

RESEARCH ARTICLE

# Genetic diversity and outlier loci detecting of shell color variation in the Pacific oyster (*Crassostrea gigas*) by SNP markers<sup>★</sup>

Junlin Song<sup>1</sup>, Qi Li<sup>1,2,\*</sup>, Xiaoxiao Zhong<sup>1</sup>, Lingfeng Kong<sup>1</sup> and Hong Yu<sup>1</sup>

<sup>1</sup> Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao 266003, China

<sup>2</sup> Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

Received 21 August 2016 / Accepted 28 February 2017

**Abstract** – Color polymorphism is relatively common in marine mollusks. Shell color affects the visual perception of products which, consequently, influences consumer preference and product value. To increase the value of the Pacific oyster (*Crassostrea gigas*) for the half-shell market, four shell color strains (golden, white, purple, and black) of *C. gigas* have been developed through successive selective breeding. To investigate genetic variation and identify breeding signatures of *C. gigas*, the four shell color strains and three wild populations were assessed through using 133 single nucleotide polymorphism (SNP) markers. The genetic diversity analysis demonstrated that the shell color strain exhibited significant reduction in observed number of alleles in comparison with the wild population. However, there is no significant differences of observed heterozygosity and expected heterozygosity were identified among the seven populations. Based on three outlier tests, a total of 16 loci were identified under selection, and four common outlier loci (CgSNP82, CgSNP273, CgSNP646, and CgSNP1131) were detected based on all the methods.  $F_{ST}$  values showed significant genetic differentiation between shell color strains, as well as between shell color strains and wild populations. The information on the genetic variation and differentiation in shell color strains and wild populations of *C. gigas* is useful for setting up suitable guidelines for founding and maintaining of cultured stocks. The detected loci might benefit the breeding programs after being tested.

**Keywords:** *Crassostrea gigas* / SNP / Outlier Locus / Shell color variation / Genetic diversity / Artificial selection

## 1 Introduction

Shells of mollusks are extremely diverse in morphology and rich in color. They have been the subject of fascination throughout human history and well sought after by conchologists and shell lovers (Liu et al., 2009). Shell color strain of bivalve mollusk is also under the same condition. The shell color is a high-valued trait which shows increasing interest for aquaculture. As a marketing characteristic, shell color is considered as a significant trait that can be used to cater to consumers and influence price (Hoang et al., 2016). Shell color is a phenotypic trait, and may reflect the potential difference of genetic variation caused by selection over generations. Previous studies demonstrated that the coloration of shell may be affected or determined by environmental factors (Kalinowski et al., 2007). However, data collected from experiments indicate that genetic factors may play a vital role in the coloration of shell in some species (Liu et al., 2009; Zou

et al., 2014). Despite these progresses, genetic researches on the shell color of bivalve mollusk are still limited.

One of the effective way to have better understanding of shell color is to detect loci involved (Qin et al., 2007). As subject to selection, loci detecting is effective and crucial for comprehending the importance of past genetic adaptations. In addition, it can also be an useful method in determining the adaptive population differentiation (Beaumont and Balding, 2004). One of the advantages of outlier loci detecting is the ability to screen a large amount of molecular markers by genome scan (Narum and Hess, 2011). Genome scans and outlier approaches that have resulted in complete or near complete fixation of beneficial alleles, also can identify genes under effective selection. Outlier analysis has been used in marine fishes (Vilas et al., 2015; Bekkevold et al., 2016), gastropods (Riquet et al., 2013) and bivalve mollusk (Murray and Hare, 2006; Zou et al., 2014; Zhong et al., 2014b). Outlier loci detection has become a common practice to identify adaptive molecular variation in species of bivalve mollusk.

Genome scan which employs numerous genetic markers offer the opportunity to assess local adaptation and recognize candidate genes in both natural populations and cultured stocks under selection. In particular, genome scans based on single

<sup>★</sup> Supporting information is only available in electronic form at [www.alr-journal.org](http://www.alr-journal.org).

\* Corresponding author: [qili66@ouc.edu.cn](mailto:qili66@ouc.edu.cn)

nucleotide polymorphism (SNP) variation have already been performed for a range of organisms (Narum and Hess, 2011). SNPs were characterized by high genome abundance and low redundancy. As they could exhibit very large differences in allele frequencies between populations, SNPs could be used as candidate population informative loci (Yu *et al.*, 2015b).

The Pacific oyster (*Crassostrea gigas*) is currently the most widely farmed oyster in the world, ranking first in production among all other cultured animals (FAO, 2016) and its coloration is of much interest to the whole oyster industry (Brake *et al.*, 2004). Through selective breeding, four shell color strains (golden, white, black, and purple) of the Pacific oyster were successfully developed (Cong *et al.*, 2014). The large variation of shell color within the cultured populations of *C. gigas* manifested that shell color of *C. gigas* could be considered as a continuous trait (Brake *et al.*, 2004; Song *et al.*, 2016).

In the present study, we used four strains of shell color and three wild populations of *C. gigas* as specimens to assess their genetic diversity level, investigate their genetic differentiation and detect outlier loci through using outlier analysis methods based on 133 SNP markers. The aim of this study is to demonstrate the effects of selective breeding on genetic diversity as well as to identify candidate SNPs under selection, which will consequently benefit the breeding of *C. gigas* with desired shell colors.

## 2 Materials and methods

### 2.1 Sample collection and DNA extraction

In this work, we surveyed three wild populations and four shell color strains of *C. gigas*, separately named the gold shell strain (GS), white shell strain (WS), black shell strain (BS), and purple shell strain (PS) (Fig. 1). The samples of wild populations were collected from three locations in Shandong province, China: Rushan (RS; 48 individuals), Weihai (WH; 38 individuals), and Dongying (DY; 48 individuals). The samples of four shell color strains of one-year-old *C. gigas* were collected in 2014 from an oyster farm in Rushan (Fig. 2). The four shell color strains (GS, WS, BS, and PS), exhibiting steadily hereditary shell color traits were the fourth generation offspring produced via successive selection. In July 2013, 92 (47 ♀ × 45 ♂), 86 (41 ♀ × 45 ♂), 107 (50 ♀ × 57 ♂), and 96 (50 ♀ × 46 ♂) individuals were selected from the third-generation strains selected for golden, white, black, and purple shell traits to serve as parents for the fourth generation selected strains, respectively. The original base populations of the selective four shell color strains of *C. gigas* were sampled in Weihai in Shandong province, China.

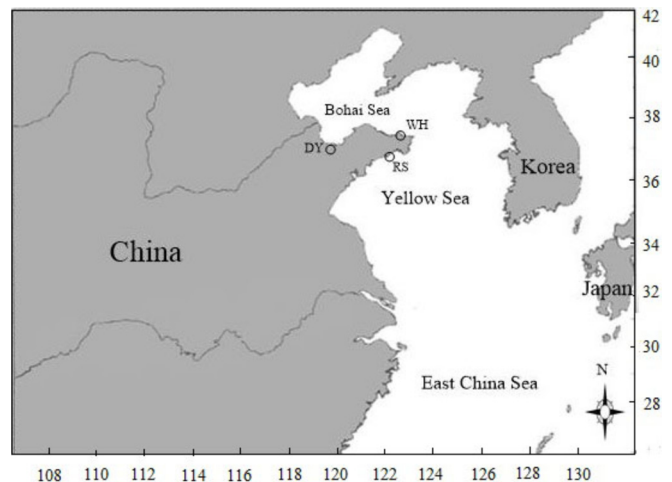
Besides, genomic DNA was extracted from the adductor muscle tissue using phenol-chloroform method described by Li *et al.* (2006). The concentration of DNA was measured with 1% agarose gel electrophoresis and Thermo NanoDrop-2000, respectively.

### 2.2 SNP genotyping

A total of 133 SNP markers were selected for genotyping with high resolution melting (HRM) assay (Zhong *et al.*, 2014a,b). PCR mixture consisted of 1 µL of template DNA



**Fig. 1.** Golden, white, black, purple, and wild shell color of *C. gigas* used in the study.



**Fig. 2.** Map showing locations and abbreviated names for three wild populations of *C. gigas*.

(10 ng), 0.05 µL of SYTO9 (Invitrogen Foster City, CA, USA), 0.4 µL of the forward and reverse primer (10 µM), 0.8 µL of dNTP mix, 0.6 µL of MgCl<sub>2</sub> and 5.65 µL of double distilled water, making up a volume of 10 µL. PCR cycling and HRM analysis were performed under the following conditions: an initial denaturation step of 95 °C for 5 min followed by 45 cycles of 95 °C for 20 s, a touch down of 68–58 °C for 20 s (0.5 °C/cycle), 72 °C for 20 s. After being denatured at 95 °C for 1 min, the probe and PCR products were allowed to anneal at 40 °C for 1 min. Subsequently, the HRM analysis of PCR products was performed within acquisitions per degree at the temperature ramping from 60 °C to 90 °C. The melting curves were analyzed based on the support of the Light Cycler 480 Gene Scanning software 1.5 (Roche Diagnostics).

### 2.3 Data analysis

Through using POPGENE 1.31 software, the observed number of alleles (*Na*), effective number of alleles (*Ne*), observed heterozygosity (*Ho*), expected heterozygosity (*He*), minor allele frequency (*MAF*), exact tests for deviations from Hardy–Weinberg equilibrium (HWE), inbreeding coefficients (*Fis*) and the Nei's genetic distance (Nei and Feldman, 1972) were calculated (Yeh *et al.*, 1999). All comparable statistics (*Na*, *Ne*, *Ho*, and *He*) were compared between the wild

**Table 1.** Statistical analysis of genetic diversity of selected shell color variants and wild populations of *Crassostrea gigas*.

Population	$N_a$	$N_e$	$H_o$	$H_e$
GS	1.917 ± 0.2529	1.471 ± 0.2949	0.332 ± 0.2156	0.292 ± 0.1478
WS	1.947 ± 0.2354	1.421 ± 0.2676	0.307 ± 0.2037	0.274 ± 0.1419
PS	1.969 ± 0.1941	1.439 ± 0.2977	0.309 ± 0.2473	0.278 ± 0.1542
BS	1.939 ± 0.2354	1.360 ± 0.3066	0.259 ± 0.2414	0.232 ± 0.1636
WH	1.985 ± 0.1387	1.423 ± 0.2686	0.310 ± 0.2076	0.276 ± 0.1364
DY	1.992 ± 0.0985	1.514 ± 0.3033	0.295 ± 0.1613	0.315 ± 0.1416
RS	1.985 ± 0.1387	1.501 ± 0.3032	0.282 ± 0.1601	0.308 ± 0.1481
Selected	1.943 ± 0.2295	1.423 ± 0.2917	0.302 ± 0.2270	0.269 ± 0.1519
Wild	1.987 ± 0.1253	1.479 ± 0.2917	0.298 ± 0.1763	0.299 ± 0.1420
Total	1.965 ± 0.1774	1.451 ± 0.297	0.300 ± 0.2017	0.284 ± 0.1470

$N_a$ , observed number of alleles;  $N_e$ , effective number of alleles;  $H_o$ , observed heterozygosity;  $H_e$ , average expected heterozygosity.

populations and shell color strains using Mann–Whitney  $U$ -test (Sokal and Rohlf, 1995) implemented with SPSS 16.0. Besides, the analysis of molecular variance (AMOVA) was also used to distinguish the molecular variance at several hierarchical levels with 1000 permutations in Arlequin 3.5 (Excoffier and Lischer, 2010). The unrooted neighbor-joining tree (NJ tree) was constructed with the software POPTREE2 (Takezaki *et al.*, 2010) based on Nei's genetic distances.

Both global and pairwise  $F_{ST}$  were evaluated to detect candidate markers under selection using three simulation methods including FDIST2 approach implemented in LOSITAN (Antao *et al.*, 2008), Bayesian analysis of BAYESCAN (Foll and Gaggiotti, 2008), and hierarchical analysis of ARLEQUIN 3.5 (Excoffier and Lischer, 2010). For hierarchical analysis based on the Beaumont test, observed loci with  $F_{ST}$  values above simulated quantile distributions were designated as 95% and 99% outlier loci accordingly. In Bayesian analysis, it provided an independent test of the outlier loci detected by reversible-jump MCMC approach and a threshold of  $\log_{10}(PO) > 0.5$  was employed in the conducted tests. FDIST2 identifies markers under selection as candidates for divergent selection at the 99% confidence level on the basis of the alternative finite island model with  $F_{ST}$ . In addition, pairwise comparisons were conducted as recommended by Vilas *et al.* (2010).

Sequence annotation was conducted using BLASTx software in NCBI database (<http://www.ncbi.nlm.nih.gov/>) and OysterBase (<http://www.oysterdb.com>). Synonymous SNPs and non-synonymous SNPs were distinguished by the NCBI ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>).

## 3 Result

### 3.1 Genetic diversity

Genetic diversity indices for the four shell color strains and three wild populations of *C. gigas* were shown in Table 1. The observed number of alleles ( $N_a$ ) ranged from 1.917 to 1.969 in the shell color strains and varied from 1.985 to 1.992 in the wild populations. Reductions in  $N_a$  values were observed in the shell color strains ( $P = 0.05$ ; Mann–Whitney  $U$ -test). The effective number of alleles ( $N_e$ ) ranged from 1.360 to 1.471 in the shell color strains and from 1.423 to 1.514 in the wild

populations, respectively. The observed heterozygosity ( $H_o$ ) values were from 0.259 to 0.332 in the shell color strains and from 0.289 to 0.310 in the wild populations. The expected heterozygosity ( $H_e$ ) values varied from 0.232 to 0.292 within shell color strains and ranged from 0.276 to 0.315 in the wild populations. The  $H_o$  were lower than those expected in the wild populations DY and RS, while  $H_o$  were greater than  $H_e$  in the other populations. Between the shell color strains and wild populations, no obvious difference in the  $H_e$ ,  $H_o$ , and  $N_e$  were observed ( $P > 0.05$ ; Mann–Whitney  $U$ -test). The information of the 133 SNPs evaluated from the four shell color strains and three wild populations were presented in Table S1. Most loci had two alleles in the seven populations, while 35 loci only had one allele in one or even more populations, including 11 loci in GS (CgSNP82, CgSNP150, CgSNP222, CgSNP230, CgSNP257, CgSNP374, CgSNP401, CgSNP653, CgSNP717, CgSNP722, and CgSNP1061), seven loci in WS (CgSNP103, CgSNP273, CgSNP472, CgSNP548, CgSNP764, CgSNP1039, and CgSNP1052), four loci in PS (CgSNP273, CgSNP401, CgSNP653, and CgSNP764), eight loci in BS (CgSNP150, CgSNP273, CgSNP225, CgSNP400, CgSNP694, CgSNP715, CgSNP791, and CgSNP909), two loci in WH (CgSNP29 and CgSNP549), one loci in RS (CgSNP779), and two loci in DY (CgSNP283 and CgSNP309). After sequential Bonferroni correction, 91 loci in the 931 single-locus showed deviation from HWE due to heterozygote deficiency or excess ( $P < 0.05/133$ ), containing 54 loci in the cultured population and 37 loci in the wild population.

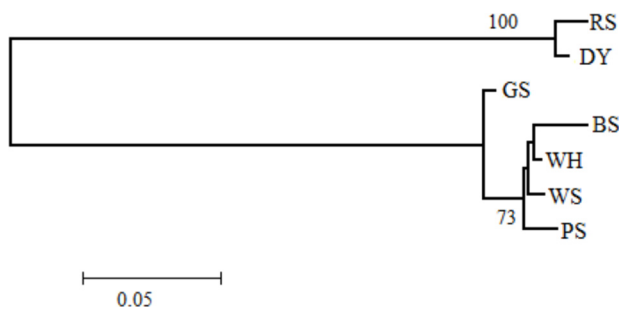
### 3.2 Genetic differentiation

The pairwise Nei's genetic distances among populations varied from 0.0091 to 0.4238 (Table 2). Particularly, the lowest value was observed between WS and WH populations, while the highest estimate was between BS and RS. The NJ tree generated from values of Nei's genetic distances is shown in Figure 3. The seven populations were separated into two groups. One group consisted of WH, GS, WS, PS, and BS, while the other consisted of DY and RS populations. When the outlier was removed from the data set, no different topologies of NJ tree was observed (data not shown). The estimation of  $F_{ST}$  values indicated high genetic differentiation among the

**Table 2.** Pairwise estimates of Nei's unbiased measures of genetic distance (above diagonal) and  $F_{ST}$  (below diagonal) between all *C. gigas* populations.

	GS	WS	PS	BS	WH	RS	DY
GS	0	0.0215	0.0323	0.0417	0.0209	0.3836	0.3773
WS	<b>0.0456</b>	0	0.0169	0.0245	0.0091	0.4040	0.3975
PS	<b>0.0668</b>	<b>0.0324</b>	0	0.0307	0.0185	0.4001	0.3976
BS	<b>0.0984</b>	<b>0.0620</b>	<b>0.0767</b>	0	0.0196	0.4238	0.4135
WH	<b>0.0430</b>	<b>0.0161</b>	<b>0.0371</b>	<b>0.0491</b>	0	0.3962	0.3877
RS	<b>0.3612</b>	<b>0.3781</b>	<b>0.3720</b>	<b>0.4255</b>	<b>0.3775</b>	0	0.1370
DY	<b>0.3710</b>	<b>0.3932</b>	<b>0.3834</b>	<b>0.4413</b>	<b>0.3932</b>	<b>0.2438</b>	0

Significant  $F_{ST}$  values are in bold ( $P < 0.05/23$ ).



**Fig. 3.** Neighbor-joining tree topology showing the genetic distance among four shell color strains and three wild populations of *C. gigas*. The numbers refer to the percentage bootstrap values from 1000 replications of resampled loci.

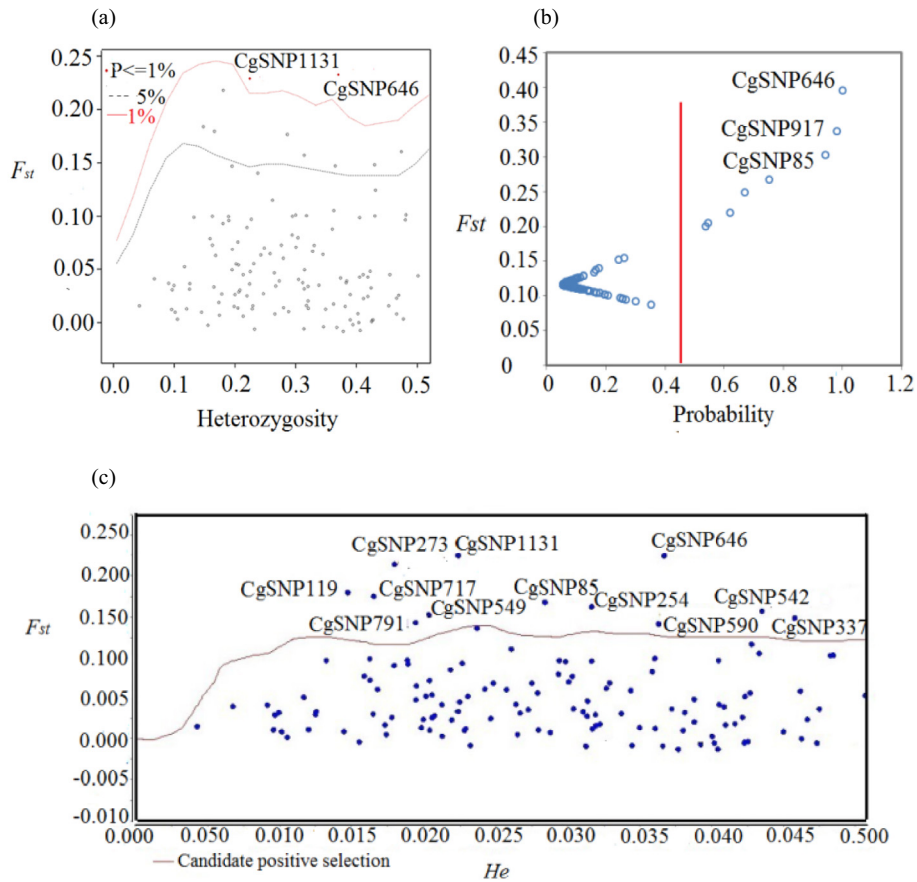
studied populations ( $F_{ST}=0.27$ ). The seven populations were treated as one group for analyzing molecular variance (AMOVA), which showed that genetic variation among populations accounted for 27% of the total variation and 73% of the variance was distributed within populations.

### 3.3 Outlier loci analysis

Among all the markers tested, a total of 16 markers were successfully detected in four shell color strains (as shown in Table 3). All three outlier detecting methods were conducted with the same data set for global analysis. Through hierarchical island model, nine loci were identified as outlier, including seven loci within the 95–99%  $F_{ST}$  level and two loci outside the 99%  $F_{ST}$  level (Fig. 4). Besides, the nine loci were also

**Table 3.** Outlier SNPs detected using Arlequin, LOSITAN, and BayeScan tests.

Locus	Hierarchical analysis			FDIST2		Bayesian			Type	Annotation
	Obs $F_{ST}$	$P$	Levels	$F_{ST}$	$P$	$P$	$\log_{10}(PO)$	$F_{ST}$		
cgSNP36						0.621	0.215	0.221	S (Gly)	TPA-inf:carosine synthase 1-like protein
cgSNP82	0.181	0.03	0.03	0.181	0.994	0.669	0.306	0.249	S (Ser)	Unknown
cgSNP85						0.943	1.215	0.303	N (Ser/Ile)	Unknown
cgSNP254	0.162	0.03	0.03	0.163	0.985				S (Arg)	Caltractin-like
cgSNP273	0.214	0.012	0.012	0.214	0.996	0.754	0.486	0.268	S (Val)	UPF0686 protein C11orf1-like protein
cgSNP337	0.15	0.031	0.031	0.149	0.986				S (Val)	<i>Plasmodium ovale</i> genome assembly
cgSNP542	0.158	0.026	0.026	0.158	0.987				S (Glu)	Kielin/chordin-like protein
cgSNP549				0.153	0.987				S (Ile)	<i>Parasteatoda tepidariorum</i> multiple epidermal growth factor-like domains protein 10
cgSNP590				0.142	0.982	0.537	0.064	0.201	S (Ser)	<i>Macaca fascicularis</i> complete genome
cgSNP621	0.172	0.015	0.015	0.171	0.993				S (Cys)	Ornithine decarboxylase
cgSNP646	0.226	0.006	0.006	0.225	0.999	1	1000	0.396	N (Glu-Gly)	Unknown
cgSNP717	0.176	0.032	0.032	0.176	0.99				S (Phe)	Acyl-CoA desaturase
cgSNP779									S (Lys)	Vacuolar protein sorting-associated protein 29
cgSNP791				0.144	0.981				N (Arg-Cys)	Isochorismatase domain-containing protein 2
cgSNP917						0.98	1.681	0.338	S (Cys)	MAM and LDL-receptor class A domain-containing protein 1-like
cgSNP1131	0.224	0.008	0.008	0.225	0.999	0.545	0.079	0.205	N (Asp-Asn)	<i>Tupaia chinensis</i> zinc finger CCCH-type, antiviral 1 (ZC3HAV1), transcript variant X3



**Fig. 4.** Results of three outlier tests. (a)  $F_{ST}$  values for SNPs detected by hierarchical analysis using Arlequin 3.5.  $F_{ST}$  are plotted against heterozygosity. (b)  $F_{ST}$  values analyzed by Bayesian approach.  $F_{ST}$  is plotted against the probability. The vertical line denotes the threshold used for identifying outlier loci. (c)  $F_{ST}$  values for SNPs simulated by FDIST2 method.

**Table 4.** Results in pairwise population comparisons and across populations are shown. Outliers were not detected in the GS/BS and WS/BS comparison

Comparison	Locus	Hierarchical analysis		FDIST2		Bayesian	
		Obs $F_{ST}$	$P$	$F_{ST}$	$P$	$F_{ST}$	$P$
GS/WS	cgSNP1131	0.303	0.0036	0.304	0.999	0.225	0.685
	cgSNP82	0.273	0.005	0.267	1	0.209	0.557
GS/PS	cgSNP 646	0.439	0.0006	0.439	1	0.341	1
	cgSNP1131	0.358	0.006	0.358	0.999	0.270	0.838
WS/PS	cgSNP273	0.32	0.0007	0.320	1	0.289	0.894
PS/BS	cgSNP273	0.32	0.011	0.320	0.999	0.280	0.657

detected by the FDIST2 method, revealing 12 loci as candidates at the 99% confidence level (Fig. 4). In addition, eight loci were detected with statistically significant patterns of divergent genetic differentiation through using the Bayesian method (Fig. 4). Among the eight loci, two loci had  $\log_{10}$  values above 1 (particularly strong) and one had a  $\log_{10}$  of 1000, considered as decisive evidence for selection. Among all the 16 loci identified, four common loci (CgSNP82, CgSNP273, CgSNP646, and CgSNP1131) were consistently detected under selection by the three methods. The four markers were statistically significant with Hierarchical, Fdist2,

and BayeScan in one or two comparisons involving samples from the color populations (Table 4). In wild populations, outliers were not detected by global and pairwise analysis. CgSNP82 and CgSNP646 showed a significantly high  $F_{ST}$  based on three methods for the GS/WS and GS/PS comparisons, respectively. CgSNP1131 was identified as an outlier in the comparisons between GS and WS, and between GS and PS. CgSNP273 represented as an outlier through analyzing the WS/PS and PS/BS comparisons. Consequently, the four markers cannot be excluded as false positive. As for the four outliers, the minimum allele frequencies ranged from

0.0000 to 0.2778, from 0.0000 to 0.3289, from 0.0132 to 0.4737, and from 0.0132 to 0.3947, respectively. In addition, CgSNP82 had one allele in golden shell color strain, while CgSNP273 had one allele in white, black, and purple shell color strains. All of the 16 SNPs were located in the coding region, including 12 synonymous and four nonsynonymous. In addition, 13 SNPs could be annotated with the support of BLASTX software (Table 3).

## 4 Discussion

It is important to develop a clear understanding of the levels of diversity in cultured populations which can be utilized for selective breeding (Wang *et al.*, 2011). In comparison to the wild populations, numerous of studies have demonstrated a reduction of genetic diversity in the hatchery populations (Yu and Guo, 2004; Xiao *et al.*, 2011). However, there are still some researches showing no major difference in genetic variability of hatchery populations in comparison with wild population (Zhao *et al.*, 2009; Yu *et al.*, 2015a).

Numbers of alleles can accurately measure genetic diversity. In this study, it can be obviously found that four shell color strains exhibited significant reduction in number of alleles than in the wild populations. The restricted number of breeding individuals used in aquaculture operations can lead to random drift in gene frequencies between generations (Hedgecock and Sly, 1990) and a loss of genetic variability in a population (Primack, 1998). In addition, artificial directional selection may lead the population genotype to become homozygous, leading to a reduction in the numbers of alleles of the populations.

Expected heterozygosity value is another useful measure to assess the genetic diversity within populations. In this study there is still no significant difference in  $H_e$  values between the shell color strains and the wild populations of *C. gigas*. This result was observed maybe because an estimate of heterozygosity could be inflated if a strain was found using heterozygous parents, and because elimination of rare alleles has not made much effect on estimates of heterozygosity (Allendorf, 1986). In this study, heterozygote deficiencies related to significant deviation from HWE were observed in both the shell color strains and wild populations. Similar phenomenon has been reported in many cultured and wild populations of marine mollusks (Li *et al.*, 2006; Lemer and Planes, 2012), which might be caused by null alleles, limited number of founders, selection, Wahlund effect, and/or small sample size (Hansen *et al.*, 2001; Crooks *et al.*, 2013). In addition, expected heterozygosity value observed in the cultured populations was markedly lower than those based on AFLP markers (Song *et al.*, 2016), indicating that SNP markers were less discriminative and polymorphic in comparison with AFLPs (Frascaroli *et al.*, 2013).

According to the estimation of pairwise  $F_{ST}$ , significant genetic differentiation was detected between shell color strains, and between shell color strains and wild populations, being consistent with the previous results based on AFLP analyses (Song *et al.*, 2016). Till now, similar results have been reported between cultured stocks and wild populations of aquaculture animals such as bay scallop (Zhao *et al.*, 2009) and topmouth culter (*Culter alburnus*) (Wang *et al.*, 2007). The

substantial genetic differentiation between shell color strains indicates that there was the effect of positive artificial selection during the successive selective breeding for shell colors. The existence of significant genetic differentiation between shell color strains and wild populations may have been resulted from several factors such as the geographical separation, bottleneck effect, selection, and genetic drift (Hedgecock and Sly, 1990; Pampoulie *et al.*, 2006). The differentiation among the three wild populations studied might be caused by strong genetic drift effects existing in *C. gigas* (Rohfritsch *et al.*, 2013). And this finding is also supported by AMOVA analysis in which genetic difference among populations was significant. In the wild RS, WH, and DY populations, WH is genetically similar to farmed populations, but very different to the other two wild populations. This may be due to the fact that the original parents of the selective four shell color strains of *C. gigas* were sampled from wild population along the coast of Weihai.

The identification of markers linked to a trait of interest has played an indispensable role in breeding through marker assisted selection. Besides, the pigmentation of shell color is also an important trait in bivalve mollusks such as abalone (Liu *et al.*, 2009; Hoang *et al.*, 2016) and pearl oysters (Zou *et al.*, 2014). In the present study, we used three approaches to identify outlier loci. However, it is not comparable to the probabilities from hierarchical, FDIST2, and  $P$ -values obtained from BayeScan. The inconsistent results obtained by the three methods were probably caused by different assumptions. Among the three methods, FDIST2 was the most sophisticated approach as it incorporates hierarchical model and simulates a distribution for neutrally distributed markers (Narum and Hess, 2011).

The four SNPs (CgSNP82, CgSNP273, CgSNP646, and CgSNP1131) were detected as outlier loci through using all the three methods. Among the four SNPs, three located within coding region were synonymous and one was nonsynonymous. The nonsynonymous outlier SNP is obvious due to the potential effect of amino acid changes on protein structure and function. The usage of synonymous codon in some genes was affected by natural selection, which may affect codon usage bias (Chamary and Hurst, 2004). Thus synonymous outlier should also not be neglected.

Allele frequency shifts at outlier loci are expected to reflect selective responses caused by strong ecological gradients leading to local adaptation, either for direct associated genes or genes in direction connection the markers (Bekkevold *et al.*, 2016). Although the mechanisms underlying outlier differentiation may be varied and are unknown in the present study, a possible explanation is that outliers reflect a selective response in one or even more genes, which may be associated with shell color. In fact, it is not surprising that shell color is controlled by simple genetic pattern. Shell color is determined by similar one-locus two-allele systems in the bluish and greenish strains of the Pacific abalone (*Haliotis discus hannai*), and rare orange and yellow bay scallops (Adamkewicz and Castagna, 1988; Kobayashi *et al.*, 2004). In Chilean scallop (*Argopecten purpuratus*), color strains (including purple, brown, orange, yellow, and white) were determined by two loci which is a simple dominant model of epistasis (Winkler *et al.*, 2001). In the noble scallop (*Chlamys nobilis*), one-locus three-allele model was proposed to explain the orange–purple, purple, and brown colors (Zheng *et al.*, 2013). Among the four outlier loci,

CgSNP1131 was closely linked to the gene coding for zinc finger protein, which is related to metal ion binding, suggesting that metal ion might be associated with the shell color formation process. Suzuki and Nagasawa (2013) reported that metal elements were involved in the formation mechanism of mollusk shells, and significant difference in contents of metal ion observed between shell color variations in some marine mollusks (Hao *et al.*, 2015). The other three outlier loci (CgSNP82, CgSNP273, and CgSNP646) were unannotated or with unknown function. A possible explanation is that the outlier loci might have been under the effect of the artificial selection process. Selection would occur at just a few loci and would not affect all loci uniformly, which is different from other demographic factors (Rohfritsch *et al.*, 2013). The recognition of genetic markers related to color variation of shell has implications for these species for the reason that they may be beneficial in selective breeding.

Admittedly, the present study is only a fundamental step to understand the genetic factor of shell colors, and test the potential genetic markers under artificial selection in *C. gigas*. Further studies on the role of the candidate outlier SNPs are still required, which will make great contributions to a better understanding of the molecular mechanisms underlying diverse color patterns.

## Supplementary Material

**Summary of the statistics for 133 SNP loci in the wild and selected *Crassostrea gigas* populations.** The Supplementary Material is available at <http://www.alr-journal.org/10.1051/alr/2017009/olm>.

**Acknowledgements.** This study was supported by the grants from National Natural Science Foundation of China (31372524), Shandong Seed Project, Shandong Province (2014GHY115002, 2016ZDJS06A06), and Qingdao National Laboratory for Marine Science and Technology (2015ASKJ02).

## References

- Adamkewicz L, Castagna M. 1988. Genetics of shell color and pattern in the bay scallop *Argopecten irradians*. *Heredity* 79: 14–17.
- Allendorf FW. 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zool Biol* 5: 181–190.
- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinform* 9: 323.
- Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 13: 969–980.
- Bekkevold D, Gross R, Arula T, Helyar SJ, Ojaveer H. 2016. Outlier loci detect intraspecific biodiversity amongst spring and autumn spawning herring across local scales. *PLoS ONE* 11: e0148499.
- Brake J, Evans F, Langdon C. 2004. Evidence for genetic control of pigmentation of shell and mantle edge in selected families of Pacific oysters, *Crassostrea gigas*. *Aquaculture* 229: 89–98.
- Chamary JV, Hurst LD. 2004. Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: evidence for selectively driven codon usage. *Mol Biol Evol* 21: 1014–1023.
- Cong R, Li Q, Ge J, Kong L, Yu H. 2014. Comparison of phenotypic traits of four shell color families of the Pacific oyster (*Crassostrea gigas*). *J Fish Sci China* 21: 494–502 (in Chinese).
- Crooks L, Carlborg Ö, Marklund S, Johansson AM. 2013. Identification of null alleles and deletions from SNP genotypes for an intercross between domestic and wild chickens. *G3* 6: 1253–1260.
- Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10: 564–567.
- FAO. 2016. Fishery and aquaculture statistics 2014. FAO report. Rome: Food and Agriculture Organization of the United Nations.
- Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 180: 977–993.
- Frascaroli E, Schrag TA, Melchinger AE. 2013. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. *Theor Appl Genet* 126: 133–141.
- Hansen MM, Ruzzante DE, Nielsen EE, Mensberg KD. 2001. Brown trout (*Salmo trutta*) stocking impact assessment using microsatellite DNA markers. *Ecol Appl* 11: 148–160.
- Hao Z, Yang L, Zhan Y, *et al.* 2015. Biochemical components of different colored strains of cultured Japanese Scallop (*Mizuhopecten yessoensis*) under different cultivation systems. *Isr J Aquac Bamidgeh* 67: 1189–1197.
- Hedgecock D, Sly F. 1990. Genetic drift and effective sizes of hatchery-propagated stocks of the Pacific oyster *Crassostrea gigas*. *Aquaculture* 88: 21–38.
- Hoang TH, Qin JG, Stone DAJ, Harris JO, Duong DN, Bansemer MS. 2016. Colour changes of greenlip abalone (*Haliotis laevis* Donovan) fed fresh macroalgae and dried algal supplement. *Aquaculture* 456: 16–23.
- Kalinowski CT, Izquierdo MS, Schuchardt D, Robaina LE. 2007. Dietary supplementation time with shrimp shell meal on red porgy (*Pagrus pagrus*) skin colour and carotenoid concentration. *Aquaculture* 272: 451–457.
- Kobayashi T, Kawahara I, Hasekura O, Kijima A. 2004. Genetic control of bluish shell color variation in the Pacific abalone *Haliotis discus hannai*. *J Shellfish Res* 23: 1153–1156.
- Lemer S, Planes S. 2012. Translocation of wild populations: conservation implications for the genetic diversity of the black-lipped pearl oyster *Pinctada margaritifera*. *Mol Ecol* 21: 2949–2962.
- 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.
- Liu X, Wu F, Zhao H, Zhang G, Guo X. 2009. A novel shell color variant of the Pacific abalone *Haliotis discus hannai* Ino subject to genetic control and dietary influence. *J Shellfish Res* 28: 419–424.
- Murray MC, Hare M. 2006. A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*. *Mol Ecol* 15: 4229–4242.
- Narum SR, Hess JE. 2011. Comparison of  $F_{ST}$  outlier tests for SNP loci under selection. *Mol Ecol Resour* 11 (Suppl. 1): 184–194.
- Nei M, Feldman MW. 1972. Identity of genes by descent within and between populations under mutation and migration pressures. *Theor Popul Biol* 3: 460–465.
- Pampoulie C, Jorundsdottir TD, Steinarrson A, Petursdottir G, Stefansson MO, Danielsdottir AK. 2006. Genetic comparison of experimental fanned strains and wild Icelandic populations of Atlantic cod (*Gadus morhua* L.). *Aquaculture* 261: 556–564.

- Primack RB. 1998. Essentials of conservation biology. Sunderland: Sinauer Associates, pp. 253-276.
- Qin Y, Liu X, Zhang H, Zhang G, Guo X. 2007. Identification and mapping of amplified fragment length polymorphism markers linked to shell color in bay scallop, *Argopecten irradians irradians* (Lamarck, 1819). *Mar Biotechnol* 9: 66–73.
- Riquet F, Daguin-Thiébaud C, Ballenghien M, Bierne N, Viard F. 2013. Contrasting patterns of genome-wide polymorphism in the native and invasive range of the marine mollusk *Crepidula fornicata*. *Mol Ecol* 22: 1003–1018.
- Rohfritsch A, Bierne N, Boudry P, Heurtebise S, Cornette F, Lapègue S. 2013. Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. *Evol Appl* 6: 1064–1078.
- Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. New York: Freeman and Co, pp. 427–434.
- Song J, Li Q, Kong L, Yu H. 2016. Identification of candidate AFLP markers for shell color of the Pacific oyster (*Crassostrea gigas*) under artificial selection. *Biochem Syst Ecol* 66: 209–215.
- Suzuki M, Nagasawa H. 2013. Mollusk shell structures and their formation mechanism. *Can J Zool* 91: 349–366.
- Takezaki N, Nei M, Tamura K. 2010. POPTREE2: software for constructing population trees from allele frequency data and computing other population statistics with Windows-interface. *Mol Biol Evol* 27: 747–752.
- Vilas R, Bouza C, Vera M, Millán A, Martínez P. 2010. Variation in anonymous and EST-microsatellites suggests adaptive population divergence in turbot. *Mar Ecol Prog Ser* 420: 231–239.
- Vilas R, Vandamme SG, Vera M, et al. 2015. A genome scan for candidate genes involved in the adaptation of turbot (*Scophthalmus maximus*). *Mar Genomics* 23: 77–86.
- Wang L, Meng Z, Liu X, Zhang Y, Lin H. 2011. Genetic diversity and differentiation of the orange-spotted grouper (*Epinephelus coioides*) between and within cultured stocks and wild populations inferred from microsatellite DNA analysis. *Int J Mol Sci* 12: 4378–4394.
- Wang W, Chen L, Yang P, et al. 2007. Assessing genetic diversity of populations of topmouth culter (*Culter alburnus*) in China using AFLP markers. *Biochem Syst Ecol* 35: 662–669.
- Winkler FM, Estevez BF, Jollan LB, Garrido JP. 2001. Inheritance of the general shell color in the scallop *Argopecten purpuratus* (Bivalvia: Pectinidae). *J Hered* 92: 521–525.
- Xiao J, Cordes JF, Moss JA, Reece KS. 2011. Genetic diversity in U.S. hatchery stocks of *Crassostrea ariakensis* (Fujita, 1913) and comparison with natural populations in Asia. *J Shellfish Res* 30: 751–760.
- Yeh FC, Yang R, Boyle T. 1999. Microsoft Windows-based free ware for population genetic analysis (Release 1.31). Edmonton: University of Alberta.
- Yu H, Gao Y, Chen AL, Kong LF, Li Q. 2015a. Genetic diversity and population structure of the ark shell *Scapharca broughtonii* along the coast of China based on microsatellite. *Biochem Syst Ecol* 58: 235–241.
- Yu Q, Li Y, Xing Q, et al. 2015b. Identification of SNPs with different allele frequencies in China and Japan population of Pacific abalone (*Haliotis discus hannai*). *Conserv Genet Resour* 7: 837–840.
- Yu Z, Guo X. 2004. Genetic analysis of selected strains of eastern oyster (*Crassostrea virginica* Gmelin) using AFLP and microsatellite markers. *Mar Biotechnol* 6: 575–586.
- Zhao C, Li Q, Kong LF. 2009. Inheritance of AFLP markers and their use for genetic diversity analysis in wild and farmed scallop (*Chlamys farreri*). *Aquaculture* 287: 67–74.
- Zheng H, Zhang T, Sun Z, Liu W, Liu H. 2013. Inheritance of shell colours in the noble scallop *Chlamys nobilis* (Bivalve: Pectinidae). *Aquac Res* 44: 1229–1235.
- Zhong X, Li Q, Guo X, Yu H, Kong L. 2014a. QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. *Aquac Int* 22: 1877–1889.
- Zhong X, Li Q, Yu H, Kong L. 2014b. Genetic variation and breeding signature in mass selection lines of the Pacific oyster (*Crassostrea gigas*) assessed by SNP markers. *PLOS ONE* 9: e 108256.
- Zou K, Zhang D, Guo H, Zhu C, Li M, Jiang S. 2014. A preliminary study for identification of candidate AFLP markers under artificial selection for shell color in pearl oyster *Pinctada fucata*. *Gene* 542: 8–15.

**Cite this article as:** Song J, Li Q, Zhong X, Kong L, Yu H. 2017. Genetic diversity and outlier loci detecting of shell color variation in the Pacific oyster (*Crassostrea gigas*) by SNP markers. *Aquat. Living Resour.* 30: 10