*Journal of Heredity* 2008:99(2):208–214
doi:10.1093/jhered/esm124
Advance Access publication January 30, 2008



# Exploiting EST Databases for the Development and Characterization of EST–SSRs in the Pacific Oyster (*Crassostrea gigas*)

Hong Yu and Qi Li

From the Department of Aquaculture, Fisheries College, Ocean University of China, Yushan Road 5, Qingdao 266003, China.

Address correspondence to Dr Qi Li at the address above, or e-mail: qili66@mail.ouc.edu.cn.

A total of 147 microsatellite-containing expressed sequence tags (ESTs) (3.63%) were detected from 4053 ESTs of the Pacific oyster (*Crassostrea gigas*) in GenBank. The average density of simple sequence repeats (SSRs) was 1 per 8.25 kb of EST after redundancy elimination. Dinucleotide repeat motifs appeared to be the most abundant type. Sixteen new polymorphic EST–SSRs were developed. The number of alleles per locus varied from 3 to 12, with an average of 5.9 alleles per locus. Marker transferability was tested on 2 other *Crassostrea* species, and 14 loci gave successful amplifications in both species. Twenty EST–SSRs were tested on 3 families of *C. gigas* for examination of inheritance mode of EST–SSRs. Thirty-five tests of segregation ratios revealed 5 significant departures from expected Mendelian ratios, 4 of which confirmed Mendelian expectations when accounting for the presence of null alleles. Null alleles were detected at 3 loci (15.0%) of the 20 loci, and the frequency of null alleles at EST–SSRs was lower than that in genomic SSRs in *C. gigas*. The results obtained in this study suggest that *C. gigas* EST–SSRs will complement the currently available genomic SSR markers and may be useful for comparative mapping, marker-assisted selection, and evolutionary studies.

Microsatellites or simple sequence repeats (SSRs) have been extensively used over the last decade because of their highly interesting properties. The uniqueness and the value of microsatellites arise from their high polymorphism, co-dominant inheritance, extensive genome coverage (occurring both in protein-coding and noncoding regions), and simple detection by polymerase chain reaction (PCR; Powell et al. 1996). These features have proved to be of great interest for population genetics, parentage analysis, and genetic mapping (Bentzen et al. 2001; Li et al. 2006). However, despite these advantages, the isolation and characterization of such markers via traditional methods (i.e., the screening of size-fractionated genomic DNA libraries) are costly and time consuming (Squirrell et al. 2003). In addition, the cross-species transferability of microsatellites derived from anonymous genomic libraries is normally quite limited due to either disappearance of the repeat region or to degeneration of the primer binding sites (Rungis et al. 2004).

In recent years, with the rapid increase of sequence information, the generation of EST-derived microsatellites becomes an attractive alternative to complement-existing SSR collections. ESTs are sequenced portions of complementary DNA copies of mRNA, and they represent part of the transcribed portion of the genome in given conditions (Poncet et al. 2006). By using some computer programs, it is easy to download the ESTs from public databases (GenBank; The European Molecular Biology Laboratory; DNA Data Bank of Japan) and scan for identification of SSRs. The EST databases provide a valuable resource for the development of SSRs. To date, EST–SSRs have been developed in a variety of plants and some important aquaculture species (Carlsson and Reece 2007; Vendramin et al. 2007). Compared with the development of traditional genomic SSRs, the development of EST–SSRs is time and cost-saving once a sufficient amount of ESTs are available. Additionally, EST–SSRs have some intrinsic advantages over genomic SSRs. One advantage is that EST–SSRs may be directly associated with a coding gene and so may be useful for association with known function genes. Also, because ESTs are evolutionary conserved, cross-species PCR amplification of EST–SSRs is expected to be more successful compared with genomic SSRs (Arnold et al. 2002). So EST–SSRs have been useful as anchor markers for comparative mapping, evolutionary studies, detection of selective sweeps, and identifying conserved genomic regions among species and genera (Vigouroux et al. 2002; Gupta et al. 2003).

For *Crassostrea gigas*, genomic microsatellites have been developed in recent years (McGoldrick et al. 2000; Li et al.

2003; Sekino et al. 2003) and used efficiently in population studies, parentage analysis, and genetic mapping (Hubert and Hedgecock 2004; Li and Kijima 2006; Li et al. 2006). However, there also appears to be some drawbacks. Firstly, one concern with regard to the use of SSRs is that they might produce a high level of null alleles. The frequency of microsatellite loci containing null alleles has been shown to be as high as 30% in humans (Callen et al. 1993), 25% in a swallow (Primmer et al. 1995), 16% in rainbow trout (Ardren et al. 1999), and 57% in the Pacific abalone (Li, Park, Kobayashi, and Kijima 2003). Many studies have also suggested the high frequency (>50%) of null alleles in *C. gigas* (McGoldrick et al. 2000; Li et al. 2003). Null alleles are often caused by mutations such as substitutions, insertions, or deletions within one or both primer annealing sites, which prevent amplification (Callen et al. 1993). The presence of null alleles can lead to confounding results when using microsatellites. If null alleles are present but not accounted for, the resulting scoring error of heterozygotes for homozygotes can create an apparent excess of homozygotes in population studies (Jones et al. 1998). In parentage analysis, an undetected null allele may cause rejection of an otherwise correctly assigned parent. In genetic linkage analysis, the null allele should be detected by the apparent noninheritance of an allele if the family structure has been confirmed by Mendelian inheritance in other loci, or at most it will cause loss of information from an assumed homozygote parent if the other parent is homozygous for the same normal allele (Callen et al. 1993). Secondly, the primers necessary for the amplification of the genomic SSRs are frequently species specific, and poor cross-species amplification of microsatellites cloned from *C. gigas* have been reported (Hedgecock et al. 2004). Thus, it is usually difficult to make some meaningful interspecific comparisons using genomic SSRs. Thirdly, although SSRs have been used in the construction of linkage mapping in *C. gigas* (Hubert and Hedgecock 2004), genomic locations of these SSRs developed from anonymous sequences cannot be compared with other species and comparative genome analysis cannot be conducted due to the lack of type I markers which can serve as anchorage points for genomic segments. Therefore, EST–SSRs as rich resources of type I markers associated with known function genes appear to be a promising alternative and can complement the shortages of genomic anonymous SSRs.

In this paper, we firstly develop and characterize a collection of EST–SSRs for *C. gigas* in terms of frequency, information content, gene annotation, transferability to related species, inheritance mode, and the extent of the distribution of the null allele.

## Materials and Methods

### Oyster Materials and DNA Extraction

To screen for polymorphic SSRs, 40 individuals of *C. gigas* collected from coastal waters in Hiroshima, Japan, were used. Sixteen *Crassostrea plicatula* individuals and 16 *Crassostrea ariakensis* individuals, which were collected from Fujian and Shandong provinces in China, respectively, were used for the examination of the transferability of EST–SSRs. *Crassostrea gigas* families were produced by single-pair matings in 2006, and 3 of them with 49–54 D-shaped larvae from each family were selected randomly to test for inheritance mode of EST–SSRs including 16 new loci and 4 loci developed in our previous study (Yu and Li 2007). The adductor muscles collected from fresh adults and 24-h D-shaped larvae were preserved in 95% ethanol at −20 °C until DNA preparation.

DNA extraction was performed by standard proteinase K digestion, phenol–chloroform extraction for adults, and by Chelex-based methods for larvae as described by Li, Park, and Kijima (2003).

### Data Mining for SSR Markers

A total of 4053 Pacific oyster EST sequences were downloaded from GenBank (The National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/sites/entrez, 10 March 2007). The EST data set was scanned and assembled using SeqMan II sequence assembly software (DNASTAR Inc., Madison, WI) and generated 2377 potential unigenes that contain contigs and singletons from all EST sequences. The SSRHUNTER program (http://www.bio-soft.net/dna/) was used to search for SSRs within this unigene set. The criteria used in SSRHUNTER to identify SSRs were as follows: 5 repeats for di-, tri-, and tetranucleotide repeats. ESTs containing SSRs were then annotated using BLAST software as described by Maneeruttanarungroj et al. (2006). The BLAST results were classified into 3 groups: known gene products, hypothetical proteins, and unknown genes.

### Primer Design and PCR Conditions

Primers were designed using the Primer Premier 5.0 program (http://www.premierbiosoft.com/). PCR was performed in 10-μL volumes containing 0.25 U *Taq* DNA polymerase (Takara Inc., Shiga, Japan), 1× PCR buffer, 0.2 mM dNTP mix, 1 μM of each primer set, 1.5 mM $MgCl_2$, and about 100 ng template DNA. PCR was performed on a GeneAmp 9700 PCR System (Applied Biosystems, Crescent City, CA) as follows: 3 min at 94 °C; 35 cycles of 1 min at 94 °C, annealing (for new developed loci, see Table 1; annealing temperatures 45 °C for *CGE006*, 52 °C for *CGE005*, *CGE007*, and *CGE009*) for 1 min, 72 °C for 1 min per cycle; followed by 5 min at 72 °C. Amplification products were resolved via 6% denaturing polyacrylamide gel and visualized by silver staining. A 10-bp DNA ladder (Invitrogen Inc., Carlsbad, CA) was used as a reference marker for allele size determination.

### Data Analysis

The number of alleles ($N$), expected heterozygosity ($H_e$), and observed heterozygosity ($H_o$) were calculated using MICROSATELLITE ANALYSER software (Dieringer and

**Table 1.** Characterization of 16 polymorphic EST–SSRs derived from *Crassostrea gigas*

| Locus | Repeat sequence | Primer sequences (5′–3′) | $T_a$ (°C) | $N_a$ (size range, bp) | Putative function | Primer location | Accession no. |
|---|---|---|---|---|---|---|---|
| *CGE011* | $(TA)_5$ | TAGGAACGGATAGAAACG | 50 | 3 (150–164) | Predicted hypothetical protein | 3′UTR | AJ565497 |
| | | TAAAGGACGGAAACTACATT | | | [*Gallus galtus*] | 3′UTR | |
| *CGE012* | $(AG)_5$ | CTTCAATAGGAAAATAAATC | 47 | 6 (220–230) | Predicted: hypothetical protein | 3′UTR | BQ426900 |
| | | ATACACATCTAAACAAACAGT | | | [*Nasonia vitripennis*] | 3′UTR | |
| *CGE013* | $(AT)_5$ | AGGTTTGCTTGTTGGTGC | 52 | 6 (170–182) | Unknown | Unknown | BQ426844 |
| | | TAACGGTGGGATGAGGAAT | | | | Unknown | |
| *CGE014* | $(TG)_5$ | TTCAGACTTTCCAACCCC | 52 | 8 (236–266) | Predicted: similar to Protein C12orf4 | 3′UTR | BQ426727 |
| | | TTTTCAATCAAATACAACTAACAC | | | [*Apis mellifera*] | 3′UTR | |
| *CGE015* | $(TA)_5$ | TAAAATCTCAATAGGCTCTTCG | 55 | 2 (116–118) | Unknown | Unknown | CX069049 |
| | | GAACGACTGGTTTCGGATG | | | | Unknown | |
| *CGE016* | $(CA)_6$ | ATGGCTACATTATGCTGA | 50 | 4 (190–196) | Unknown | Unknown | CX068967 |
| | | ATCTGGTGCTTTACGAGT | | | | Unknown | |
| *CGE017* | $(AG)_5$ | CGCCATCAAAAGTGTTCA | 52 | 7 (231–249) | Unknown | Unknown | CX068814 |
| | | GGGTATGTCATTGCCAGC | | | | Unknown | |
| *CGE018* | $(CA)_5$ | ACAACATCCAACAACTCTC | 50 | 6 (144–154) | Unknown | Unknown | CK172347 |
| | | CTTCTCCATTTATCCAGC | | | | Unknown | |
| *CGE019* | $(TTG)_5$ | TAACCTCCTGTTCCCCTGT | 55 | 6 (387–408) | SJCHGC01721 protein | 5′UTR | CK172320 |
| | | TCAGCATTTCAAAGATACCCT | | | [*Schistosoma japonicum*] | 3′UTR | |
| *CGE020* | $(TA)_6$ | TGTTTTATTATCATACGCC | 47 | 13 (202–244) | Unknown | Unknown | DV736352 |
| | | TCCAATCTAATACAAAATACC | | | | Unknown | |
| *CGE021* | $(TG)_5$ | ATAGAAAACAGAAGCCAG | 50 | 6 (250–266) | Unknown | Unknown | DV736668 |
| | | GCAGGTGATGTATTGATT | | | | Unknown | |
| *CGE022* | $(TA)_5$ | TGTTACCAGTTCTTCGTG | 52 | 5 (242–254) | c-src tyrosine kinase | 3′UTR | BQ426966 |
| | | TATTTGAGACAAGGAGGC | | | [*Xenopus laevis*] | 3′UTR | |
| *CGE023* | $(TG)_{13}$ | GGAAACCGACCAATCAAAA | 52 | 4 (380–390) | Unknown | Unknown | DV736642 |
| | | GCAGCAAAGGAGAATACAGG | | | | Unknown | |
| *CGE024* | $(AT)_5$ | ACATCATCCAAGACCACG | 52 | 10 (218–244) | Unknown | Unknown | BQ427306 |
| | | ATGCTTATAGTTGAGGTAAATCT | | | | Unknown | |
| *CGE025* | $(ATA)_5$ | TGAGTTACATTCATATCAAATC | 52 | 4 (147–168) | Unknown | Unknown | CX069052 |
| | | AACAGCGAAGTACCAGTC | | | | Unknown | |
| *CGE026* | $(TA)_6$ | GTAGCAGTGGCATTGACG | 54 | 4 (222-230) | Predicted protein | 3′UTR | CX069252 |
| | | ATGGGGAAGATGGCATAT | | | [*Nematostella vectensis*] | CDS | |

Schlötterer 2003). Exact tests for deviations from Hardy–Weinberg equilibrium (HWE) were performed using the GENEPOP program (http://genepop.curtin.edu.au/). Significant levels were calculated per locus using Bonferroni method (Rice 1989). Inheritance ratios were analyzed at 20 EST–SSRs loci in this study. The *G*-test was used to measure the goodness-of-fit for expected Mendelian segregation ratios (1:1, 1:2:1, and 1:1:1:1) at the 0.01 probability level.

## Results

### EST–SSRs Isolated from ESTs Database of *C. gigas*

A total of 147 microsatellite-containing EST sequences (3.63%) were detected from 4053 ESTs in the Pacific oyster dbEST. After clustering and assembly of the ESTs, 2377 unigenes were identified, including 1919 singletons and 458 contigs, of which the total length was 1147.352 kb. One hundred and thirty-nine unique SSRs were found within 2377 unigenes, which represented an average density of 1 SSR per 8.25 kb. The proportion of microsatellite unit sizes was not evenly distributed: dinucleotide repeats were the most abundant within *C. gigas* ESTs, accounting for 122 loci (87.77%). Tri- and tetranucleotide repeats were found at lower frequencies with 14 loci (10.07%) and 3 loci (2.16%), respectively (Table 2). Regarding dimeric SSRs, the AG/CT microsatellite type was predominant (Figure 1), whereas CG microsatellite type was not detected.

One hundred and thirty microsatellite-containing EST sequences (after clustering and assembly) were BLAST-searched against GenBank. Twenty-eight ESTs (21.54%) were known gene products, 15 ESTs (11.54%) were hypothetical proteins, and 87 ESTs (66.92%) were unknown genes.

### Development of EST–SSRs and Level of Polymorphism

Thirty of the 139 SSR loci with sufficiently long flanking sequences were selected for microsatellite marker optimization. Of the 30 potential microsatellite markers, 12 were not easily amplified, 2 were monomorphic, and 16 were found to be polymorphic among 40 individuals of *C. gigas*. Of the 16 polymorphic loci, 2 (*CGE019* and *CGE023*) produced products approximately 150–200 bp larger than expected.

**Table 2.** Summary of EST–SSRs from the EST database of *Crassostrea gigas*

| Parameter | Value |
|---|---|
| Total number of ESTs searched | 4053 |
| Total number of ESTs with SSR | 147 |
| Total number of unigenes | 2377 |
| Total number of contigs | 458 |
| Total number of singletons | 1919 |
| Total length of unigenes (kb) | 1147.352 |
| Total number of unique SSRs | 139 |
| Frequency of SSRs | 1 per 8.25 kb |
| Repeat type | |
| Dinucleotide | 122 (87.77%) |
| Trinucleotide | 14 (10.07%) |
| Tetranucleotide | 3 (2.16%) |

The basic information obtained with each EST–SSR was shown in Table 1. The number of alleles per locus ranged from 2 to13, with an average of 5.9 alleles per locus. GenBank (BLAST) searches indicated that 6 of 16 SSR–ESTs matched to genes of known functions at E values less than $10^{-4}$, whereas the other 10 had no significant matches to known genes (Table 1).

### Transferability of EST–SSR Primer Pairs

Transferability of the 16 EST–SSRs was examined by testing for amplification of the SSR loci in 2 other *Crassostrea* species: *C. plicatula* and *C. ariakensis*. The basic information obtained with each EST–SSR was shown in Table 3. All the loci could give successful amplification in both *C. plicatula* and *C. ariakensis*, except for locus *CGE019* that did not give successful amplifications in both *C. plicatula* and *C. ariakensis*, and locus *CGE022* that could not be amplified in *C. plicatula*. Significant deficits of heterozygotes and departures from HWE were observed in 14 single-locus exact tests after Bonferroni correction in the both species (Table 3).

### Segregation Analysis

The 16 new loci developed in this study and 4 loci (*CGE005*, *CGE006*, *CGE007*, and *CGE009*) developed in our previous study were analyzed in 3 *C. gigas* families. Genotypic frequencies in parents and offspring in each family at each of 20 loci were shown in Table 4. Of the 60 genotypic ratios observed (20 loci × 3 families), 24 came from crosses between homozygous parents and thus resulted in offspring identical to the parents or with the expected heterozygote genotype. Thirty genotypic ratios were in accordance with Mendelian expectations ($P > 0.01$). At *CGE022*, genotypes of offspring could not be determined in families PG2 because the sire could not be amplified. Five genotypic ratios were not in agreement with Mendelian segregation (Table 4). Four of the five genotypic ratios (*CGE020* in family PG2, *CGE022* in family PG3, and *CGE012* in families PG1 and PG2) confirmed Mendelian expectations when accounting for the presence of null alleles (Table 4). Locus *CGE025* in family PG1, which exhibited distorted segregation ratio, was heterozygote deficient for the genotype A/B.

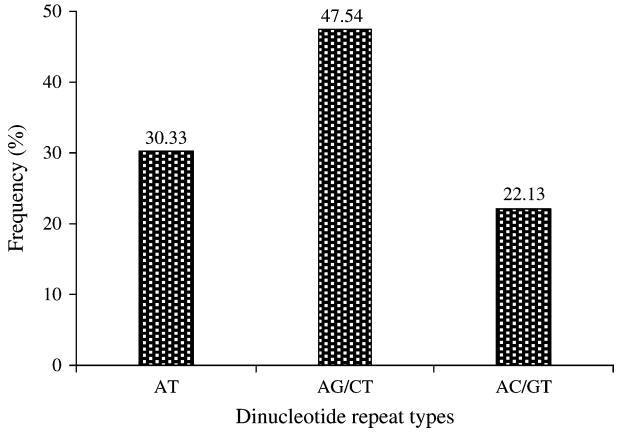


**Figure 1.** Distribution of dinucleotide microsatellites.

## Discussion

### ESTs: A Rich Source of SSRs in *C. gigas*

A total of 147 microsatellite-containing EST sequences (3.63%) were detected from 4053 EST sequences in *C. gigas* EST database. The proportion of microsatellite-containing ESTs in *C. gigas* is slightly higher than in Chinese shrimp (2.2%, Wang et al. 2005); lower than in the black tiger shrimp (13.7%), catfish (11.2%), and zebrafish (6.2%) (Serapion et al. 2004; Maneeruttanarungroj et al. 2006); and similar to that in various plant species (~5%, Varshney et al. 2005).

Random sequencing within cDNA libraries leads to a high proportion of redundant ESTs. In our study, redundancy was eliminated prior to analysis in order to reduce the data set size. The frequency of SSRs in nonredundant ESTs more accurately reflects the density of SSRs in the transcribed portion of the genome. Before redundancy elimination, the average density of EST–SSRs was 1 per 12.69 kb, whereas after redundancy elimination, the average density was 1 per 8.25 kb.

In *C. gigas* ESTs, dinucleotide repeats were found to be the most abundant (87.77%). The similar phenomenon was also found in other marine species: catfish and Chinese shrimp (Serapion et al. 2004; Wang et al. 2005). However, in plant species, trinucleotide repeats are the most common (Varshney et al. 2005). Of the dimeric EST–SSRs in *C. gigas*, CG microsatellite type was not detected, which may be a rare type in *C. gigas* ESTs. The similar phenomenon was also detected in catfish ESTs (Serapion et al. 2004). Further research is needed to determine if GC type holds true for the entire *C. gigas* genome.

### EST–SSRs Development and Polymorphism

Among the 16 EST–SSRs developed in our study, 14 primer pairs amplified products of the expected length. The exceptions (*CGE019* and *CGE023*) produced products ranging from approximately 150–200 bp larger than expected and were most likely due to the presence of

**Table 3.** Cross-species amplification of 16 EST–SSRs from *Crassostrea gigas* in other *Crassostrea* species including *Crassostrea plicatula* and *Crassostrea ariakensis*

| | *C. plicatula* ($n$ = 16) | | | | *C. ariakensis* ($n$ = 16) | | | |
|---|---|---|---|---|---|---|---|---|
| Locus | $N_a$ | $H_O$ | $H_E$ | $P$ | $N_a$ | $H_O$ | $H_E$ | $P$ |
| *CGE011* | 2 | 0.000 | 0.457 | 0.0000* | 3 | 0.133 | 0.251 | 0.0320 |
| *CGE012* | 3 | 0.093 | 0.298 | 0.0000* | 3 | 0.065 | 0.197 | 0.0000* |
| *CGE013* | 2 | 0.032 | 0.493 | 0.0000* | 4 | 0.312 | 0.672 | 0.0200 |
| *CGE014* | 4 | 0.643 | 0.771 | 0.0084 | 5 | 0.800 | 0.894 | 0.5560 |
| *CGE015* | 1 | 0.000 | 0.000 | — | 1 | 0.000 | 0.000 | — |
| *CGE016* | 2 | 0.029 | 0.033 | 0.0000* | 2 | 0.000 | 0.378 | 0.0000* |
| *CGE017* | 4 | 0.278 | 0.808 | 0.0001* | 3 | 0.019 | 0.432 | 0.0000* |
| *CGE018* | 3 | 0.098 | 0.438 | 0.0000* | 3 | 0.132 | 0.498 | 0.0005* |
| *CGE019* | — | — | — | — | — | — | — | — |
| *CGE020* | 6 | 0.614 | 0.849 | 0.0050 | 5 | 0.533 | 0.841 | 0.0010 |
| *CGE021* | 3 | 0.346 | 0.669 | 0.0430 | 4 | 0.640 | 0.706 | 0.0718 |
| *CGE022* | — | — | — | — | 2 | 0.000 | 0.371 | 0.0000* |
| *CGE023* | 2 | 0.094 | 0.640 | 0.0000* | 2 | 0.101 | 0.323 | 0.0160 |
| *CGE024* | 6 | 0.600 | 0.787 | 0.2349 | 6 | 0.333 | 0.614 | 0.0110 |
| *CGE025* | 3 | 0.291 | 0.523 | 0.0510 | 4 | 0.438 | 0.611 | 0.1940 |
| *CGE026* | 2 | 0.067 | 0.287 | 0.0000* | 2 | 0.000 | 0.337 | 0.0000* |

* Statistically significant after Bonferroni correction ($P < 0.01/15$).

introns and insertions–deletions, which are not present in EST sequences.

Because of greater EST sequences conservation in transcribed regions, EST–SSRs are usually predicted to be less polymorphic compared with genomic SSRs, and several reports about plants have revealed that the level of EST–SSR polymorphism was lower than that of genomic SSRs (Cho et al. 2000; Gupta et al. 2003). However, some recent studies reported high levels of polymorphism with EST–SRRs (Eujayl et al. 2004; Saha et al. 2005), with cases where EST–SSR markers were associated with equivalent or even higher levels of polymorphism than genomic SSR (e.g., Liewlaksaneeyanawin et al. 2004; Varshney et al. 2005). Our results also supported these observations. The number of alleles per locus developed in this study (16 loci) and our previous study (10 loci) ranged from 3 to18, with an average of 7.55 alleles per locus. The range of number of alleles and average number of allele per locus are larger than that reported by Li et al. (2003), with 79 genomic microsatellites analyzed on *C. gigas* (range of number of alleles: 2–10; number of alleles per locus: 5.7), similar to that reported by Sekino et al. (2003) (2–20; 10.6).

## Interspecific Transferability

The high interspecific transferability is an important feature of EST–SSRs due to the virtue of the sequence conservation of transcribed regions of ESTs. Not surprisingly, high interspecific transferability across *Crassostrea* species was detected in our study. Twenty-three of the 26 (88.5%) EST–SSR primer pairs developed for *C. gigas* produced strong amplification products in both *C. plicatula* and *C. ariakensis*. This interspecific transferability of EST–SSRs in *C. gigas* is obviously higher than that of genomic SSRs. Hedgecock et al. (2004) tested 86 genomic SSRs developed for *C. gigas* in cross-species amplification, and only 31 (36.0%) were likely useful for *C. ariakensis*. Sekino et al. (2003) analyzed the interspecific transferability of 8 genomic SSRs of *C. gigas* and revealed 4 (50%) loci in *C. ariakensis*. The EST–SSRs developed from *C. gigas* can be easily transferred to *C. plicatula* and *C. ariakensis* for which very little molecular information is available. They constitute a set of conserved orthologous markers, which would be ideal for assessing genetic diversity in related species as well as for cross-referencing transcribed sequences in comparative genomics studies. Significant deficits of heterozygotes and deviation from HWE were detected for 9 loci in *C. plicatula* and *C. ariakensis*, suggesting the occurrence of null alleles in the related species.

## Mode of Inheritance and Null Alleles

In this study, 14 of the 20 tested EST–SSRs codominantly segregated as expected in a Mendelian fashion in all families. Null alleles in SSRs appear to be common in bivalves. In fact, null alleles have been evidenced at the *C. gigas* microsatellites from controlled crosses (McGoldrick et al. 2000; Launey and Hedgecock 2001; Li et al. 2003). In our study, null alleles were also detected. Of the 20 loci, null alleles were observed at 3 loci (*CGE020*, *CGE022*, and *CGE012*) (15.0%). Overall, 5 of the 240 parental alleles (20 loci × 6 parents × 2) were null alleles (2.1%). By contrast with genomic SSR markers, the frequency of the *C. gigas* SSRs containing null alleles has been reported to be 47.4% (9 of 19), 51.9% (41 of 79), and 66.7% (16 of 24) (McGoldrick et al. 2000; Launey and Hedgecock 2001; Li et al. 2003). As McGoldrick et al. (2000) surveyed, of 172 parental alleles, 30 (17%) were null alleles. The frequency of troublesome null alleles in *C. gigas* EST–SSRs was lower than that of the genomic SSR markers, presumably owing to greater sequence conservation in EST–SSR markers that are adjacent to coding genes. The results obtained in this study demonstrated that the SSRs developed from EST in the Pacific oyster might be uniquely useful in population genetic

**Table 4.** Segregation analysis of microsatellite alleles in 3 *Crassostrea gigas* families

| Locus | Family | Dam | Sire | Genotypes of progeny | Expected ratio | Observed ratio | *P* value |
|---|---|---|---|---|---|---|---|
| *CGE005* | PG1 | A/B | A/A | A/A:B/A | 1:1 | 20:30 | 0.156 |
| | PG2 | A/C | A/D | A/A:A/C:D/A:D/C | 1:1:1:1 | 11:14:12:14 | 0.911 |
| | PG3 | A/A | A/D | A/A:A/D | 1:1 | 29:21 | 0.257 |
| *CGE006* | PG1 | A/A | A/A | A/A | 1 | 51 | — |
| | PG2 | A/B | A/B | A/A:A/B:B/B | 1:2:1 | 15:24:12 | 0.775 |
| | PG3 | B/B | A/C | A/B:C/B | 1:1 | 27:23 | 0.571 |
| *CGE007* | PG1 | A/C | B/B | A/B:C/B | 1:1 | 34:19 | 0.038 |
| | PG2 | C/D | B/C | C/B:C/C:B/D:C/D | 1:1:1:1 | 12:9:22:8 | 0.032 |
| | PG3 | A/A | B/B | A/B | 1 | 49 | — |
| *CGE009* | PG1 | B/C | A/B | A/C:A/B:B/B:B/C | 1:1:1:1 | 18:8:13:11 | 0.241 |
| | PG2 | D/C | D/E | D/D:C/D:D/E:C/E | 1:1:1:1 | 10:12:11:16 | 0.654 |
| | PG3 | F/F | B/G | F/B:F/G | 1:1 | 22:28 | 0.396 |
| *CGE011* | PG1 | A/A | A/A | A/A | 1 | 51 | — |
| | PG2 | A/A | A/A | A/A | 1 | 49 | — |
| | PG3 | A/A | A/A | A/A | 1 | 49 | — |
| *CGE012* | PG1 | A/X | A/B | (A/A+A/X):A/B:B/X | 2:1:1 | 24:8:18 | 0.134 |
| | PG2 | A/B | B/X | A/B:(B/X+B/B):A/X | 1:2:1 | 11:19:20 | 0.062 |
| | PG3 | A/B | A/B | A/A:A/B:B/B | 1:2:1 | 18:19:12 | 0.157 |
| *CGE013* | PG1 | A/A | A/A | A/A | 1 | 49 | — |
| | PG2 | B/B | B/C | B/B:B/C | 1:1 | 29:21 | 0.257 |
| | PG3 | B/C | B/B | B/B:B/C | 1:1 | 25:29 | 0.586 |
| *CGE014* | PG1 | B/B | A/A | A/B | 1 | 50 | — |
| | PG2 | A/B | A/B | A/A:A/B:B/B | 1:2:1 | 18:23:11 | 0.301 |
| | PG3 | A/B | A/A | A/A:A/B | 1:1 | 29:21 | 0.257 |
| *CGE015* | PG1 | A/A | A/B | A/A:A/B | 1:1 | 26:24 | 0.777 |
| | PG2 | A/A | A/A | A/A | 1 | 51 | — |
| | PG3 | A/A | A/B | A/A:A/B | 1:1 | 24:26 | 0.777 |
| *CGE016* | PG1 | A/B | B/B | A/B:B/B | 1:1 | 26:23 | 0.668 |
| | PG2 | C/C | C/C | C/C | 1 | 50 | — |
| | PG3 | A/C | C/C | A/C:C/C | 1:1 | 34:17 | 0.016 |
| *CGE017* | PG1 | A/A | A/A | A/A | 1 | 49 | — |
| | PG2 | B/C | B/D | B/B:B/C:B/D:C/D | 1:1:1:1 | 8:9:15:22 | 0.030 |
| | PG3 | D/D | D/D | D/D | 1 | 51 | — |
| *CGE018* | PG1 | B/B | A/B | A/B:B/B | 1:1 | 29:23 | 0.405 |
| | PG2 | A/A | A/A | A/A | 1 | 50 | — |
| | PG3 | B/C | A/B | A/B:A/C:B/B:B/C | 1:1:1:1 | 10:19:11:14 | 0.321 |
| *CGE019* | PG1 | A/A | A/A | A/A | 1 | 50 | — |
| | PG2 | A/A | A/A | A/A | 1 | 50 | — |
| | PG3 | A/A | A/A | A/A | 1 | 49 | — |
| *CGE020* | PG1 | A/B | A/C | A/A:A/B:B/C:A/C | 1:1:1:1 | 16:16:8:9 | 0.193 |
| | PG2 | D/X | A/D | A/D:(D/D+D/X):A/X | 1:2:1 | 9:32:10 | 0.182 |
| | PG3 | A/A | A/A | A/A | 1 | 49 | — |
| *CGE021* | PG1 | A/A | B/B | A/B | 1 | 49 | — |
| | PG2 | A/B | C/D | A/C:A/D:B/C:B/D | 1:1:1:1 | 8:14:14:13 | 0.531 |
| | PG3 | A/C | B/B | A/B:C/B | 1:1 | 27:23 | 0.571 |
| *CGE022* | PG1 | A/A | A/A | A/A | 1 | 31 | — |
| | PG2 | A/A | NA | | | | |
| | PG3 | B/X | A/B | A/B:A/X:(B/B+B/X) | 1:1:2 | 14:15:21 | 0.517 |
| *CGE023* | PG1 | A/C | A/B | A/A:AC:A/B:C/B | 1:1:1:1 | 14:17:4:16 | 0.016 |
| | PG2 | C/C | C/C | C/C | 1 | 49 | — |
| | PG3 | B/B | B/B | B/B | 1 | 50 | — |
| *CGE024* | PG1 | A/B | A/A | A/B:A/A | 1:1 | 26:23 | 0.668 |
| | PG2 | B/C | D/D | B/D:C/D | 1:1 | 27:23 | 0.571 |
| | PG3 | B/B | B/B | B/B | 1 | 49 | — |
| *CGE025* | PG1 | A/B | A/B | A/A:A/B:B/B | 1:2:1 | 14:14:22 | 0.003 |
| | PG2 | A/A | C/C | AC | 1 | 49 | — |
| | PG3 | A/A | A/A | A/A | 1 | 50 | — |
| *CGE026* | PG1 | A/C | B/D | A/B:A/D:C/B:C/D | 1:1:1:1 | 11:15:15:12 | 0.808 |
| | PG2 | C/C | A/A | A/C | 1 | 50 | — |
| | PG3 | B/C | A/A | A/B:A/C | 1:1 | 28:23 | 0.483 |

analysis, pedigree analysis, and in anchoring linkage maps across *Crassostrea*.

## Funding

National High Technology Research and Development Program (2006AA10A409); Natural Science Foundation of China (30571442).

## Acknowledgments

We would like to thank Shirasu Kunio from Nippon Suisan Kaisha, Ltd., for providing the Pacific oyster samples. We would also like to thank 2 anonymous reviewers for their valuable suggestions and comments on this paper.

## References

Ardren WR, Borer S, Thrower F, Joyce JE, Kapuscinski AR. 1999. Inheritance of 12 microsatellite loci in *Oncorhynchus mykiss*. J Hered. 90:529–536.

Arnold C, Rossetto M, McNally J, Henry RJ. 2002. The application of SSRs characterized for grape (*Vitis vinifera*) to conservation studies in Vitaceae. Am J Bot. 89:22–28.

Bentzen P, Olsen JB, McLean JE, Seamons TR, Quinn TR. 2001. Kinship analysis of Pacific salmon: insights into mating, homing, and timing of reproduction. J Hered. 92:127–136.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR. 1993. Incidence and origin of "null" alleles in the (AC)n microsatellite markers. Am J Hum Genet. 52:922–927.

Carlsson J, Reece KS. 2007. Eight PCR primers to amplify EST-linked microsatellites in the Eastern oyster, *Crassostrea virginica* genome. Mol Ecol Notes. 7:257–259.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, 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.

Dieringer D, Schlotterer C. 2003. MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes. 3:167–169.

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC, Mian MA. 2004. Medicago truncatula EST-SSRs reveal cross-species genetic markers for *Medicago* spp. Theor Appl Genet. 108:414–422.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. 2003. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. Mol Genet Genomics. 270:315–323.

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

Hubert S, Hedgecock D. 2004. Linkage maps of microsatellite DNA markers for the Pacific oyster *Crassostrea gigas*. Genetics. 168:351–362.

Jones AG, Stockwell CA, Walker D, Avise JC. 1998. The molecular basis of a microsatellite null allele from the white sands pupfish. J Hered. 89:339–342.

Launey S, Hedgecock D. 2001. High genetic load in the Pacific oyster *Crassostrea gigas*. Genetics. 159:255–265.

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

Li Q, Kijima A. 2006. Microsatellite analysis of gynogenetic families in the Pacific oyster, *Crassostrea gigas*. J Exp Mar Biol Ecol. 331:1–8.

Li Q, Park C, Kobayashi T, Kijima A. 2003. Inheritance of microsatellite DNA markers in the Pacific abalone *Haliotis discus hannai*. Mar Biotechnol. 5:331–338.

Li Q, Park C, Kijima A. 2003. Allelic transmission of microsatellites and application to kinship analysis in newly hatched Pacific abalone larvae. Fish Sci. 69:883–889.

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

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K. 2004. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet. 109:361–369.

Maneeruttanarungroj C, Pongsomboon S, Wuthisuthimethavee S, Klinbunga S, Wilson KJ, Swan J, Li Y, Whan V, Chu KH, Li CP, et al. 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.

McGoldrick DJ, Hedgecock D, English LJ, Baoprasertkul P, Ward RD. 2000. The transmission of microsatellite alleles in Australian and North American stocks of the Pacific oyster (*Crassostrea gigas*): selection and null alleles. J Shellfish Res. 19:779–788.

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, Kochko A, Hamon P. 2006. SSR mining in coffee tree EST databases: potential use of EST–SSRs as markers for the *Coffea* genus. Mol Gen Genomics. 276:436–449.

Powell W, Machray GC, Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1:215–222.

Primmer CR, Møller AP, Ellegren H. 1995. Resolving genetic relationships with microsatellite markers: a parentage testing system for the swallow *Hirundo rustica*. Mol Ecol. 4:493–498.

Rice WR. 1989. Analyzing tables of statistical tests. Evolution. 43:223–225.

Rungis D, Berube Y, Zhang J, Ralph S, Ritland CE, Ellis BE, Douglas C, Bohlmann J, Ritland K. 2004. Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. Theor Appl Genet. 109:1283–1294.

Saha MC, Mian R, Zwonitzer JC, Chekhovskiy K, Hopkins AA. 2005. An SSR- and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). Theor Appl Genet. 110:323–336.

Sekino M, Hamaguchi M, Aranishi F, Okoshi K. 2003. Development of novel microsatellite DNA markers from the Pacific oyster *Crassostrea gigas*. Mar Biotechnol. 5:227–233.

Serapion J, Kucuktas H, Feng J, Liu Z. 2004. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). Mar Biotechnol. 6:364–377.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W. 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol. 12:1339–1348.

Varshney RK, Graner A, Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23:48–54.

Vendramin E, Dettori MT, Giovinazzi J, Micali S, Quarta R, Verde I. 2007. A set of EST-SSRs isolated from peach fruit transcriptome and their transportability across *Prunus* species. Mol Ecol Notes. 7:307–310.

Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc Natl Acad Sci USA. 99:9650–9655.

Wang H, Li F, Xiang J. 2005. Polymorphic EST–SSR markers and their mode *Fenneropenaeus chinensis*. Aquaculture. 249:107–114.

Yu H, Li Q. 2007. EST-SSR markers from the Pacific oyster, *Crassostrea gigas*. Mol Ecol Notes. 7:860–862.

Received June 5, 2007
Accepted November 24, 2007

Corresponding Editor: Jill Pecon-Slattery