Korean J. Malacol. 30(3): 197-210 2014
http://dx.doi.org/10.9710/kjm.2014.30.3.197

Transferability of Cupped Oyster EST (Expressed Sequence Tag)-Derived SNP (Single Nucleotide Polymorphism) Markers to Related *Crassostrea* and *Ostrea* Species

Woo-Jin Kim[1], Hyungtaek Jung[2], Eun-Ha Shin[1], and Ilseon Baek[3]

[1]*Biotechnology Research Division, National Fisheries Research and Development Institute, Busan 619-705, Korea*
[2]*Science and Engineering Faculty, Queensland University of Technology, Brisbane 4001, Australia*
[3]*Division of Marine Technology, Chonnam National University, Yeosu 550-749, Korea*

ABSTRACT

Single nucleotide polymorphisms (SNPs) are widely acknowledged as the marker of choice for many genetic and genomic applications because they show co-dominant inheritance, are highly abundant across genomes and are suitable for high-throughput genotyping. Here we evaluated the applicability of SNP markers developed from *Crassostrea gigas* and *C. virginica* expressed sequence tags (ESTs) in closely related *Crassostrea* and *Ostrea* species. A total of 213 putative interspecific level SNPs were identified from re-sequencing data in six amplicons, yielding on average of one interspecific level SNP per seven bp. High polymorphism levels were observed and the high success rate of transferability show that genic EST-derived SNP markers provide an efficient method for rapid marker development and SNP discovery in closely related oyster species. The six EST-SNP markers identified here will provide useful molecular tools for addressing questions in molecular ecology and evolution studies including for stock analysis (pedigree monitoring) in related oyster taxa.

Key words: SNPs, EST, *Crassostrea*, *Ostrea*, Transferability, Type I markers

INTRODUCTION

A number of oyster species are of global economic significance in wild fisheries and in aquaculture (Jonas *et al*., 2007), apart from providing important components of many aquatic ecosystems. While they provide models for genetic and environmental studies (Jenny *et al*., 2007), more importantly oyster species have shown the highest production rates of any farmed aquatic animal since the mid 90's, and provided approximately one-tenth of global aquaculture production in 2011 (FAO, 2013). Growth in demand for cultured edible oysters has seen genetic improvement programs for some species initiated in several countries (Langdon *et al*., 2003) but most stock improvement programs are still in the early stage of development and lack the large numbers of genetic markers required.

For several decades, mitochondrial DNA (mtDNA) and microsatellite genetic markers have been the most common types used for molecular ecological studies (Beheregaray, 2008; Kim *et al*., 2014). In recent years, however, with the rapid increase in the availability of genomic information on oysters from expressed sequence tag (EST) collections, EST-derived simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs), a genomic approach to molecular marker development has become an

Received: August 16, 2014; Revised: September 19, 2014; Accepted: September 23, 2014
Corresponding author : Woo-Jin Kim
Tel: +82 (51) 720-2465 e-mail: wj2464@korea.kr
Corresponding author : Hyungtaek Jung
Tel: +61 (7) 3138-4111 e-mail: h7.jung@qut.edu.au
1225-3480/24530

attractive alternative to complement-existing markers because then can reduce time and cost compared with development of traditional genomic libraries. To date, a large number of EST-derived markers in oysters have been developed successfully from public sequence databases (Quilang *et al*., 2007; Sauvage *et al*., 2007; Zhang and Guo 2010; Kim *et al*., 2011) and in particular EST-SSRs have been used successfully for cross-species amplification in some *Crassostrea* species (Li *et al*., 2009). This approach reduces potential problems including potential for substantial null (non-amplifying) alleles, time and cost of development (Li and Guo, 2004). Recently the availability of a full genome sequence for an oyster species is likely to accelerate this development utilizing EST data (Zhang *et al*., 2012). SSRs and SNPs markers have been used widely for genome mapping, linkage and association studies, parentage analysis, and other applications in molecular ecological studies because they provide powerful co-dominant markers. While SSRs are very popular markers because of their high level of variation, SNPs are now considered to the marker of choice because of their relative abundance across the genome and their suitability for high-throughput genotyping. An attractive feature of EST-derived markers is their direct association with transcripts that may be identified and associated with cellular and molecular functions (Li *et al*., 2009). Some ESTs may also include evolutionarily conserved motifs, allowing more robust cross-species PCR amplification than is possible with traditional genomic library markers (Bouck and Vision, 2007). While a few studies have utilized oyster EST databases for SNP development (Quilang *et al*., 2007; Sauvage *et al*., 2007; Zhang and Guo, 2010), only one study to date has considered testing cross-species amplification of EST-derived SNP markers in closely related oyster species (Kim *et al*., 2011).

SNP-based markers are typically mined from whole-genome sequences or from EST data sets developed in genetically diverse individuals (Barbazuk *et al*., 2007). EST analysis is not only the most efficient approach for gene discovery, but also provides an effective approach for identification or discovery of polymorphic DNA markers including SNPs (Morin *et al*., 2004; Amaral *et al*., 2009). In non-model organisms, however, where availability of nucleotide sequence is often limited or absent, SNPs have to be identified via laboratory screening (Sauvage *et al*., 2007; Kim *et al*., 2011). Given the growth in public EST databases for a diverse array of taxa including oysters (Jenny *et al*., 2007; Fleury *et al*., 2009), SNP marker development has become a subject of great interest as these markers provide considerable advantages over microsatellite markers (Kim *et al*., 2011). This is because they are distributed abundantly across the genome, are often functionally relevant, are less mutable than SSRs, and in general are more robust in the laboratory and for data interpretation (Morin and McCarthy, 2007; Kim *et al*., 2011). Thus, the aim of the current study was to evaluate suitability of EST-SNP markers for cross-species amplification in closely related *Crassostrea* and *Ostrea* species. In addition, we explore the potential of Type I SNP loci for addressing species-level systematic relationships in oysters.

MATERIAL AND METHODS

1. Oyster samples

A total of 52 oysters comprising 44 *Crassostrea* and eight *Ostrea* species were used for primer testing and DNA sequencing (Table 1). *C. gigas* (*Cgig*, n = 2), *C. ariakensis* (*Cari*, n = 8), *C. nippona* (*Cnip*, n = 8), and *O. denselamellosa* (*Oden*, n = 8) from Korea were collected by the authors in 2003, *C. angulata* (*Cangu*, n = 2) from Portugal were provided by Prof. Diarmaid Ó Foighil at University of Michigan, *C. iredalei* (*Cired*, n = 8) from Vietnam and *C. hongkongensis* (*Chong*, n = 8) from China were provided by Prof. Kimberly Reece at Virginia Institute of Marine Science, and *C. sikamea* (*Csik*, n = 4) and *C. virginica* (*Cvir*, n = 4) from USA were provided by Prof. Dennis Hedgecock at University of Southern California. All tissue samples were stored in 95% ethanol prior to DNA extraction.

2. DNA extraction and Sequencing

DNA was extracted from adductor muscle or gill tissue using a DNeasy extraction kit (Qiagen),

Table 1. Sample collection sites and sample sizes

Species	Site	Abbreviation	Sample size
C. gigas	Tongyoung City (Korea)	*Cgig*	2
C. angulata	Rio Mira (Portugal)	*Cangu*	2
C. ariakensis	Seomjin River (Korea)	*Cari*	8
C. hongkongensis	Yamen River (China)	*Chong*	8
C. iredalei	Nah Trang (Vietnam)	*Cired*	8
C. nippona	Donghae City (Korea)	*Cnip*	8
C. sikamea	Washington State (USA)	*Csika*	4
C. virginica	Wachapreague, VA (USA)	*Cvir*	4
O. denselamellosa	Donghae City (Korea)	*Oden*	8

[a]*Csik* indicate the *C. sikamea* culture line from Washington State. Other oysters are from wild populations.

following the extraction protocol for animal tissues. EST sequences were obtained from *C. gigas* (www.ifremer.fr/GigasBase) and *C. virginica* website (www.marinegenomic.org). A total of six amplicons (EST-SNP markers), selected from previous studies (Jung *et al*., 2006; Kim *et al*., 2011), were re-sequenced to confirm that the target loci had been amplified successfully, and to survey patterns of interspecific polymorphisms. However, Cyclophilin F (Cyc F), Muted protein (MP), Nucleoside disphosphate kinase (NDK) and Ribosomal protein large 13 (RPL13) loci could not be amplified in *C. iredalei*, *C. virginica* and *O. denselamellosa* because of potential low homology in primer binding sites. In addition, an unknown gene was detected in the RPL13 trial for *C. nippona* (Table 2). Therefore, sequences from B cell translocation gene (BTG) and Elongation factor 1α (EF-1α) loci that amplified successfully across all species were used to explore systematic relationships among several *Crassostrea* and *Ostrea* species.

Amplification was tested in a panel of 52 individuals sampled from several oyster species to determine the potential transferability of the markers across species and genera (Tables 1 and 2). PCRs were performed in a 25 μL volume containing 25 ng genomic DNA, 10 mM Tris-HCl (pH, 8.0), 0.1% Triton X-100, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 5 pmol of each primer, and 0.5 U Ex-Taq DNA polymerase (Takara, Kyoto, Japan). The amplication was carried out using a PTC-220 thermocycler (MJ Research, Watertown, MA, USA) programmed for 5 min at 95°C followed by 35 cycles of 30 s at 95°C, 30 s at 52-60°C, and 30 s at 72°C, and a final extension of 10 min at 72°C. Amplified products were purified using AMPure beads (Agencourt Bioscience, Beverly, MA, USA) according to the manufacturer's protocol for sequencing. Cycle sequencing of each sample was conducted in one or both directions with the primers used for amplification and the ABI BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) in the same PTC-220 thermocycler under the following conditions: 50 cycles of 94°C for 15 s, 48°C for 20 s, and 60°C for 4 min. Sequences were analyzed on an ABI 3100xl automated sequencer (Applied Biosystems). Lasergene v.9 (DNASTAR Inc., Madison, WI, USA) was used to visualize sequences. A multiple sequence alignment of individual consensus sequences including the sequence used to design the primers was constructed to detect SNPs and insertions/deletions (indels). SNPs were scored as heterozygous where two peaks differing by < 25% in intensity were present in the sequence trace and were repeatable in two individuals and could be repeated in a single individual (forward and reverse direction sequences). Haplotypes were reconstructed from diploid sequences using the PHASE algorithm implemented in Dnasp v.5.10 (Librado and Rozas, 2009). Transition (Ts) to transversion (Tv) ratio was calculated in MEGA 5 (Tamura *et al*., 2011) to identify possible genes that could potentially be targets of historical selection (Morton *et al*., 2006).

3. Phylogenetic analyses

To address potential systematic relationships among

Table 2. Summary of candidate SNP markers, derived from *C. gigas* EST database, amplification across oysters

Locus†	Reference	Tested Species‡	Forward/reverse primer (5'-3')	T (°C) range	Expected size range (bp)* Intron + Exon	Expected size range (bp)* 3'UTR	Putative homolog amplified?
BTG§	BQ426863	A to I	F: §CTC**D**AGCCAATTCATGCAAAGGAC R: §ATATGTACAGATGA**Y**TGGCAGC**R**A	56-60	62/65 (H)[a]	27-41	All yes
Cyc F	AY441092	A to I	F: GGCATGAATGTCGTCAAAGC R: ACAAATGGCAGGATTGATACAAC	58-60	73	121-132	All yes but PCR fail for H, I
EF-1α§	CB617441	A to I	F: §GTCCTTGATTGCCA**Y**ACTGCTC R: §GGTTCTTTCCGACGTA**Y**TTCTT	57-60	287/290 (E, I)[a]	145-154	All yes
MP	BQ426621	A to I	F: AGCAGGAATTCTGGCTAGAC R: TAATCAACAAATAAGTTTATTGCAC	57-59	42-53 (G)[b]	123-183	All yes but no amplification for H and I
NDK	BQ426836	A to I	F: TCATGGTAGTGATTCCGTAG R: TATTTGACTGTCCACAACACTGATG	53-57	89	53-57	All yes but no amplification for H and I
RPL13	BQ426257	A to I	F: AGTGCTTTCACTGCTCTCCGTCAAG R: CGGCAAAAACTGGATATGATCTG	52-58	86 (E)[c]	194-209	All yes but no amplifycation for F, H and I. Other gene detected in E

†Locus name: B cell translocation gene (BTG), cyclophilin F (Cyc F), elongation factor 1α (EF-1α), muted protein (MP), nucleoside disphosphate kinase (NDK), and ribosomal protein large 13 (RPL13); T: annealing temperature; ‡Tested species name: *C. gigas* (A), *C. ariakensis* (B), *C. angulata* (C), *C. hongkongensis* (D), *C. nippona* (E), *C. iredalei* (F), *C. sikamea* (G), *C. virginica* (H) and *O. denselamellosa* (I).

§Used degenerate bases from *C. gigas* and *C. virginica* EST database (Bold and underline characters; D is for A/G/T but not C, R is for A/G, Y is for T/C).

*Actual expected size ranges within each region without primer sites.

[a]One amino acid deletion occurred compared to other oysters.

[b]Missing sequence from base and peak quality check not deletion or insertion.

[c]Based on coding sequence translation, different gene was detected compared to other oysters.

oyster taxa, basal oyster sequences (placed in the center of the unrooted tree) from the CDS regions of BTG and EF-1α loci were selected and two gene fragments were concatenated after application of the homogeneity test in PAUP 4.0b.10 . Bayesian phylogenetic (Bayesian Index [BI]) analyses were performed using the Mr. Bayes 3.1.2 package (Ronquist and Huelsenbeck, 2003). A GTR + G model was implemented and this model was run for 10^6 generations. A Metropolis-coupled Markov Chain Monte Carlo (MCMC) process was undertaken for each data partition running simultaneously with a cold chain and three incrementally heated chains. Maximum likelihood (ML) tree building was also performed with RAxML 7.0.3 (Stamatakis, 2006) using the GTRMIX model. Non-parametric bootstrapping with 1000 pseudoreplicates was used to estimate confidence intervals for tree topologies. Gaps were treated as missing data.

RESULTS

All six primer pairs yielded detectable PCR products in *C. gigas*, *C. angulata*, *C. ariakensis*, and *C. hongkongensis*, but amplified less successfully in the remaining oyster taxa screened (Table 2). Despite high polymorphism in the 3'UTR across species, results of cross-species amplification in related *Crassostrea* oysters showed a high success rate (100%) in *C. angulata*, *C. ariakensis*, *C. hongkongensis* and *C. sikamea* for the six target loci. 98 putative interspecific level SNPs (28 amino acid replacements), 6 indels (*C. virginica* in BTG and *C. nippona* and *O. denselamellosa* in EF-1α) and 11 missing sequences (*C. sikamea* in MP) were identified from a total of 656 bp of coding sequence in six sequenced amplicons for cross-species amplification. Average CDS nucleotide diversity (π) was 0.0499 and 0.0505 (excluding gaps and missing sequences) overall. Average frequency of interspecific level SNPs was estimated to be one SNP

every seven bp. Six amplicons contained 3' UTR, yielding 260 SNPs (125, without MP amplicons because of missing sequences presence), 100 indels (64, without MP amplicons) and 43 missing sequences (0, without MP ampliconsin *C. iredalei* and *C. sikamea*) were observed in a total of 776 bp (593 without MP amplicons). Average CDS nucleotide diversity (π) was 0.0843 and 0.0941 (excluding gaps and missing sequences) overall. The average frequency of interspecific level SNPs was estimated to be one SNP every five bp in a 593 bp of fragment (without MP sequences). Overall average density of interspecific level SNPs was estimated to be one SNP every seven bp from a total of 213 putative SNPs revealed in the study. Regardless of whether comparisons were at the intra- or interspecific level studies, higher polymorphisms (SNPs and indels) were detected in 3' UTR regions (introns or non-coding regions in other studies) than in CDS.

PCR success rate was higher for Asian *Crassostrea* than for Eastern congeners (*C. virginica*) or *O. denselamellosa*, consistent with the phylogenetic relationship inferred from the SNP locus sequence data (Fig. 1). The homogeneity (incongruence) test showed that the two data sets were congruent ($P > 0.05$), and simultaneous analysis of the total sequence data partitioned by ML and BI converged into a single tree (Fig. 1). The monophyletic relationships of the cupped oyster clades based on data from the two loci confirmed Portuguese oyster to be a sister species to Pacific oyster. Based on these data, the Eastern and European oysters both constitute outgroups to the Asian species examined here. The tree indicated that C. iredealei was nested with high bootstrap values (100 [ML] and 96 [BI]) between Asian and Eastern oysters as a subclade with other Asian groups. This is an interesting result and will require further investigation. The current study did not, however, fully recover a clear relationship among *C. ariakensis*, *C. hongkongensis*, *C. nippona* and *C. sikamea*.

1) B cell translocation gene (BTG)

BTG is a nuclear gene that has been implicated in cell differentiation processes including proliferation of microglia and their sensitivity to apoptogenic agents (Lee *et al*., 2003). BTG primers used here amplified a BTG-like sequence in all oyster species examined. While primers amplified a 109 bp fragment (CDS and 3' UTR) encoding 20 amino acids in the Eastern oyster (including a deletion), 21 amino acids were amplified in the other cupped oysters. Thirteen base-pair polymorphisms were evident (seven amino acids replacement) in CDS and 25 in 3' UTR (including indels) among the nine oyster taxa compared here (Fig. 2). The Ts/Tv ratio was 1.2 in CDS and 1.1 in 3' UTR, respectively. Average nucleotide diversity (π) in CDS was 0.0780 (0.0753 excluding gaps and missing sequences) and 0.1396 (0.1646 excluding gaps and missing sequences) in 3' UTR, respectively.

2) Cyclophilin F (Cyc F)

Cyclophilins are a family of cytosolic ubiquitous proteins (Galat, 1999) present in all subcellular compartments that play a pivotal role in protein folding via enzymatic catalysis of the peptidyl-prolyl *cis-trans* isomerisation reaction (Takahashi *et al*., 1989). Cyc F primers used here amplified a Cyc F product in all sampled oyster species except for *C. virginica* and *O. denselamellosa*. A 205 bp fragment (CDS and 3' UTR) encoded 24 amino acids in the study species. Thirteen interspecific sites (two amino acids replacement) in CDS and 25 in 3' UTR (including indels) were variable among the seven oyster taxa compared here (Fig. 2). The Ts/Tv ratio was 0.6 for both CDS and 3' UTR. SNP density varied among genes, and this may be due in part to strong historical selection. Ts/Tv ratio can help to identify such genes affected by selection (Morton *et al*., 2006). Average nucleotide diversity (π) in CDS was 0.0582 (no gaps and missing sequences) and 0.0386 (0.0403 excluding gaps and missing sequences) in 3' UTR, respectively.

3) Elongation factor 1α (EF-1α)

EF-1α is a core element of the translation apparatus and is a member of the GTPase protein family so this gene has been used widely as a phylogenetic marker in eukaryotes to resolve phylogenetic patterns as it is highly conserved

(Baldauf *et al*., 1996). The EF-1α primers amplified an EF-1α product in all cupped oyster species here. A 444 bp fragment (CDS and 3' UTR) encoded 95 amino acids in *C. nippona* and *O. denselamellosa* (a deletion occurred) but 96 amino acids were present in other cupped oysters. 63 interspecific sites (14 amino acids replacement) in CDS and 41 in 3' UTR (including indels) were variable among the nine oyster taxa compared here (Fig. 2). The Ts/Tv ratio was 1.0 in CDS and 0.7 in 3' UTR, respectively. Average nucleotide diversity (π) in CDS was 0.0649 (0.0663 excluding gaps and missing sequences) and 0.0610 (0.0616 excluding gaps and missing sequences) in 3' UTR.

4) Muted protein (MP)

The MP gene encodes a component of BLOC-1, which influence the biogenesis of lysosome-related organelle complex. Components of this complex are involved in biogenesis of organelles including melanosomes and platelet-dense granules (Li *et al*., 2004). Primers used here amplified BLOC-1 like product in all cupped oyster species except for *C. virginica* and *O. denselamellosa*. MP primers amplified a 236 bp fragment (CDS and 3' UTR) that encoded 14 amino acids in *C. sikamea* (missing sequence) and 17 amino acids in other cupped oyster species. Eight interspecific sites (two amino acids replacement) in CDS were variable among the seven oyster taxa compared here (Fig. 2). Polymorphic sites in 3' UTR

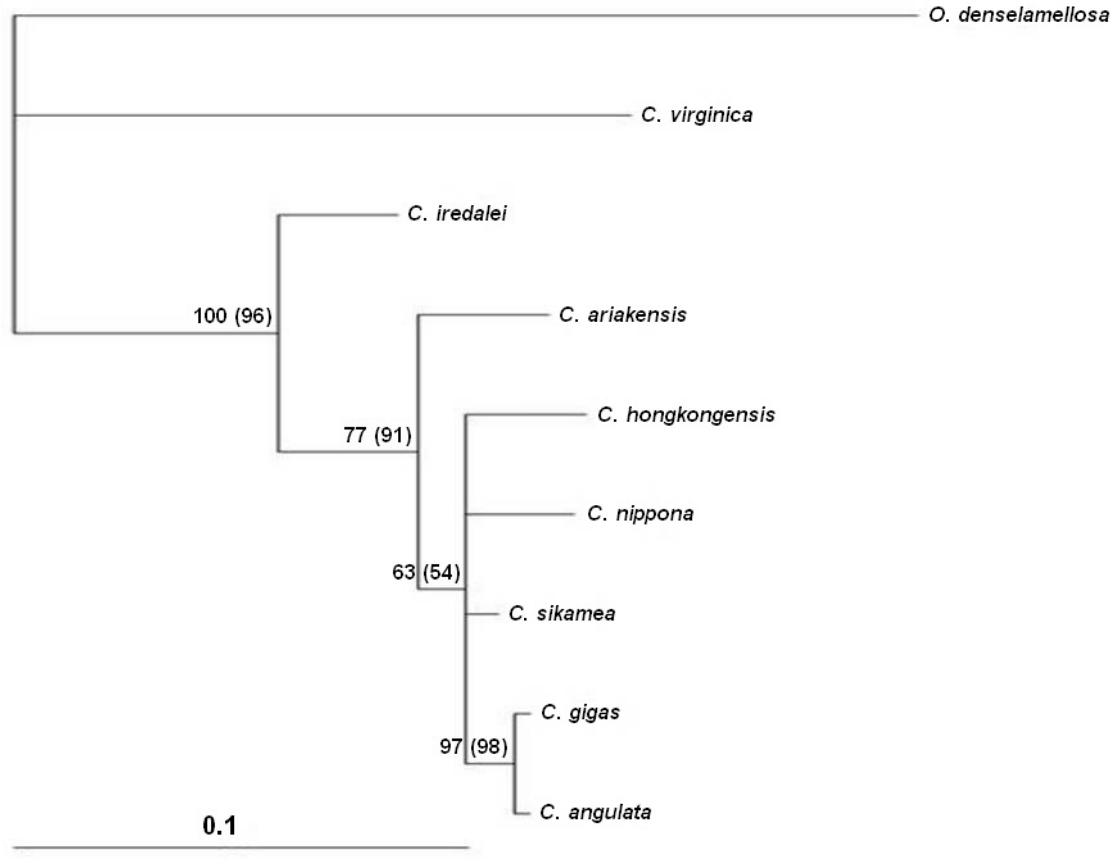
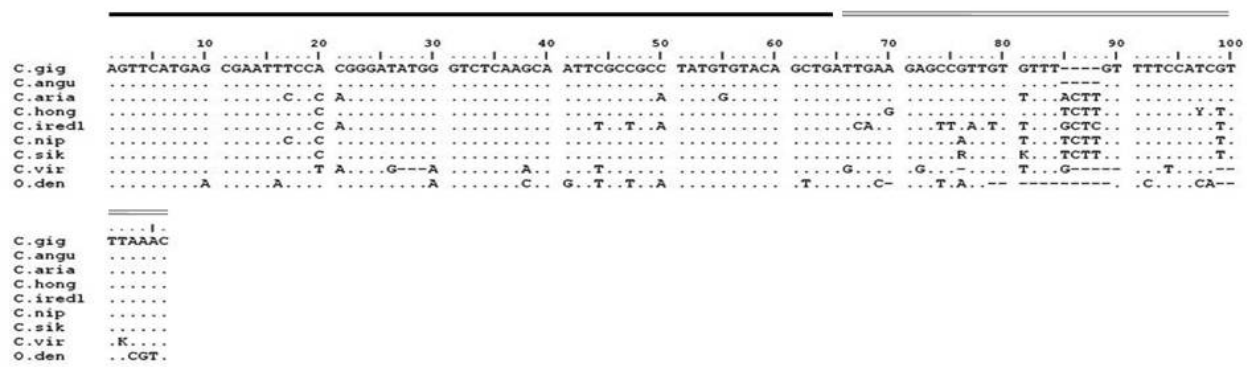
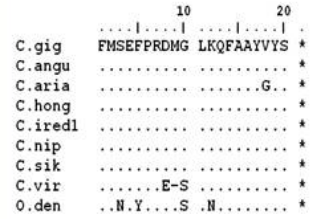


Fig. 1. Bayesian method phylogenetic tree of concatenated coding region sequence data set of BTG and EF-1α using RAxML and MrBayes programs. The numbers above the internal branches represent percent bootstrap support values based on 1000 pseudoreplicates by maximum likelihood and the posterior probability with Bayesian methods in brackets.

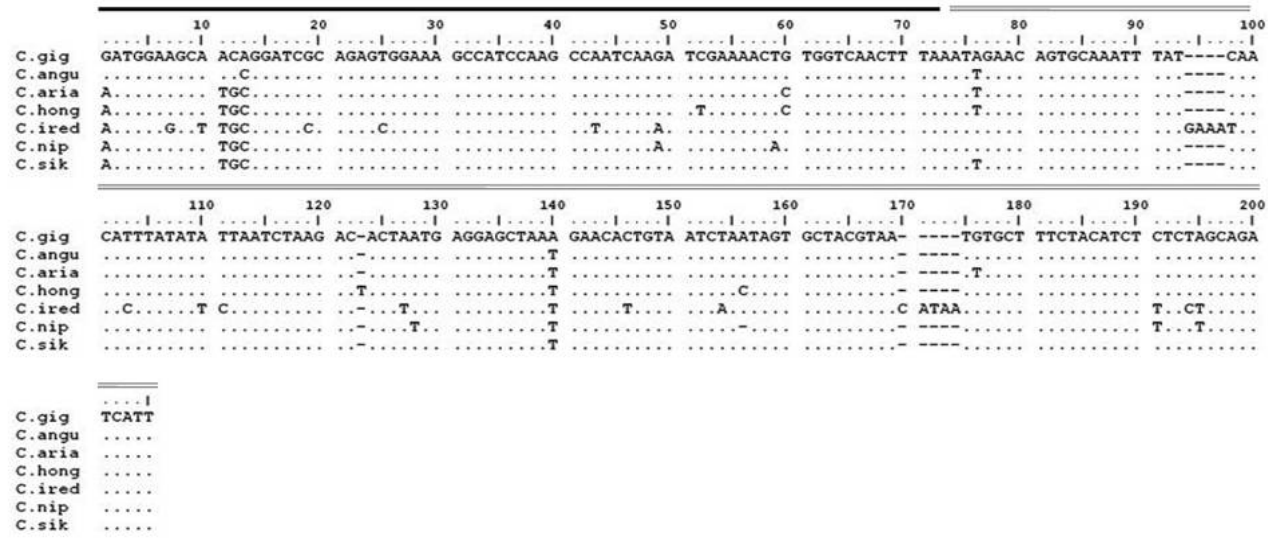
1. BTG
(A)



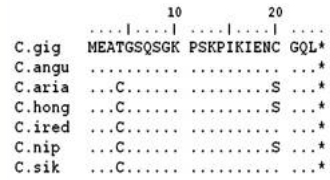
(B)



2. CycF
(A)

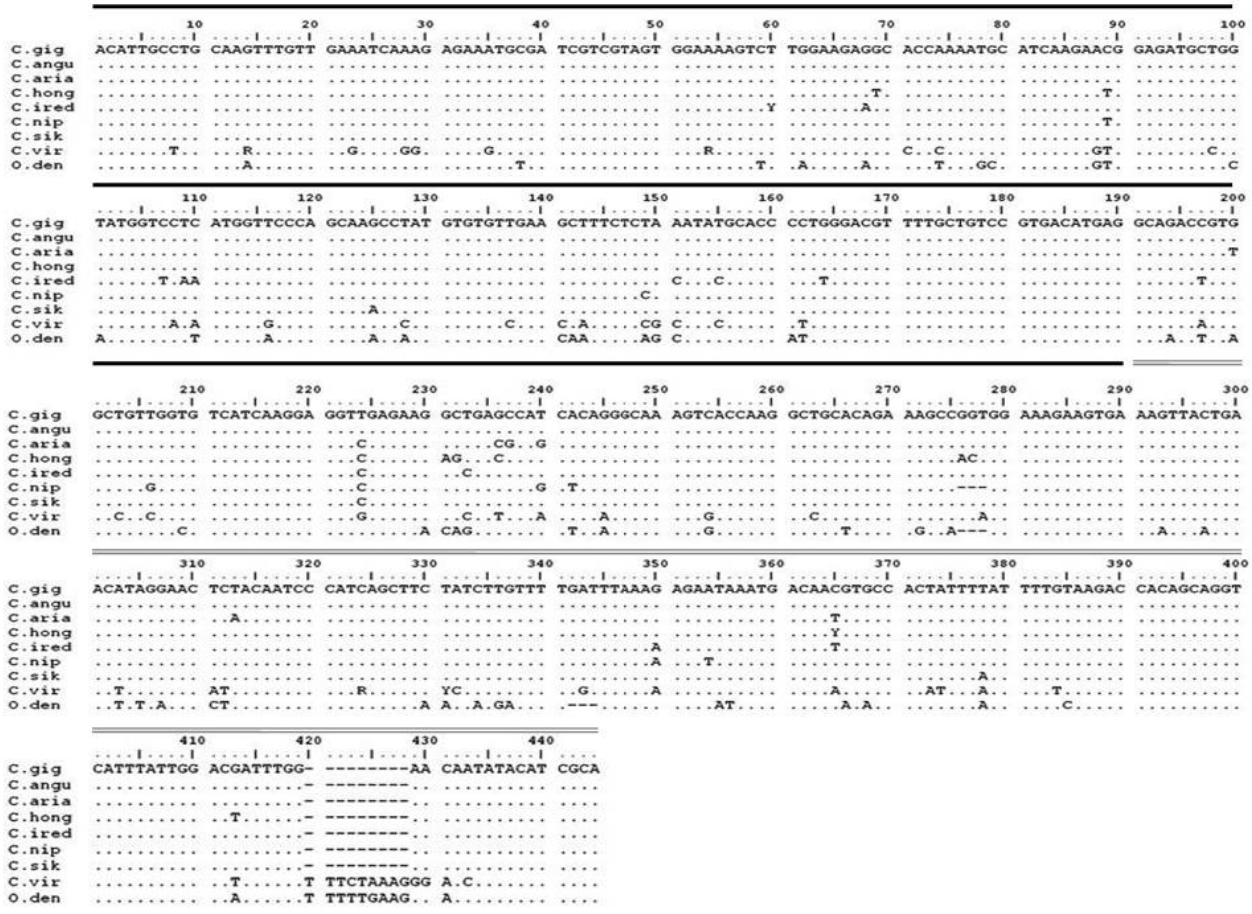


(B)

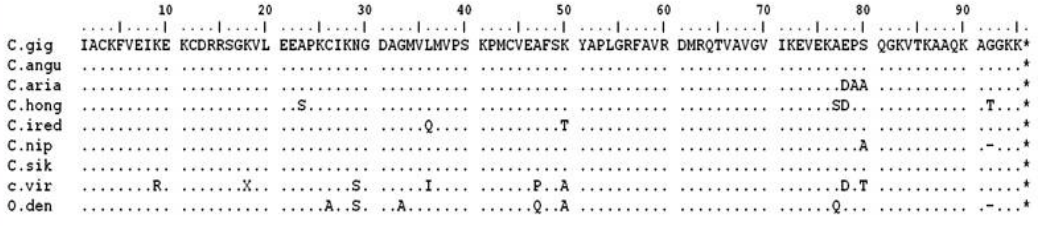


3. EF-1α

(A)

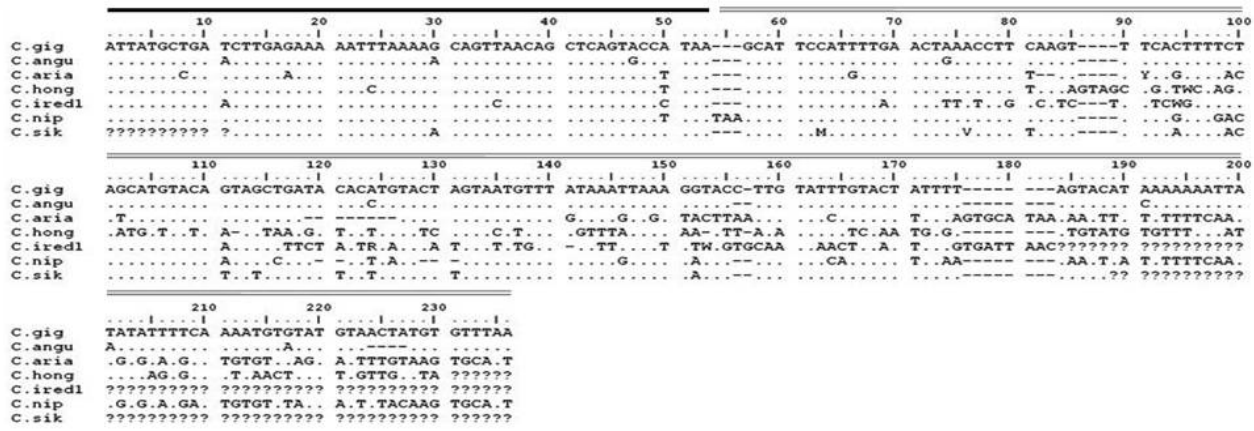

C.gig ACATTGCCTG CAAGTTTGTT GAAATCAAAG AGAAATGCGA TCGTCGTAGT GGAAAAGTCT TGGAAGAGGC ACCAAAATGC ATCAAGAACG GAGATGCTGG
C.angu
C.aria
C.hong
C.ired
C.nip
C.sik
C.vir .....T.. ...R..... ..G....GG. ....G..... ....R..... .... C..C..... .....GT. .....C..
O.den ....A..... ....T.. ....T. .A.....A.. ...T..GC. .....GT. C
C.gig TATGGTCCTC ATGGTTCCCA GCAAGCCTAT GTGTGTTGAA GCTTTCTCTA AATATGCACC CCTGGGACGT TTTGCTGTCC GTGACATGAG GCAGACCGTG
C.angu
C.aria .....T
C.hong
C.ired .....T.AA ..........C...C..... ...T..... .....T...
C.nip .....C.
C.sik ....A.....
C.vir .....A.A .....G.... .....C.. .....C... C.A....CG C...C..... .T..... .....A...
O.den A.....T .....A.... ....A..A.. .... CAA....AG C.......... AT..... .....A..T..A
C.gig GCTGTTGGTG TCATCAAGGA GGTTGAGAAG GCTGAGCCAT CACAGGGCAA AGTCACCAAG GCTGCACAGA AAGCCGGTGG AAAGAAGTGA AAGTTACTGA
C.angu
C.aria ...C..... .....CG..G
C.hong ...C..... AG...C....
C.ired ...C..... ..C.....
C.nip .....G.... ...C..... .....G .T..... .....---..
C.sik ...C.....
C.vir ..C..C.... ...G..... ..C..T...A ....A..... ...G..... ..C..... .....A..
O.den .....C. .....A CAG..... .T..A..... ...G..... .....T.... .G..A---..
C.gig ACATAGGAAC TCTACAATCC CATCAGCTTC TATCTTGTTT TGATTTAAAG AGAATAAATG ACAACGTGCC ACTATTTTAT TTTGTAAGAC CACAGCAGGT
C.angu
C.aria ..A..... ....T.....
C.hong .....Y.....
C.ired .....T.....
C.nip .....A ...T.....
C.sik .....A..
C.vir ..T..... AT..... ...R..... YC..... ..G.....A .....A..... ..AT...A.. ...T.....
O.den ..T.T.A... CT..... .....A A..A.GA... .---..... ....AT.... .....A.A.. .....A.. ....C.....
C.gig CATTTATTGG ACGATTTGG- --------AA CAATATACAT CGCA
C.angu .....- --------.. .....
C.aria .....- --------.. .....
C.hong ..T.....- --------.. .....
C.ired .....- --------.. .....
C.nip .....- --------.. .....
C.sik .....- --------.. .....
C.vir ..T.....T TTCTAAAGGG A.C..... ....
O.den ..A.....T TTTTGAAG.. A..... ....

(B)


C.gig IACKFVEIKE KCDRRSGKVL EEAPKCIKNG DAGMVLMVPS KPMCVEAFSK YAPLGRFAVR DMRQTVAVGV IKEVEKAEPS QGKVTKAAQK AGGKK*
C.angu .....*
C.aria .....DAA .....*
C.hong .....S..... .....SD.. .....T...*
C.ired .....Q.... .....T ....*
C.nip .....A ..... .-...*
C.sik .....*
c.vir .....R. .....X.. .....S. .....I.... .....P..A ....D.T ....*
O.den ....A..S. ..A..... .....Q..A ....Q... ..... .-...*

4. MP

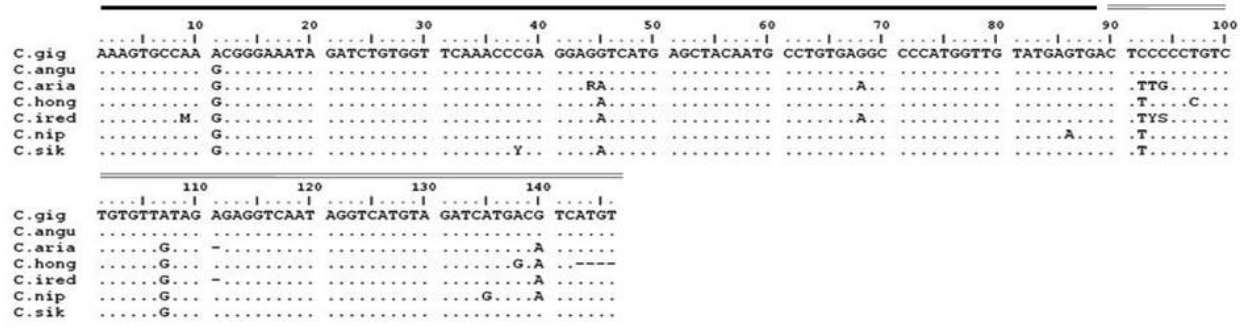
(A)



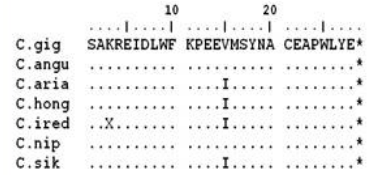
(B)

5. NDK

(A)

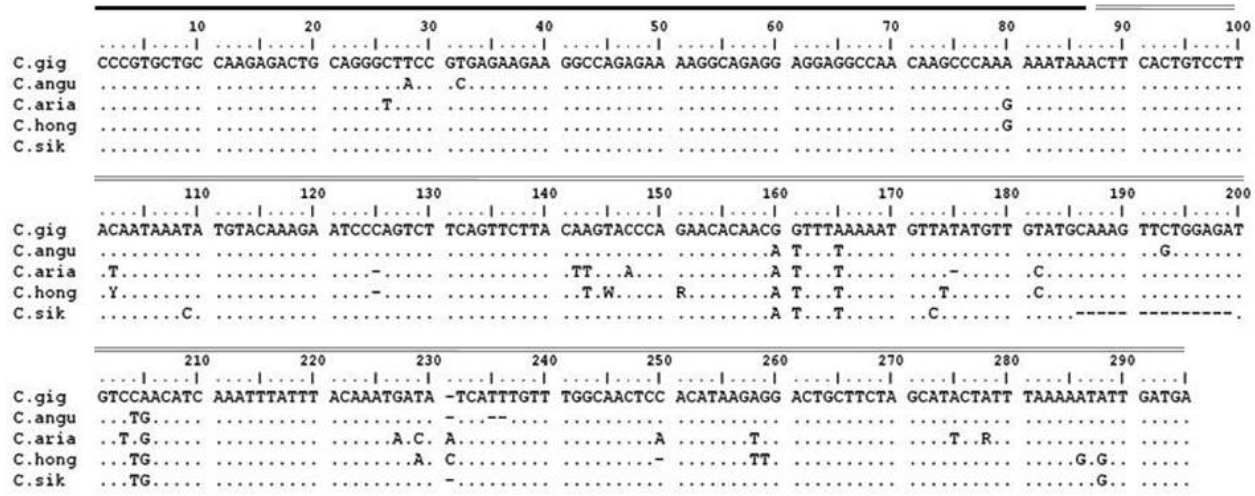


(B)



6. RPL13

(A)



(B)

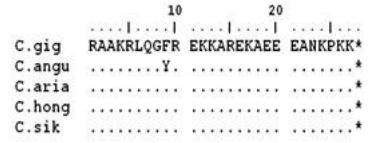


Fig. 2. Comparison of alignment of sequenced amplicons for nucleotide and amino acid translation. (**A**) Alignment of sequenced amplicons showing representative genotypes and haplotypes. Sequenced amplicons represent individuals from several accessions: *C. gigas* from Korea (C. gig), *C. angulata* from Portugal (C. angu), *C. ariakensis* from Korea (C. ari), *C. hongkongensis* from China (C. hong), *C. iredalei* from Vietnam (C. ired), *C. nippona* from Korea (C. nip), *C. sikamea* from USA (C. sik), *C. virginica* from USA (C. vir) and *O. denselamellosa* from Korea (O. den). Solid black line is coding sequencing region; gray double line is 3' untranslated region; - is insertion and deletion site; ? is missing sequence and ? is identical sequence site, respectively. (**B**) Alignment of translated amino acid from the sequence amplicons. * is stop codon in the amplicon frame; - is insertion and deletion site; ? is missing sequence and X is two amino acid translations (AAA [Lys] & GAA [Glu] in EF-1α of *C. virginica* and AAG [Lys] & GAA [Glu] in NDK of *C. iredalei*), respectively.

(including indels) were not estimated because many sequences were missing. The Ts/Tv ratio was 2.1 in CDS and 0.5 in 3' UTR, respectively. Average nucleotide diversity (π) in CDS was 0.0556 (0.0605 excluding gaps and missing sequences) and 0.1609 (0.1955 excluding gaps and missing sequences) in 3' UTR, respectively.

5) Nucleoside disphosphate kinase (NDK)

NDK catalyses the transfer of phosphoryl groups from nucleoside triphosphates to nucleoside diphosphates (Lambeth *et al*., 1997). NDK primers used here amplified an NDK-like product in all cupped oyster species except for *C. virginica* and *O. denselamellosa*. They produced a 146 bp fragment (CDS and 3' UTR) that encoded 27 amino acids in all species screened. Seven interspecific sites (two amino acids replacement) in CDS and thirteen in 3' UTR (including indels) were variable among the seven oyster taxa compared here (Fig. 2).The Ts/Tv ratio was not relevant because no Tv was identified. Average nucleotide diversity (π) in CDS was 0.0216 (no gaps and missing sequences) and 0.0562 (0.0531 excluding gaps and missing sequences) in 3' UTR, respectively.

6) Ribosomal protein large 13 (RPL13)

RPL 13, interacts primarily with RNA, and is one of the major components of the 60S ribosomal subunit

that is involved in protein biosynthesis (Ban *et al*., 2000; Peters *et al*., 2007). RPL13 primers amplified a PRL13-like product in all cupped oyster species except for *C. nippona* (unknown gene detected), *C. iredalei* (no PCR product), *C. virginica* (no PCR product) and *O. denselamellosa* (no PCR product). The RPL13 primers amplified a 267 bp fragment (CDS and 3' UTR) that encoded 28 amino acids in the oyster species examined here. Four interspecific sites (one amino acid replacement) in CDS and 45 in 3' UTR (including indels) were variable among the five oyster taxa compared here (Fig. 2). The Ts/Tv ratio was 3.5 in CDS and 1.0 in 3' UTR. Average nucleotide diversity (π) in CDS was 0.0209 (no gaps and missing sequences) and 0.0494 (0.0495 excluding gaps and missing sequences) in 3' UTR, respectively. The homogeneity (incongruence) test showed that two data sets were congruent ($P > 0.05$), and simultaneous analysis of the total sequence data partitioned by ML and BI converged into a single tree (Fig. 1).

DISCUSSION

While only a relatively small number of oyster species were examined here for EST-derived SNP markers, the results suggest several species-specific genetic markers for distinguishing among cupped oyster species. While the monophyletic relationship of the cupped oyster clades here was consistent with a previous finding (Reece *et al*., 2008), limited informative sites from small gene fragments and a small number of inter-species sequenced were not sufficient to elucidate a clear relationship among *C. ariakensis*, *C. hongkongensis*, *C. nippona* and *C. sikamea*. Adding sequences from additional genes (mitochondrial and nuclear genes), which have been used successfully to resolve relationships among several other species groups (Gadagkar *et al*., 2005), could provide a possible solution to resolving the phylogeny of cupped oysters. In particular, developing more EST-SNP markers could help to explain unresolved clades among closely related *Crassostrea* species because the mtDNA genome evolves as a single unit, shows maternal inheritance in oyster species (Obata *et al*., 2008), and yields only a single gene tree, no matter how many base pairs or genes are sequenced (Amaral *et al*., 2009).

A set of conserved ortholog markers based on BTG and EF-1α would be ideal for assessing genetic diversity in related species as well as for cross-referencing transcribed sequences in comparative genomics studies. Although a considerable amount of DNA data are available for inferring phylogenetic relationships, those developed from known functional genes (type I markers) are often the most highly valued (Kim *et al*., 2011), as they allow biologists to link genomic information with biological information. Data mining from the public oyster EST database can provide more conserved ortholog type I markers that can complement currently available genomic markers for comparative mapping, marker-assisted selection, and ecological and evolutionary studies.

Average CDS frequency of interspecific level SNPs was estimated to be one SNP every seven bp an outcome that showed higher polymorphism rate than the observed intraspecific level of SNPs in Eastern oyster (one per 24 bp in Zhang and Guo, 2010) and Pacific oyster (one per 60 bp in Sauvage *et al*., 2007; one per 30 bp in Kim *et al*., 2011). In addition, the average 3' UTR frequency of interspecific level SNPs was estimated to be one SNP every five bp in a 593 bp of fragment (without MP sequences). This was also more polymorphic than observed intraspecific levels of SNPs in Eastern oyster (one per 16 bp [introns] in Zhang and Guo 2010) and Pacific oyster (one per 14 bp [introns] and one per 16 bp [3' UTR] in Kim *et al*., 2011). Regardless of whether comparisons were at the intra- or interspecific level studies, higher polymorphisms (SNPs and indels) were detected in 3' UTR regions (introns or non-coding regions in other studies) than CDS a result that is consistent with previous oyster studies (Sauvage *et al*., 2007; Zhang and Guo 2010; Kim *et al*., 2011). With one interspecific level SNP per seven bp identified here and approximately one intraspecific level SNP per 20-30 bp in previous studies (Zhang and Guo 2010; Kim *et al*., 2011), oysters are likely to be among the most polymorphic organisms that have been evaluated to date. Similar intra-species levels of high

polymorphism, about one SNP per 20 bp, have been reported in the nematode *Caenorhabditis remanei* (Cutter *et al*., 2006) and the sea squirt *Ciona savignyi* (Small *et al*., 2007), and one SNP per 40 bp in the giant freshwater prawn *Macrobrachium rosenbergii* (Jung *et al*., 2014).

In summary, the genic EST-derived SNP markers, developed here from a public dbEST from Pacific and Eastern oysters for other related cupped oyster species, showed high levels of polymorphism and a high rate of among species transferability. The study also validated six new interspecific level SNP markers via mining existing ESTs and re-sequencing for the first time. In total 213 putative SNPs were identified from re-sequencing data in six amplicons (without SNPs detected in 3' UTR of MP amplicons) and revealed an average of one SNP per seven bp in interspecific comparisons. This very high polymorphic SNP frequency was much higher than intraspecific level comparisons of SNPs in Eastern and Pacific oysters, making oysters one of the most polymorphic organisms. Candidate SNPs and SNP markers evaluated by re-sequencing yielded a high success rate for transferability. The success rate for finding transferable EST-derived SNP markers can be improved if the quality of conserved sequences and contigs are enhanced by mining oyster dbESTs. Undoubtedly, this approach will represent an efficient alternative for rapid marker development and SNP discovery in closely related oyster species and will be helpful for addressing phylogenetic, population genetic and stock analysis questions (pedigree monitoring) in oyster species. Type I SNP markers and identified SNPs will be especially informative for genome mapping studies and characterization of gene function (candidate gene analysis) in oyster species in general.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Kimberly Reece and Dr. Dennis Hedgecock for providing oyster samples. We are also thankful to Dr. Peter Mather and Dr. Patrick Gaffney in regard to constructive comments and valuable suggestions on the manuscript. The authors also thank Dr. Patrick Gaffney for sequencing of a few oyster samples. This project was supported by Korea Science and Engineering Foundation (Grant No.2005-215-F00009) and by a partial grant from the National Fisheries Research and Development Institute (RP-2014-BT-028) in Korea.

REFERENCES

Amaral, A.R., Coelho, M.M., Marugan-Lobon, J. and Rohlf, F.J. (2009) Cranial shape differentiation in three closely related delphinid cetacean species: insights into evolutionary history. *Zoology*, **112**: 38-47.

Behereqaray, L.B. (2008) Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology*, **17**: 3754-3774.

Baldauf, S.L., Palmer, J.D. and Doolittle, W.F. (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proceedings of the National Academy Sciences of the United States of America*, **93**: 7749-7754.

Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**: 905-920.

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L. and Schnable, P.S. (2007) SNP discovery via 454 transcriptome sequencing. *Plant Journal*, **51**: 910-918.

Bouck, A. and Vision, T. (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology*, **16**: 907-924.

Cutter, A.D., Baird, S.E. and Charlesworth, D. (2006) High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. *Genetics*, **174**: 901-913.

FAO. (2013) Fisheries and Aquaculture Statistical Databases, Global Aquaculture Production (online query).

Fleury, E., Huvet, A. Lelong, C., de Lorgeril, J., Boulo, V., Gueguen, Y., Bachère, E., Tanguy, A., Moraga, D., Fabioux, C., Lindeque, P., Shaw, J., Reinhardt, R., Prunet, P., Davey, G., Lapègue, S., Sauvage, C., Corporeau, C., Moal, J., Gavory, F., Wincker, P., Moreews, F., Klopp, C., Mathieu, M., Boudry, P. and Favrel, P. (2009) Generation and analysis of a 29,745 unique Expressed Sequence Tags from the Pacific oyster (*Crassostrea gigas*) assembled into a publicly accessible database: the Gigas Database. *BMC Genomics*, **10**: 341-355.

Gadagkar, S.R., Rosenberg, M.S. and Kumar, S. (2005) Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology Part B: Molecular and Development Evolution*, **304**: 64-74.

Galat, A. (1999) Variations of sequences and amino acid compositions of proteins that sustain their biological

functions: An analysis of the cyclophilin family of proteins. *Archives of Biochemistry and Biophysics*, **371**: 149-162.

Jenny, M., Chapman, R., Mancia, A., Chen, Y.-A., McKillen, D., Trent, H., Lang, P., Escoubas, J.-M., Bachere, E., Boulo, V., Liu, J., Cunningham, C., Cupit, P., Tanguy, A., Guo, X., Moraga, D., Boutet, I., Huvet, A., De Guise, S., Almeida, J. and Warr, G.A. (2007) A cDNA Microarray for *Crassostrea* virginica and *C. gigas*. *Marine Biotechnology*, **9**: 577-591.

Jung, H., Kim, W.J. and Gaffney, P.M. (2006) Development of single nucleotide polymorphisms (SNPs) in *Crassostrea ariakensis* and related *Crassostrea* species. *Journal of Shellfish Research*, **25**: 742.

Jung, H., Lyons, R.E., Li, Y., Thanh, N.M., Dinh, H., Hurwood, D.A., Salin, K.R. and Mather, P.B. (2014) A candidate association study for growth performance in an improved giant freshwater prawn (*Macrobrachium rosenbergii*) culture line. *Marine Biotechnology*, **16**: 161-180.

Kim, W.J., Jung, H. and Gaffney, P.M. (2011) Development of type I genetic markers from expressed sequence tags in highly polymorphic species. *Marine Biotechnology*, **13**: 127-132.

Kim, W.J., Dammannagoda, S.T., Jung, H., Baek, I.S., Yoon, H.S. and Choi, S.D. (2014) Mitochondrial DNA sequence analysis from multiple gene fragments reveals genetic heterogeneity of *Crassostrea ariakensis* in East Asia. *Genes and Genomics*, **36**: 611-627

Lambeth, D.O., Mehus, J.G., Ivey, M.A. and Milavetz, B.I. (1997) Characterization and cloning of a nucleotide-diphosphate kinase targeted to matrix of mitochondrial in Pigeon. *Journal of Biological Chemistry*, **272**: 24604-24611.

Langdon, C., Evans, F., Jacobson, D. and Michael. B. (2003) Yields of cultured Pacific oysters *Crassostrea gigas* Thunberg improved after one generation of selection. *Aquaculture*, **220**: 227-244.

Lee, H., Cha, S., Lee, M., Cho, G.J., Choi, W.S. and Suk, H. (2003) Role of antiproliferative B cell translocation gene-1 as an apoptotic sensitizer in activation-induced cell death of brain microglia. *Journal of Immumology*, **171**: 5802-5811.

Li, L. and Guo, X. (2004) AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology*, **6**: 26-36.

Li, W., Rusiniak, M.E., Chintala, S., Gautam, R., Novak, E.K. and Swank, R.T. (2004) Murine Hermansky-Pudlak syndrome genes: regulators of lysosome-related organelles. *BioEssays*, **26**: 616-628.

Li, Q. and Wan, J. M. (2005) SSRHunter: development of a local searching software for SSR sites. *Hereditas*, **27**: 808-810.

Li, Q., Liu, S. and Kong, L. (2009) Microsatellites within genes and ESTs of the Pacific oyster *Crassostrea gigas* and their transferability in five other *Crassostrea* species. *Electronic Journal of Biotechnology*, **12**: 1-7.

Librado, P. and Rozas, J. (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**: 1451-1452.

Morin, P.A., Luikart, G., Wayne, R.K. and the SNP workshop group. (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution*, **19**: 208-216.

Morin, P. A. and McCarthy, M. (2007) Highly accurate SNP genotyping from historical and low-quality samples. *Molecular Ecology Notes*, **7**: 937-946.

Morton, B.R., Bi, I.V., McMullen, M.D. and Gaut, B.S. (2006) Variation in mutational dynamics across the maize genome as a function of regional and flanking base composition. *Genetics*, **172**: 569-577.

Obata, M., Shimizu, M., Sano, N. and Komaru, A. (2008) Maternal inheritance of mitochondrial DNA (mtDNA) in the Pacific oyster (*Crassostrea gigas*): a preliminary study using mtDNA sequence analysis with evidence of random distribution of MitoTracker-stained sperm mitochondira in fertilized eggs. *Zoological Science*, **25**: 248-254.

Peters, I.R., Peeters, D., Helps, C.R. and Day, M.J. (2007) Development and application of multiple internal reference (housekeeper) gene assays for accurate normalization of canine gene expression studies. *Veterinary Immunology Immunopathology*, **117**: 55-66.

Quilang, W., Wang, S., Li, P., Abernathy, J., Peatman, E., Wang, Y., Wang, L., Shi, Y., Wallace, R., Guo, X. and Liu, Z. (2007) Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers. *BMC Genomics*, **8**: 157-167.

Reece, K.S., Cordes, J.F., Stubbs, J.B., Hudson, K.L. and Francis, E.A. (2008) Molecular phylogenies help resolve taxonomic confusion with Asian *Crassostrea* oyster species. *Marine Biotechnology*, **153**: 709-721.

Ronquist, F. and Huelsenbeck, J.P. (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**: 1572-1574.

Sauvage, C., Bierne, N. and Boudry, P. (2007) Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene*, **406**: 13-22.

Small, K.S., Brudno, M, Hill, M.M. and Sidow, A. (2007) Extreme genomic variation in a natural population. *Proceedings of the National Academy Sciences of the United States of America*, **104**: 5698-5703.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**: 2688-2690.

Takahashi, N., Hayano, T. and Suzuki, M. (1989) Peptidyl-prolylcis-trans isomerase is the cyclosporin A-binding protein cyclophilin. *Nature*, **337**: 473-475.

Tamura, K., Peterson, D., Peterson, N., Nei, G. and Kumar, S. (2011) MEGA5: molecular evolutionary

genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology Evolution*, **28**: 2731-2739.

Zhang, L.S. and Guo, X.M. (2010) Development and validation of single nucleotide polymorphism markers in the eastern oyster *Crassostrea virginica* Gmelin by mining ESTs and resequencing. *Aquaculture*, **302**: 124-129.

Zhang, G., Fang, Z., Guo, X., Luo, R., Xu, F., Yang, P. and Zhang, L. *et al*. (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**: 49-54.