCT) ₅	NP_001002470	Surfeit gene 4
CA967614	(CTG) ₆	NP_997832	Dihydrolipoamide S-acetyltransferase
AU301583	(ATT) ₇	XP_001115912	NADH:ubiquinone oxidoreductase MLRQ subunit homolog
AU300986	(ATG) ₅	NP_001003427	Zgc:92726
AU279303	(TCAGCA) ₄ , (CAT) ₇	NP_001001400	ARP8 actin-related protein 8 homolog (yeast)
CF662879	(TATC) ₁₄	XP_684937	Zinc finger protein 648
CF662630	(ATTC) ₄ , (TATC) ₂₆ , (TTATAAT) ₉ , (TTATATAT) ₆	NP_001032489	Zgc:123275
CA969499	(TTTA) ₈	XP_001057313	FBP interacting repressor
CA967146	(CTTT) ₆	AAO65266	Thioredoxin reductase 1
CF661473	(TTTG) ₄	XP_692098	LOC398632 protein
CA966905	(TTTA) ₅	NP_956379	Solute carrier family 16
CF661367	(TTTG) ₆	XP_236599	Armadillo repeat containing 8
CF660418	(AAAT) ₄	NP_998215	Zgc:56053
CF662641	(TTAT) ₄	NP_001034366	Signal recognition particle receptor
CA966288	(TATT) ₆	XP_695213	Autoantigenic hnRNP-associated with lethal yellow
CA965791	(TTTA) ₄	AAH76251	S100 calcium binding protein, beta (neural)
CA965724	(TATG) ₄	CAH69031	Synaptosome-associated protein 25a
CA965042	(ATTT) ₅	AAH59655	Zgc:73337
CA965028	(TTTG) ₄	XP_542752	LOC485633
CA964614	(GTCC) ₅ , (TTCT) ₄	XP_709635	Zgc:158409
CA969807	(AAAT) ₄	NP_001032669	Zgc:123203
CA969766	(TGTT) ₄ , (TTTTATT) ₄	NP_001003482	Zgc:91810
CA969635	(GTTT) ₅	XP_684451	LOC556535
CA969577	(TTTG) ₄	XP_683649	Short transient receptor potential channel 4
CA969141	(TCTT) ₄	XP_696537	Alpha-tectorin precursor

(continued on next page)

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA968247	(GAT) ₂₃ , (TTTG) ₆	XP_689519	High mobility group protein 4
CO729446	(AGAC) ₅	NP_001003993	Zgc:91930
AU279363	(TATT) ₆	AAO26406	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha b
CF660711	(AAATT) ₆	XP_694607	Human dynein heavy chain (DHC)
CF662230	(GTGATG) ₃ , (TTCAA) ₁₀	NP_001027761	Warm temperature acclimation-related 65kDa protein
CA967019	(ATTCT) ₇	NP_001034197	Ripply1
CF661881	(AAATA) ₄	CAC83659	Parvalbumin
CF660601	(CCCAG) ₃	NP_996963	Zgc:76867
CF660585	(GGAGA) ₃	XP_697344	LOC568891
CF662652	(AATTC) ₃	XP_688980	LOC560483
CF662367	(CCCCT) ₃	XP_511489	Isolog of yeast sui1 and rice gos2; putative
CF662350	(GGATG) ₃	NP_955879	Voltage-dependent anion channel 2
CA965640	(GTTTT) ₆	NP_958448	Inhibitor of DNA binding 2
CA964884	(TTTAC) ₃	XP_700558	LOC571836
CA964379	(GATGG) ₄	NP_955879	Voltage-dependent anion channel 2
CA970368	(GGTCA) ₃	NP_998694	Transmembrane protein 50A
CA969918	(TTTTC) ₄	NP_001002690	Zgc:91826
CA969902	(TTATA) ₃	XP_683936	LOC556114
CA969262	(CCCTT) ₃	NP_001038412	LOC560935
CA968904	(ATTTT) ₄	NP_956701	Zgc:64005
CA968885	(ATTTT) ₃	NP_991248	Zgc:77636
CA968399	(TGATC) ₃	AAH53305	Zgc:64204
CA967400	(TCTCC) ₃	CAI11851	Inter-alpha (globulin) inhibitor H2
CA965262	(AATC) ₇	Unknown	Unknown
CA965974	(TC) ₁₄	Unknown	Unknown
CA967010	(GTTT) ₆	Unknown	Unknown
CF662729	(TG) ₂₄	Unknown	Unknown
CA966927	(CAT) ₉ , (CT) ₈	Unknown	Unknown
CF661036	(AC) ₄₀	Unknown	Unknown
CF663110	(TTTC) ₁₀ , (TTTA) ₅	Unknown	Unknown
CA966616	(ATT) ₁₂	Unknown	Unknown
CF662969	(CA) ₁₅	Unknown	Unknown
CF662952	(TC) ₁₇	Unknown	Unknown
CF662926	(CA) ₁₅	Unknown	Unknown
CF662811	(TAA) ₁₄	Unknown	Unknown
CF662748	(TAA) ₁₀	Unknown	Unknown
AU312527	(AC) ₁₆ , (CA) ₂₄ , (AG) ₁₃ , (AC) ₁₃ , (CT) ₁₀ , (TA) ₁₁	Unknown	Unknown
CF661118	(TCA) ₇	Unknown	Unknown
CA968041	(AC) ₁₁	Unknown	Unknown
CF661085	(AC) ₂₄₄	Unknown	Unknown
CF661239	(TC) ₁₈	Unknown	Unknown
CF661175	(CT) ₁₈	Unknown	Unknown
CF660356	(TG) ₂₇	Unknown	Unknown
CF663004	(TG) ₁₅	Unknown	Unknown
CF661039	(ATAAA) ₁₀ , (TG) ₁₅	Unknown	Unknown
CF660911	(TC) ₁₇	Unknown	Unknown
CF662964	(GT) ₃₀	Unknown	Unknown
CF662856	(TC) ₁₄ , (CA) ₁₀	Unknown	Unknown
CF662727	(GT) ₂₈ , (GA) ₁₇	Unknown	Unknown
CF662666	(CT) ₁₃	Unknown	Unknown
CF662477	(AC) ₂₁	Unknown	Unknown
CA966751	(GT) ₁₄	Unknown	Unknown
EC394874	(AT) ₁₆	Unknown	Unknown
CA966207	(TG) ₁₅	Unknown	Unknown
CA965819	(AC) ₂₄	Unknown	Unknown
CA965591	(AC) ₁₅	Unknown	Unknown

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA965579	(CA) ₁₃	Unknown	Unknown
CA964281	(TC) ₁₇ ,(TCAAGT) ₄	Unknown	Unknown
CA970012	(GT) ₁₇	Unknown	Unknown
CA969745	(CA) ₁₅	Unknown	Unknown
CA969558	(TG) ₁₈ , (AG) ₁₁	Unknown	Unknown
CA969253	(CT) ₉ ,(ATATAC) ₆ , (TA) ₁₉	Unknown	Unknown
CA968600	(AC) ₁₂	Unknown	Unknown
CA968276	(AC) ₁₆ ,(TTTTA) ₄	Unknown	Unknown
CA967897	(TG) ₁₉	Unknown	Unknown
AU301579	(TC) ₈ ,(TG) ₁₃	Unknown	Unknown
AU301578	(AT) ₂₃	Unknown	Unknown
AU301034	(TG) ₁₆	Unknown	Unknown
AU279309	(GT) ₃₃	Unknown	Unknown
AU240295	(AT) ₁₇	Unknown	Unknown
EC394430	(TC) ₁₀	Unknown	Unknown
CA966937	(CA) ₁₇	Unknown	Unknown
CF661820	(AC) ₈	Unknown	Unknown
CF661750	(GT) ₁₂	Unknown	Unknown
CF662107	(AC) ₁₁	Unknown	Unknown
CF661992	(CT) ₈	Unknown	Unknown
CF661945	(CA) ₁₅	Unknown	Unknown
CF661664	(TG) ₁₁	Unknown	Unknown
CF661540	(AT) ₁₀	Unknown	Unknown
CF661372	(CT) ₁₀	Unknown	Unknown
CF661320	(CA) ₁₅	Unknown	Unknown
CF661254	(AC) ₈	Unknown	Unknown
EC394840	(AT) ₁₈	Unknown	Unknown
CF661014	(TG) ₁₁	Unknown	Unknown
CF660583	(AT) ₁₂	Unknown	Unknown
CF660577	(TC) ₁₁	Unknown	Unknown
CF660503	(TATG) ₄ ,(AT) ₁₁	Unknown	Unknown
CF662963	(GT) ₁₀	Unknown	Unknown
CF662854	(AG) ₉	Unknown	Unknown
CF662813	(TG) ₁₀	Unknown	Unknown
CF662805	(TA) ₈	Unknown	Unknown
CF662786	(AC) ₁₉	Unknown	Unknown
CF662767	(TC) ₁₁ ,(CA) ₁₂	Unknown	Unknown
CF662711	(GT) ₁₁	Unknown	Unknown
CF662703	(TAAATA) ₆ ,(TA) ₁₇	Unknown	Unknown
CF662699	(GT) ₁₁	Unknown	Unknown
CF662693	(TG) ₉	Unknown	Unknown
CF662576	(CA) ₁₄	Unknown	Unknown
EC394778	(TA) ₈	Unknown	Unknown
CF662415	(TA) ₁₁	Unknown	Unknown
CF662362	(AT) ₈	Unknown	Unknown
CA967218	(TC) ₉ ,(TA) ₈ ,(TTTA) ₆ (ATTTA) ₅	Unknown	Unknown
CA967049	(AC) ₁₂	Unknown	Unknown
CA966906	(TG) ₈	Unknown	Unknown
CA966772	(AG) ₈	Unknown	Unknown
CA966559	(GT) ₁₀	Unknown	Unknown
CA966546	(GT) ₁₁	Unknown	Unknown
CA966507	(AT) ₁₂	Unknown	Unknown
CA966463	(CA) ₁₃	Unknown	Unknown
CA966371	(TG) ₉	Unknown	Unknown
CA966295	(GA) ₁₂	Unknown	Unknown
EC394780	(AG) ₈	Unknown	Unknown

(continued on next page)

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA966278	(CA) ₁₁	Unknown	Unknown
CA966154	(TG) ₁₂	Unknown	Unknown
CA965982	(GT) ₁₃	Unknown	Unknown
CA965954	(TG) ₁₂	Unknown	Unknown
CA965798	(CT) ₁₃	Unknown	Unknown
CA965670	(TG) ₁₁	Unknown	Unknown
CA965125	(GT) ₁₀	Unknown	Unknown
CA964743	(AT) ₈	Unknown	Unknown
CA964693	(GT) ₉	Unknown	Unknown
CA964652	(TG) ₁₄	Unknown	Unknown
CA964480	(AT) ₈	Unknown	Unknown
CA964446	(GA) ₁₈	Unknown	Unknown
CA970416	(GT) ₁₀	Unknown	Unknown
CA970311	(TA) ₁₁	Unknown	Unknown
CA970268	(GT) ₁₀	Unknown	Unknown
CA970199	(CT) ₁₁	Unknown	Unknown
CA970127	(TC) ₇	Unknown	Unknown
CA970093	(AC) ₁₀	Unknown	Unknown
CA970021	(GT) ₇	Unknown	Unknown
CA969947	(AT) ₁₀	Unknown	Unknown
CA969907	(TG) ₈	Unknown	Unknown
CA969836	(TA) ₁₀	Unknown	Unknown
CA969694	(AC) ₁₆	Unknown	Unknown
CA969663	(TG) ₁₁	Unknown	Unknown
CA969599	(GAAAGA) ₄ , (AG) ₁₃	Unknown	Unknown
CA969563	(GT) ₁₁	Unknown	Unknown
CA969482	(CG) ₈	Unknown	Unknown
CA969443	(GA) ₇	Unknown	Unknown
CA969315	(TG) ₁₁	Unknown	Unknown
CA968932	(AT) ₈	Unknown	Unknown
CA968851	(CT) ₈	Unknown	Unknown
CA968839	(TG) ₈	Unknown	Unknown
CA968256	(AT) ₈	Unknown	Unknown
CA968215	(TCATC) ₄ , (AC) ₉	Unknown	Unknown
CA968030	(GT) ₉	Unknown	Unknown
CA967842	(GT) ₉	Unknown	Unknown
CA967677	(AT) ₉	Unknown	Unknown
CA967606	(CA) ₁₄	Unknown	Unknown
CA967595	(TG) ₁₁	Unknown	Unknown
CA967540	(AG) ₁₂	Unknown	Unknown
AU300957	(TC) ₁₃ , (CTCTTT) ₆	Unknown	Unknown
AU300951	(CA) ₈	Unknown	Unknown
AU279279	(GT) ₈ , (TCCTTT) ₅	Unknown	Unknown
AU240311	(GA) ₁₀	Unknown	Unknown
CF661412	(CTTCCT) ₄ , (CAT) ₈ , (AC) ₇	Unknown	Unknown
CF660544	(TTA) ₁₀	Unknown	Unknown
EC394891	(CAA) ₁₀	Unknown	Unknown
CF663085	(ATT) ₁₄	Unknown	Unknown
CF662823	(AAT) ₈ , (TAT) ₁₂	Unknown	Unknown
CF662665	(AAC) ₇ , (AAT) ₉	Unknown	Unknown
CA967254	(ATA) ₆ , (AAT) ₁₃ , (TTA) ₁₁ , (TTATTT) ₅	Unknown	Unknown
CA966871	(TGA) ₁₀ , (GAT) ₉	Unknown	Unknown
CA970182	(CAT) ₉	Unknown	Unknown
CA969696	(TTA) ₉	Unknown	Unknown
CA969664	(AGA) ₃₇	Unknown	Unknown
CA969266	(TCA) ₁₁	Unknown	Unknown
CA968077	(CAG) ₉	Unknown	Unknown

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA967688	(ATA) ₁₀	Unknown	Unknown
AU312506	(AAC) ₇	Unknown	Unknown
CF662209	(TCA) ₈ ;(CAG) ₅	Unknown	Unknown
CA967201	(TCA) ₆	Unknown	Unknown
CF661373	(ATA) ₈	Unknown	Unknown
CF661229	(ACT) ₈	Unknown	Unknown
EC394877	(GAT) ₁₁	Unknown	Unknown
CF660981	(AGC) ₈	Unknown	Unknown
CF660821	(TTAATA) ₈ ;(CAA) ₇	Unknown	Unknown
CF660696	(CCA) ₈	Unknown	Unknown
CA968433	(TAT) ₅	Unknown	Unknown
CF660367	(GTT) ₅ ;(AT) ₈	Unknown	Unknown
CF662825	(TTA) ₁₁	Unknown	Unknown
CF662467	(CTC) ₇	Unknown	Unknown
CF662396	(ATA) ₅	Unknown	Unknown
CA967207	(GTT) ₅	Unknown	Unknown
EC394775	(GAG) ₇	Unknown	Unknown
CA966859	(TTA) ₈	Unknown	Unknown
CA966713	(GAG) ₆	Unknown	Unknown
CA966662	(AGC) ₅	Unknown	Unknown
CA966370	(TTA) ₇	Unknown	Unknown
CA966326	(TAA) ₆	Unknown	Unknown
CA966226	(TTG) ₇	Unknown	Unknown
EC394821	(TAA) ₇	Unknown	Unknown
CA966155	(TTG) ₈	Unknown	Unknown
CA965977	(TGT) ₈	Unknown	Unknown
CA965011	(TGA) ₅	Unknown	Unknown
CA964998	(AAT) ₇ ;(CAT) ₈	Unknown	Unknown
CA964948	(AAT) ₅	Unknown	Unknown
CA964674	(TTA) ₁₁	Unknown	Unknown
CA964635	(TCA) ₆	Unknown	Unknown
CA964615	(AAG) ₇	Unknown	Unknown
CA964612	(ACC) ₉	Unknown	Unknown
CA964552	(GAG) ₇	Unknown	Unknown
CA964514	(AAG) ₁₁	Unknown	Unknown
CA964507	(ATG) ₆	Unknown	Unknown
EC394879	(GAA) ₈	Unknown	Unknown
CA964369	(TTA) ₁₀	Unknown	Unknown
CA964284	(GCA) ₆	Unknown	Unknown
CA964207	(AAT) ₆	Unknown	Unknown
CA970282	(TGT) ₆	Unknown	Unknown
CA969987	(TTTC) ₄ ;(TTG) ₅	Unknown	Unknown
CA969779	(TCA) ₇	Unknown	Unknown
EC394771	(AAT) ₆	Unknown	Unknown
CA969762	(CGG) ₅	Unknown	Unknown
CA969531	(TTA) ₇	Unknown	Unknown
CA969028	(ATT) ₅	Unknown	Unknown
CA968798	(TAT) ₈	Unknown	Unknown
CA968785	(GAG) ₇	Unknown	Unknown
CA968727	(CAT) ₈ ;(CAG) ₆	Unknown	Unknown
CA968354	(TTA) ₇	Unknown	Unknown
EC394841	(ATT) ₅	Unknown	Unknown
CA968272	(AAC) ₆	Unknown	Unknown
CA968008	(TAT) ₈	Unknown	Unknown
CA967748	(AAG) ₁₄ ;(TCA) ₈	Unknown	Unknown
CA970434	(GGA) ₅	Unknown	Unknown
AU300985	(GAT) ₅	Unknown	Unknown

(continued on next page)

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA967624	(AGG) ₇	Unknown	Unknown
AU279317	(GGT) ₆	Unknown	Unknown
EC394810	(TTG) ₅	Unknown	Unknown
CF662060	(GATA) ₅₀	Unknown	Unknown
CF661552	(TGTT) ₄ , (TCTA) ₁₇	Unknown	Unknown
CF662764	(CAT) ₉ , (TATC) ₁₇	Unknown	Unknown
CF662378	(AT) ₁₁ , (TTTC) ₇	Unknown	Unknown
CA965453	(TAGA) ₁₁ , (TTATTT) ₄	Unknown	Unknown
CA964781	(CT) ₁₉ , (TCTA) ₉	Unknown	Unknown
CA970294	(GATA) ₂₆	Unknown	Unknown
CA968409	(TATC) ₇	Unknown	Unknown
AU301166	(GGAT) ₈ , (GGAT) ₁₀ , (GACA) ₂₃ , (AGAT) ₂₁ , (GACAGATAGATA) ₁₁	Unknown	Unknown
CF662039	(TTTG) ₆ , (TTGTTTAT) ₄	Unknown	Unknown
CF662033	(AAAG) ₆	Unknown	Unknown
CF662019	(ATT) ₅	Unknown	Unknown
CF661896	(AAAC) ₆	Unknown	Unknown
CA967189	(AAAT) ₄	Unknown	Unknown
CF661309	(TTAT) ₆	Unknown	Unknown
CF661252	(TTTT) ₆	Unknown	Unknown
CF661184	(TGAA) ₆ , (TTTCAT) ₄ (TTTCA) ₅	Unknown	Unknown
CF660878	(TTTG) ₄	Unknown	Unknown
CF660854	(TTTC) ₆ , (TTTA) ₅	Unknown	Unknown
CF660475	(TTTC) ₇ , (ACAAA) ₃	Unknown	Unknown
CF660426	(TTGT) ₄	Unknown	Unknown
CF662746	(TGAA) ₅	Unknown	Unknown
CF662712	(GT) ₁₀ , (TGTT) ₆ , (CT) ₇	Unknown	Unknown
CF662512	(ATAA) ₅	Unknown	Unknown
CF662236	(AAACA) ₃	Unknown	Unknown
CF662228	(ATTT) ₃	Unknown	Unknown
CA967087	(TTTA) ₄ , (TG) ₁₁	Unknown	Unknown
CA967003	(TTGT) ₆	Unknown	Unknown
CA966811	(AAAT) ₄	Unknown	Unknown
CA966790	(TGAA) ₄ , (TA) ₁₄	Unknown	Unknown
CA966434	(AGGC) ₄	Unknown	Unknown
CA966206	(TTTA) ₆	Unknown	Unknown
CA965902	(TAAA) ₄	Unknown	Unknown
CA965885	(AAAT) ₆	Unknown	Unknown
CA965506	(TGAA) ₄	Unknown	Unknown
CA965107	(TTTG) ₅	Unknown	Unknown
CA964933	(TAAA) ₅	Unknown	Unknown
CA964672	(AATA) ₅	Unknown	Unknown
CA964619	(CCTCT) ₃	Unknown	Unknown
CA964586	(TTTA) ₅	Unknown	Unknown
CA964436	(CAAA) ₄	Unknown	Unknown
CA964286	(TTTA) ₄	Unknown	Unknown
CA964242	(TTCT) ₉	Unknown	Unknown
CA970112	(TCTT) ₅	Unknown	Unknown
CA969654	(AATA) ₆	Unknown	Unknown
CA969426	(TATT) ₄	Unknown	Unknown
CA969353	(AAAC) ₄	Unknown	Unknown
CA969254	(TTTA) ₅	Unknown	Unknown
CA969056	(TGAC) ₄	Unknown	Unknown
EC394783	(AATA) ₆	Unknown	Unknown
CA968419	(TTTG) ₆	Unknown	Unknown
CA968309	(TATT) ₆	Unknown	Unknown

Appendix B (continued)

Accession no.	Repeat motif	Accession no. of closest homology	Gene name
CA968281	(ATTT) ₆	Unknown	Unknown
CA968234	(AAAT) ₅	Unknown	Unknown
CA968225	(AAAT) ₅	Unknown	Unknown
CA968153	(AATT) ₄	Unknown	Unknown
CA968064	(TA) ₁₄ (AC) ₁₂ (TATT) ₅	Unknown	Unknown
CA968028	(TTTA) ₆	Unknown	Unknown
CA967604	(TCAG) ₆	Unknown	Unknown
AU301090	(GTTG) ₅	Unknown	Unknown
AU301212	(AAAC) ₉	Unknown	Unknown
AU300938	(GTTT) ₅	Unknown	Unknown
CF661212	(TCACT) ₉	Unknown	Unknown
CF662819	(CTCTT) ₆	Unknown	Unknown
CF662436	(AGTGAAA) ₃₅	Unknown	Unknown
CF661592	(TTTGT) ₄	Unknown	Unknown
CF661460	(TTTTG) ₄	Unknown	Unknown
EC394888	(GAGATG) ₄	Unknown	Unknown
CA967181	(TATGATA) ₄ , (ATATG) ₆ , (AAATA) ₇	Unknown	Unknown
CF661296	(TTTGA) ₃	Unknown	Unknown
CF661068	(TAAAA) ₁₀	Unknown	Unknown
CF660830	(AAAAC) ₄	Unknown	Unknown
CF660525	(ATTTT) ₄	Unknown	Unknown
CF660399	(TTTCG) ₃	Unknown	Unknown
CF662907	(TTTAA) ₃ , (CTTT) ₅	Unknown	Unknown
CF662890	(TTCGT) ₃	Unknown	Unknown
CF662625	(CTCGG) ₄	Unknown	Unknown
CF662571	(TTTTG) ₅	Unknown	Unknown
CF662288	(GTTTT) ₅	Unknown	Unknown
CA965820	(CTCAT) ₄	Unknown	Unknown
CA965695	(GAGGA) ₃	Unknown	Unknown
CA964943	(ATAAA) ₃	Unknown	Unknown
CA964706	(GCTGT) ₃	Unknown	Unknown
CA963994	(ACTGT) ₃	Unknown	Unknown
CA969826	(TTTGT) ₃	Unknown	Unknown
CA969469	(TGTTT) ₄ , (TC) ₁₂	Unknown	Unknown
CA969447	(CATGT) ₃	Unknown	Unknown
CA968252	(TTATT) ₃	Unknown	Unknown
CA968037	(TATTG) ₃	Unknown	Unknown
CA967933	(TGTTT) ₃	Unknown	Unknown
CA967755	(ATTTA) ₃	Unknown	Unknown
CA967613	(TAGCT) ₃	Unknown	Unknown
AU300967	(AAAAC) ₃	Unknown	Unknown
AU279275	(TTTTTG) ₃ , (TGTTT) ₄	Unknown	Unknown

References

- Aljanabi, S.M., Martinez, I., 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res.* 25, 4692–4693.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- Chen, C.X., Zhou, P., Choi, Y.A., Huang, S., Gmitter, F.G., 2006. Mining and characterizing microsatellites from citrus ESTs. *Theor. Appl. Genet.* 112, 1248–1257.
- Chen, S.L., Liu, Y.G., Xu, M.Y., Li, J., 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.
- Cho, Y.G., Ishii, T., Temnykh, S., Chen, X., Lipovich, L., McCouch, S.R., Park, W.D., Ayres, N., Cartinhour, S., 2000. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 100, 713–722.
- Cordeiro, G.M., Casu, R., McIntyre, C.L., Manners, J.M., Henry, R.J., 2001. Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. *Plant Sci.* 160, 1115–1123.
- David, L., Rajasekaran, P., Fang, J., Hillel, J., Lavi, U., 2001. Polymorphism in ornamental and common carp strains (*Cyprinus*

- carpio* L.) as revealed by AFLP analysis and a new set of microsatellite marker. *Mol. Genet. Genomics* 266, 353–362.
- Edwards, Y.J., Elgar, G., Clark, M.S., Bishop, M.J., 1998. The identification and characterization of microsatellites in the compact genome of the Japanese pufferfish, *Fugu rubripes*: perspectives in functional and comparative genomic analyses. *J. Mol. Biol.* 278, 843–854.
- Guo, S.W., Thompson, E.A., 1992. Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48, 361–372.
- Gupta, P.K., Varshney, R.K., 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* 113, 163–185.
- Kantety, R.V., Rota, M.L., Matthews, D.E., Sorrells, M.E., 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48, 501–510.
- Liao, X.L., Yu, X.M., Tong, J.G., 2006. Genetic diversity of common carp from two largest Chinese lakes and the Yangtze River revealed by microsatellite markers. *Hydrobiologia* 568, 445–453.
- Liu, Z.J., Cordes, J.F., 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37.
- Maneeruttanarungroj, C., Pongsomboon, S., Wuthisuthimethavee, S., Klinbunga, S., Wilson, K.J., Swan, J., Li, Y., Whan, V., Chu, K.H., Li, C.P., 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.
- Neff, B.D., Gross, M.R., 2001. Microsatellite evolution in vertebrates: inference from AC dinucleotide repeats. *Evolution* 55, 1717–1733.
- Nonneman, D., Waldbieser, G.C., 2005. Isolation and enrichment of abundant microsatellites from a channel catfish (*Ictalurus punctatus*) brain cDNA library. *Anim. Biotechnol.* 16, 103–116.
- Pérez, F., Ortiz, J., Zhinaula, M., Gonzabay, C., Calderón, J., Volckaert, F.A.M.J., 2005. Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus birdy*. *Mar. Biotechnol.* 7, 554–569.
- Rexroad III, C.E., Rodriguez, M.F., Coulibaly, I., Gharbi, K., Danzmann, R.G., Dekoning, J., Phillips, R., Palti, Y., 2005. Comparative mapping of expressed sequence tags containing microsatellites in rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics* 6, 54.
- Schneider, S., Roessli, D., Excoffier, L., 2000. Arlequin: a software for population genetics data analysis, version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.
- Serapion, J., Kucuktas, H., Feng, J.N., Liu, Z.J., 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* 6, 364–377.
- Stallings, R.L., Ford, A.F., Nelson, D., Torney, D.C., Hildebrand, C.E., Moyzis, R.K., 1991. Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes. *Genomics* 10, 807–815.
- Sun, X.W., Liang, L.Q., 2004. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* 238, 165–172.
- Tong, J., Lehnert, S.A., Byrne, K., Kwan, H.S., Chu, K.H., 2002. Development of polymorphic EST markers in *Penaeus monodon*: applications in Penaeid genetics. *Aquaculture* 208, 69–79.
- Tong, J., Yu, X., Liao, X., 2005. Characterization of a highly conserved microsatellite marker with utility potentials in cyprinid fishes. *J. Appl. Ichthyol.* 21, 232–235.
- Wang, H.X., Li, F.H., Xiang, J.H., 2005. Polymorphic EST-SSR markers and their mode of inheritance in *Fenneropenaeus chinensis*. *Aquaculture* 249, 107–114.
- Wohlfarth, G., 1984. In: Mason, I.L. (Ed.), Common carp, in evolution of domesticated animals. Longman, London.
- Yeh, F.C., Yang, R.C., Boyle, T., 1999. POPGENE, Version 1.3.1: microsoft windows-based freeware for population genetic analysis. Available at (www.ualberta.ca/~fyeh). University of Alberta and the Centre for International Forestry Research.
- Yue, G.H., Ho, M.Y., Orban, L., Komen, J., 2004. Microsatellites within genes and ESTs of common carp and their applicability in silver crucian carp. *Aquaculture* 234, 85–98.
- Zhan, A.B., Bao, Z.M., Wang, X.L., Hu, J.J., 2005. Microsatellite markers derived from bay scallop *Argopecten irradians* expressed sequence tags. *Fish. Sci.* 71, 1341–1346.