

RESEARCH ARTICLE

Genetic Variation and Breeding Signature in Mass Selection Lines of the Pacific Oyster (*Crassostrea gigas*) Assessed by SNP Markers

Xiaoxiao Zhong, Dandan Feng, Hong Yu, Lingfeng Kong, Qi Li*

Key Laboratory of Mariculture Ministry of Education, Ocean University of China, Qingdao, China

* qili66@ouc.edu.cn



OPEN ACCESS

Citation: Zhong X, Feng D, Yu H, Kong L, Li Q (2016) Genetic Variation and Breeding Signature in Mass Selection Lines of the Pacific Oyster (*Crassostrea gigas*) Assessed by SNP Markers. PLoS ONE 11(3): e0150868. doi:10.1371/journal.pone.0150868

Editor: Filippos A. Aravanopoulos, Aristotle University of Thessaloniki, GREECE

Received: December 1, 2015

Accepted: February 5, 2016

Published: March 8, 2016

Copyright: © 2016 Zhong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by grants from the National High Technology Research and Development Program (2012AA10A405-6), National Natural Science Foundation of China (31372524), and Seed Improvement Project of Shandong Province.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In breeding industries, a challenging problem is how to keep genetic diversity over generations. To investigate genetic variation and identify breeding signatures in mass selected lines of Pacific oyster (*Crassostrea gigas*), three sixth-generation selected lines and four wild populations were assessed using 103 single nucleotide polymorphism (SNP) markers. The genetic diversity data indicated that the selected lines exhibited a significant reduction in the observed heterozygosity and observed number of alleles per locus compared with the wild populations ($P \leq 0.05$), indicating the selected lines tended to lose genetic diversity contrasted with the wild populations. The unweighted pair-group method with arithmetic mean (UPGMA) analysis showed that the wild populations and selected lines were not separated into two groups. Using four outlier tests, a total of 17 loci were found under selection at two levels. The global outlier detection suggested that 4 common outlier loci were subject to selection using both the hierarchical island model and Bayesian likelihood approaches. At regional level, 3 SNPs were detected as outlier using at least two outlier tests and one outlier SNP (CgSNP309) was overlapped in the two wild-selected population comparisons. The candidate outlier SNPs provide valuable resources for future association studies in *C. gigas*.

Introduction

The Pacific oyster (*Crassostrea gigas*), naturally distributed around Japan, China and Korea, is now an important cultivated oyster species worldwide [1]. Many countries has started introducing *C. gigas* since 1940s, mainly because of its rapid growth rate, high disease resistance and strong environmental adaptability [2]. The *C. gigas* is one of the most popular oyster species in China, and its main places of production are Shandong and Liaoning provinces [3, 4]. In 2012, China produced 3.94 million tons of oysters with *C. gigas* as one of the most dominant species [4, 5]. However, the broodstock used today remains largely unselected and thus *C. gigas* has gained little from heredity improvement by selective breeding [6].

Several selective breeding projects have been launched in couple of countries and obtained encouraging results [7–9]. With the aim of improving the productivity traits of *C. gigas*, a breeding program selected for fast growth has also been initiated in China. The first generation

of selection was carried out by mass selection on three cultured stocks from China, Japan and Korea in 2007 and the average improvement (increase in shell height) was 10% [4]. An increase in growth rate is detected over the successive six-generation selection, but little is known about the effects of strong directional selection on genetic level in the process of cultivation. The main focus of selected lines is on inbreeding and the decrease of genetic diversity [10, 11]. Yu and Guo [11] detected rare alleles of the 4 selected strains decreased significantly compared with a wild population in Eastern oyster (*Crassostrea virginica*). Moreover, Cruz et al. [12] found selected strains through breeding programs tended to lose genetic diversity compared with wild populations in Pacific white shrimp (*Litopenaeus vannamei*). The reduction of genetic diversity in a population may possibly reduce disease resistance and decrease adaptability to environmental changes [13, 14]. Therefore, it is crucial to survey the genetic diversity within or between selected lines and wild populations for successful hatchery management of *C. gigas*.

In recent years, the single nucleotide polymorphisms (SNPs) and simple sequence repeats (SSRs) developed from EST sequences have been extensively used in population genetic studies. Microsatellites have been chosen in many studies for their codominant inheritance and high allelic variability [15]. However, SNP genotyping results can be compared across different platforms and laboratories more easily than microsatellite data, therefore facilitating the integration and interpretation of genotyping data across different databases. Moreover, mutations resulting in some SNPs can be responsible for an adaptive phenotype or the direct target of selection and have been shown to be correlated with economic traits in many aquatic animals [16–18]. Therefore, SNP markers offer a valuable chance to explore the genetic basis of phenotypic variation during the selective breeding program in China.

In the present study, we used 103 SNP markers to evaluate the genetic diversity level, assess the genetic differentiation, and identify putative loci under selection in three sixth-generation selected lines and four wild populations of *C. gigas*. The main goal is to demonstrate the effects of cultivation on genetic diversity in the selective breeding program of *C. gigas*, and to identify candidate SNPs under selection through the use of outlier analysis methods.

Materials and Methods

Ethics Statement

The *C. gigas* does not belong to endangered species and the sample locations are not in protection.

Oyster Collections and DNA Extraction

Three selected lines and four wild populations of *C. gigas* were surveyed in our study. Collection details for the 7 *C. gigas* populations were shown in Fig 1 and Table 1. The cultivated oysters were randomly selected from three sixth-generation selected lines. In 2007, Pacific oysters from three cultivated populations in Pusan, South Korea (Stock K, 35.1°N, 129.1°E), Onagawa Bay, Miyagi Prefecture, Japan (Stock J, 38.3°N, 141.3°E), and Rushan, Shandong, China (Stock C, 36.4°N, 121.3°E) were used to establish selected lines for fast growth [4]. In each line, 30–60 pairs of these individuals from the top end of the shell height distribution were used as brood-stock oysters for the next generation [19]. In July 2012, 91 (50 ♀×41 ♂) and 107 (59 ♀×48 ♂) individuals from the fifth-generation selected lines of Japan and Korea (JS5 and KS5) were selected as parental oysters for the sixth-generation selected lines of Japan and Korea (JS6 and KS6), respectively. In July 2013, 85 (50 ♀×35 ♂) individuals in the fifth-generation selected lines of China (CS5) were selected as parental oysters for the sixth-generation selected lines of China (CS6). Fertilization was conducted at a ratio of 50 sperm per egg with 10⁷ oocytes for

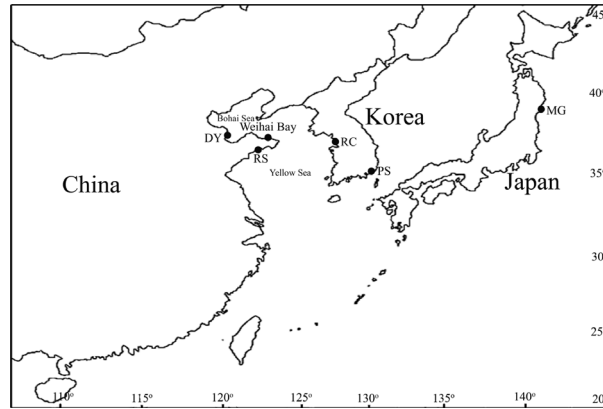


Fig 1. Approximate location of sampling sites shown with shaded circles. Populations are marked by abbreviations that correspond to [Table 1](#).

doi:10.1371/journal.pone.0150868.g001

each female [4, 19]. Oysters were cultivated in Weihai Bay, Shandong, China (37.3°N, 122.1°E). One-year-old oysters of the three sixth-generation selected lines (CS6, 48 individuals; JS6, 56 individuals and KS6, 48 individuals) were randomly selected for the analysis. The wild oysters were sampled from four locations: Rushan, Shandong Province, China (RS; 36.4°N, 121.3°E; 48 individuals; shell length, 3.45±0.48 cm, shell height, 5.63±0.71 cm, shell width, 1.87±0.35 cm); Dongying, Shandong Province, China (DY; 37.6°N, 119.0°E; 48 individuals; shell length, 3.93±0.68 cm, shell height, 4.96±0.90cm, shell width, 1.67±0.39 cm); Miyagi Prefecture, Japan (MG; 38.3°N, 141.3°E; 40 individuals; shell length, 4.22±0.78 cm, shell height, 7.32±0.91 cm, shell width, 2.01±0.37 cm) and Inchon, Korea (RC; 37.4°N, 126.6°E; 48 individuals; shell length, 3.62±0.78 cm, shell height, 5.73±0.98 cm, shell width, 1.63±0.39 cm). The adductor muscle was sampled and kept at -80°C. DNA was extracted using the phenol–chloroform method with a modification [20].

SNP Genotyping

A total of 103 SNP markers (from CgSNP4 to CgSNP457) were used in the study [21, 22]. SNPs were genotyped using the high resolution melting (HRM) method on the LightCycler 480 real-time PCR machine (Roche Diagnostics, Burgess Hill, UK) according to the procedure described by Zhong et al. [21]. The 10-µl mixture included 0.25 U Taq DNA polymerase (Takara, Dalian, China), 10 × PCR buffer, 0.2 mM dNTP mix, 0.2 µM of each primer set, 1.5 mM MgCl₂, 5 µM SYTO9 (Invitrogen Foster City, CA, USA) and 10 ng DNA [21]. The PCR cycling conditions were as follows: an activation step at 95°C, 5 min; 45–50 cycles of 95°C, 20 s; a touch down of 68°C to 58°C, 20 s (0.5°C /cycle) and extension for 72°C, 20 s. Before melting,

Table 1. Sample information for the Pacific oyster *Crassostrea gigas*.

Source	Abbreviation	Sample type (size)	Geographical coordinates
Rushan, Shandong, China	Stock C, RS	Base stock, wild population (48)	36.4°N, 121.3°E
Onagawa Bay, Miyagi Prefecture, Japan	Stock J, MG	Base stock, wild population (40)	38.3°N, 141.3°E
Pusan, South Korea	Stock K/PS	Base stock	35.1°N, 129.1°E
Dongying, Shandong, China	DY	Wild population (48)	37.6°N, 119.0°E
Inchon, Korea	RC	Wild population (48)	37.4°N, 126.6°E
Weihai Bay, Shandong, China	CS6, JS6, KS6	Selected lines (48, 56, 48)	37.3°N, 122.1°E

doi:10.1371/journal.pone.0150868.t001

the amplicons were heated at 95°C for 60s, and then cooled at 40°C for 60s. Melting curves were produced by heating amplicons from 60°C to 90°C [21].

Data Analysis

The observed number of alleles (N_a), effect number of alleles (N_e), shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), minor allele frequency (MAF), inbreeding coefficients (F_{is}) and chi-square analysis of Hardy–Weinberg equilibrium (HWE) were evaluated by POPGENE 1.32 [23]. All comparable statistics (N_a , N_e , I , H_o and H_e) were calculated between wild populations and selected lines using Mann-Whitney U-Test [24] implemented with SPSS 16.0.

Pairwise F_{st} values were calculated with Arlequin 3.5.1.3 [25]. F_{st} values were calculated for significance using 10000 permutations. Gene flow among wild populations was calculated following the formula [26]:

$$Nm = \frac{(1 - F_{st})}{4F_{st}}$$

The Nei's genetic distance [27] was also estimated by POPGENE 1.32. The unweighted-pair-group method with arithmetic-mean (UPGMA) tree was constructed by POPTREE2 [28]. Bootstrap analyses were conducted with 1000 replicates.

The population genetic structure analysis was estimated by Structure 2.3.4 with a model-based Bayesian procedure [29, 30]. The analysis was conducted using a burn-in period of 10^5 iterations and a run length of 10^6 MCMC replications. The number of clusters (K) was set from 1 to 10 with three replicates. The most probable K value was evaluated by STRUCTURE HARVESTER [31, 32].

To explore candidate SNPs under selection, F_{st} -based outlier tests were conducted using two island models including a finite island model (FDIST approach) and a hierarchical island model implemented in Arlequin 3.5.1.3 [25, 33], and Bayesian likelihood approach implemented in BayeScan [34]. To detect additional evidence of selection, lnRH tests were also conducted in population pairwise comparisons.

The outlier analysis was implemented at two levels. (1) An overall analysis encompassed all populations under the hierarchical island models. Three groups were set (CS6+RS+DY+RC vs. JS6+MG vs. KS6) based on the Structure result. The analysis was conducted using 50000 coalescent simulations with 10 groups of 100 demes. (2) An analysis was conducted between selected lines and wild populations from which the selected lines originated (CS6 and RS; JS6 and MG) under a finite island-model. We conducted 50000 coalescent simulations with 100 demes to generate the joint distribution of F_{st} versus heterozygosity. Loci which were outside the 99% confidence intervals were treated as outliers potentially under selection.

The outlier tests at the two levels were also calculated with the BayeScan software. The analyses were estimated using a pilot run of 20, a burn-in of 50000 with 100000 iterations each, a sample size of 5000, a thinning interval of 10, and an FDR of 0.05. Loci with \log_{10} values of the posterior odds (PO) >0.5 and 2.0 were regarded as candidate SNPs under selection with substantial and decisive evidence [35].

The lnRH test was applied to estimate the ratio of gene diversity (heterozygosity) of all loci in the population pairwise comparisons as follows:

$$\ln RH = \ln \left(\frac{\left(\frac{1}{1-H_{pop1}} \right)^2 - 1}{\left(\frac{1}{1-H_{pop2}} \right)^2 - 1} \right)$$

where H_{pop1} and H_{pop2} denote expected heterozygosity for population 1 and population 2 [36]. The $\ln RH$ values were standardized for each population comparisons and therefore the standardized distributions had a mean of zero and a standard deviation of one [37]. Those SNPs fell out of the 95% interval were taken as candidate loci under selection.

Sequence annotation was conducted by BLASTx software in NCBI database (<http://www.ncbi.nlm.nih.gov/>) and OysterBase (<http://www.oysterdb.com/>), and the critical E value was set as 1.0×10^{-6} . The NCBI ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>) was used to distinguish synonymous SNPs, non-synonymous SNPs or SNPs from untranslated regions (UTRs). The putative function of genes was identified by using the Gene Ontology (GO) annotation by mining the Swiss-Prot database (<http://us.expasy.org/sprot/sprot-top.html>) and OysterBase.

Results

Genetic Diversity

Genetic variability indices for the four wild populations and three selected lines of *C. gigas* were shown in Table 2. For all the genetic diversity parameters, the selected lines had lower values (mean N_a , N_e , I , H_o and H_e) than the wild populations, but no significant loss of N_e , I or H_e were detected ($P > 0.05$). The observed heterozygosities ranged from 0.2703 to 0.2939 with a mean of 0.2806 in the wild populations, and ranged from 0.2479 to 0.2733 with a mean of 0.2599 in the selected lines. There is a significant reduction of observed heterozygosities in the selected lines in contrast with the wild populations ($P = 0.05$). The observed number of alleles per locus varied from 1.9417 to 1.9709 in the selected lines, and varied from 1.9806 to 2.000 in the wild populations. There was a significant reduction of the observed number of alleles per locus in the selected lines compared with the wild populations ($P < 0.05$).

Information of the 103 SNPs evaluated from the three selected lines and four wild populations were summarized in S1 Table. All loci had two alleles in the 7 populations except that 19 loci had only one allele in one or more populations, including 3 loci (CgSNP252, CgSNP265 and CgSNP283) in CS6, 5 loci (CgSNP149, CgSNP176, CgSNP254, CgSNP265 and CgSNP305) in JS6, 6 loci (CgSNP14, CgSNP140, CgSNP176, CgSNP187, CgSNP252 and CgSNP319) in KS6, 1 locus (CgSNP283) in RS, 2 loci (CgSNP254 and CgSNP309) in RC and 2 loci (CgSNP176 and CgSNP283) in DY. A total of 83 of the 721 single-locus exact tests showed significant departures

Table 2. Genetic diversity parameters of selected lines and wild populations.

Population	N_a	N_e	I	H_o	H_e
CS6	1.9709±0.1690	1.5031±0.3206	0.4599±0.2034	0.2733±0.1766	0.3053±0.1585
JS6	1.9515±0.2160	1.4780±0.3472	0.4332±0.2247	0.2586±0.1846	0.2867±0.1729
KS6	1.9417±0.2354	1.4539±0.3464	0.4146±0.2346	0.2479±0.1969	0.2743±0.1772
RS	1.9903±0.0985	1.5138±0.3033	0.4774±0.1748	0.2939±0.1613	0.3154±0.1416
MG	2.000±0.0000	1.5362±0.3025	0.4892±0.1742	0.2801±0.1576	0.3261±0.1426
RC	1.9806±0.1387	1.5052±0.3206	0.4639±0.1945	0.2703±0.1637	0.3070±0.1548
DY	1.9806±0.1387	1.5012±0.3032	0.4664±0.1879	0.2780±0.1601	0.3083±0.1481
Selected	1.9547±0.2068	1.4783±0.3381	0.4359±0.2209	0.2599±0.1860	0.2888±0.1695
Wild	1.9879±0.0940	1.5141±0.3074	0.4742±0.1829	0.2806±0.1607	0.3142±0.1468
Total	1.9713±0.1504	1.4962±0.3228	0.4551±0.2019	0.2703±0.1734	0.3015±0.1582

Note: N_a , observed number of alleles; N_e , effective number of alleles; I , shannon's information index; H_o , observed heterozygosity; H_e , expected heterozygosity.

doi:10.1371/journal.pone.0150868.t002

from *HWE* after sequential Bonferroni correction ($P < 0.05/103$), and all loci showed heterozygote deficiencies except for CgSNP261 in CS6 and JS6, CgSNP36 in KS6 and MG, and CgSNP402 in RS.

Genetic Differentiation

Most pairwise *Fst* values were significant ($P < 0.001$), except for that estimated between wild populations RS and RC. The lowest value was observed between RS and RC ($Fst = 0.00689$), however the highest value was detected between JS6 and KS6 ($Fst = 0.11737$) (Table 3). Within the selected lines, the values varied from 0.06941 (KS6 and CS6) to 0.11737 (JS6 and KS6). Within the wild populations, the values ranged from 0.00689 (RS and RC) to 0.07046 (RC and MG). Moreover, gene flow among the wild populations ranged from 3.29811 (MG and RC) to 36.03447 (RS and RC).

The pairwise Nei's genetic distances among populations were also shown in Table 3. The lowest value was observed between RS and RC (0.0083), while the highest value was detected between KS6 and JS6 (0.0571). The UPGMA tree was constructed based on pairwise genetic distance (Fig 2). The UPGMA tree separated the 7 populations into two clusters. The KS6 originated from Korea belonged to one cluster while the other cluster included the other 6 *C. gigas* populations. Moreover, this cluster was further divided into 2 subgroups; the first subgroup contained JS6 and MG originated from Japan, and the second subgroup contained CS6, RS and DY originated from China, and RC from Korea.

Genetic Structure

The genetic structure test indicated that the most likely number of genetic groups was $K = 3$. The first group consisted of CS6, RS and DY originated from China, and RC from Korea. The second group consisted of JS6 and MG originated from Japan. Moreover, the selected line KS6 originated from Korea constituted the third genetic group (S1 Fig).

Outlier SNPs

Both the hierarchical island model and Bayesian likelihood approach revealed 4 loci (CgSNP140, CgSNP158, CgSNP176 and CgSNP209) under selection across all populations (Table 4; Fig 3). To detect specific footprint of artificial selection at regional level, we conducted pairwise tests between selected lines and wild populations from which the selected lines originated. The *Fdist* method revealed 5 candidate SNPs under selection, including 3 loci (CgSNP40, CgSNP225 and CgSNP236) for populations originated from China (CS6+RS) and 2 loci (CgSNP176 and CgSNP197) for populations originated from Japan (JS6+MG) (Table 4). The same two SNPs (CgSNP 225 and CgSNP 176) were also detected by BayeScan. The *lnRH*

Table 3. Pairwise *Fst* values (lower diagonal) and Nei's genetic distance (upper diagonal) among seven populations.

	CS6	JS6	KS6	RS	DY	MG	RC
CS6	0	0.0407	0.0349	0.0190	0.0323	0.0451	0.0193
JS6	0.07944	0	0.0571	0.0407	0.0357	0.0226	0.0443
KS6	0.06941	0.11737	0	0.0382	0.0542	0.0552	0.0346
RS	0.03067	0.07742	0.07460	0	0.0160	0.0413	0.0083
DY	0.05764	0.06847	0.10685	0.02380	0	0.0354	0.0191
MG	0.07851	0.04069	0.10513	0.06930	0.06006	0	0.0408
RC	0.03149	0.08571	0.06802	0.00689	0.03035	0.07046	0

doi:10.1371/journal.pone.0150868.t003

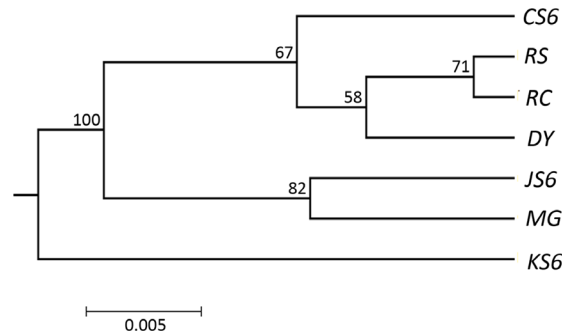


Fig 2. Phylogenetic tree of four wild populations (RS, DY, RC and MG) and three selected lines (CS6, KS6 and JS6) using the unweighted pair-group method with arithmetic mean (UPGMA) based on Nei's genetic distance derived from 103 SNPs. Numbers above branches indicate bootstrap values.

doi:10.1371/journal.pone.0150868.g002

Table 4. Outlier SNPs detected using Arlequin, BayeScan and InRH tests.

	Arlequin				BayeScan				LnRH test		
	Locus	Ho	Fst	P	Log ₁₀ (PO)	Fst	Probi	Alpha	LnRH	Type	Annotation
All populations	CgSNP140	0.640	0.837	0.000	1000	0.476	1.000	3.268		S (Ala)	CGI_10015432
	CgSNP158	0.500	0.311	0.002	1000	0.179	1.000	1.566		UTR	Unknown
	CgSNP176	0.121	0.363	0.002	3.097	0.205	0.999	1.732		N (Cys/Tyr)	CGI_10028477
CS6+RS	CgSNP209	0.631	0.809	0.000	1000	0.341	1.000	2.588		S (Glu)	Unknown
	CgSNP23	-	-	-	-	-	-	-	-2.258	N (The/Ala)	Unknown
	CgSNP40	0.570	0.239	0.008	-	-	-	-	-	S (Arg)	Unknown
	CgSNP82	-	-	-	-	-	-	-	2.265	S (Ser)	Unknown
	CgSNP83	-	-	-	-	-	-	-	2.265	S (Met)	Unknown
	CgSNP225	0.328	0.266	0.001	0.628	0.110	0.809	1.124	-3.628	S (Leu)	Hypothetical protein flotillin 2
JS6+MG	CgSNP236	0.379	0.250	0.003	-	-	-	-	2.947	N (Pro-Thr)	Hypothetical protein CGI_10002462
	CgSNP309	-	-	-	-	-	-	-	-2.084	S (Asn)	Collagen alpha-5(VI) chain
	CgSNP33	-	-	-	-	-	-	-	-3.137	N (Gln/Arg)	Unknown
	CgSNP140	-	-	-	-	-	-	-	-2.092	S (Ala)	CGI_10015432
	CgSNP176	0.415	0.497	0.000	3.398	0.241	1.000	1.995	Not fit	N (Cys/Tyr)	CGI_10028477
	CgSNP194	-	-	-	-	-	-	-	2.150	S (Lys)	Unknown
	CgSNP197	0.438	0.230	0.007	-	-	-	-	-	N (Glu/Lys)	CGI_10017873
	CgSNP283	-	-	-	-	-	-	-	-2.249	S (Ser)	NADH dehydrogenase [ubiquinone] iron-sulfur protein 2
	CgSNP309	-	-	-	-	-	-	-	-2.370	S (Asn)	Collagen alpha-5(VI) chain
	CgSNP375	-	-	-	-	-	-	-	-2.928	S (His)	UPF0451 protein C17orf61-like protein
CgSNP397	-	-	-	-	-	-	-	-2.092	S (Ser)	Hypothetical protein CGI_10010736	

Note: Ho, observed heterozygosity; PO, posterior odds; Alpha, positive values indicate positive selection, while negative values indicate putative balancing selection; Probi, probability; Not fit, CgSNP176 was monomorphic in JS6, thus not fit for the InRH test; S, synonymous; N, non-synonymous; UTR, untranslated region.

doi:10.1371/journal.pone.0150868.t004

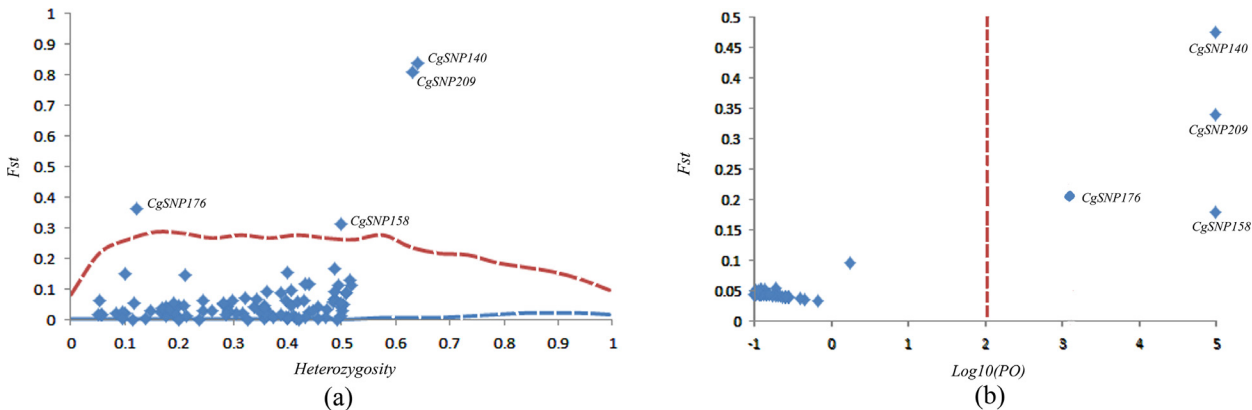


Fig 3. Results of two outlier tests in all populations. Locus names of putative outliers potentially affected by selection are indicated. (a) Hierarchical island model: empirical distribution of F_{st} against heterozygosity. The upper and lower lines are the 99% confidence intervals. (b) BayeScan: F_{st} estimates plotted against \log_{10} of the posterior odds (PO). The dashed lines correspond to the posterior odds 100 ($\log_{10}(PO) = 2$).

doi:10.1371/journal.pone.0150868.g003

test detected 12 outlier SNPs, including 5 SNPs (CgSNP23, CgSNP82, CgSNP83, CgSNP225 and CgSNP236) for CS6-RS comparison, 6 SNPs (CgSNP33, CgSNP140, CgSNP194, CgSNP283, CgSNP375 and CgSNP397) for JS6-MG comparison and one common SNP (CgSNP 309) for the two population comparisons. Totally, 17 candidate outlier SNPs were found under selection using Arlequin, BayeScan and InRH tests. Among the 17 SNPs, 16 were detected in the coding region, and 1 in the UTR. Among the 16 SNPs located in the coding region, 11 SNPs were synonymous and 5 nonsynonymous. Nine SNPs could be annotated by BLASTx software. The CGI_10015432, CGI_10017873, NADH dehydrogenase [ubiquinone] iron-sulfur protein 2 and Collagen alpha-5(VI) chain were involved in G-protein coupled receptor protein signaling pathway, lipid transport, oxidation-reduction process and lipoprotein metabolic process, however, the putative function of other five proteins (CGI_10028477, flotillin 2, CGI_10002462, UPF0451 protein C17 or f61-like protein and CGI_10010736) could not be identified using GO searches.

Discussion

Genetic Diversity

Maintenance of genetic variation is known to be important for long-term survival of populations because the level of variation determines their adaptability to environmental changes [38, 39]. Heterozygosity in many instances is considered as the primary parameter to reflect overall genetic variability of populations [40, 41]. In our study, the selected lines showed a significant reduction in the observed heterozygosity contrasted with the wild populations. Similar results have been detected in cultivated Atlantic salmon (*Salmo salar*) and Pacific abalone (*Haliotis discus hannai*) populations [42, 43].

Many researches have indicated loss of alleles is often more easily observed than decrease of heterozygosity in cultivated or selected populations [10, 11, 44]. Yu and Guo [11] detected the decrease in allele number was significant for selected lines NEH (by 36.8%) and CTS (by 35.1%) contrasted with the wild population of *C. virginica*. Wang et al. [45] found the cultured populations lost 9 of the 45 alleles (20%) compared with the wild population of bay scallop (*Argopecten irradians*). Compared with wild populations, 20% to 48% fewer alleles were detected in farmed salmon strains from Ireland and Norway [46]. In our study, the selected lines exhibited a significant reduction in observed number of alleles in contrast with the wild

populations ($P < 0.05$). Moreover, 0.98% (2/205) and 2.43% (5/206) fewer alleles were found in selected lines CS6 and JS6 than their wild progenitors (RS and MG), respectively. The CS6 stock seemed to be missing two rare alleles at 2 loci (CgSNP252 and CgSNP265) compared with RS. No individuals with the rare allele of the CgSNP265 found rarely in RS, were sampled from CS6. There was a possibility that this rare allele existed in CS6, but was not sampled in the population. As this rare allele frequency was 0.0521 in RS, the probability of completely missing this rare allele in the CS6 (48 samples) should be estimated as $0.9479^{96} = 0.0059$. Therefore, it is most likely that the CS6 has indeed lost this rare allele at the CgSNP265. Similarly, the JS6 may have indeed lost rare alleles at 4 loci (CgSNP149, CgSNP176, CgSNP254 and CgSNP305). Taken together, we can make a conclusion that the selected lines through breeding programs tended to lose genetic diversity compared with the wild populations. The decrease of genetic diversity may result from genetic drift, bottlenecks, and inbreeding caused by a reduced effective population size.

Significant deviations from *HWE* were detected both in selected lines and wild populations. In selected lines, heterozygote deficiency may be mainly because of a limited number of founders and artificial selection [47]. In the wild populations, heterozygote deficiency may be caused by Wahlund effect and natural selection. As heterozygotes with null alleles may be taken as homozygotes, null alleles may also account for the heterozygote deficiency.

Genetic Differentiation

The UPGMA analysis indicated that the selected lines and wild populations were not separated into two groups and even the wild populations and selected lines from the same origin country were clustered into the same subgroup except for KS6 and RC, suggesting there was no clear division between the wild populations and selected lines. The result is also supported by the Structure analysis.

Within the selected lines, moderate genetic differentiation ($0.05 < F_{st} < 0.15$) was detected. The UPGMA analysis also demonstrated the three selected lines fell into three groups. The significant genetic differentiation among the three selected lines may result from three different founder populations, thus providing the genetic basis for the establishment of selected lines for faster growth using three *C. gigas* stocks from China (Rushan), Japan (Miyagi Prefecture) and Korea (Pusan).

Within the wild populations, slight genetic differentiations ($F_{st} < 0.05$) were detected among the three populations RS, DY and RC. The N_m values among these populations ranged from 7.98723 (RC and DY) to 36.03447 (RS and RC), which appeared to be sufficiently large to swamp potential for large genetic differences. Since oyster adults are sedentary, larval dispersal becomes a decisive factor affecting gene flow among wild populations. Therefore, the long pelagic larval phase (14–21 days) and high fecundity (10 to 50×10^6 oocytes per female) of the Pacific oyster may account for the relatively large gene flow detected among RS, DY and RC [48, 49]. However, moderate genetic differentiations were found between population MG and other three populations RS, DY and RC, suggesting potential barriers to gene exchange may exist. Many studies showed a significant relationship between gene flow and geographic distance, suggesting that a pattern of isolation by distance (IBD) explained a great proportion of gene flow in *Macra chinensis* and *Crassostrea ariakensis* [50, 51]. The IBD pattern indicates long geographic distance can restrict migration among wild populations, which may explain the restricted gene flow between population MG and other three populations RS, DY and RC. Overall, the genetic relationship among the four wild populations could be visualized using the UPGMA tree, which indicated that RS and RC had the closest relationship with DY being the sister group, while MG was in another clade.

Outlier SNPs

In order to improve the productivity traits of *C. gigas*, successive generations of mass selection for fast growth were conducted in three *C. gigas* stocks from China, Japan and Korea since 2007. As there are similarities in controlled selective regimes, the selected lines may provide a valuable opportunity to study parallel evolution, which is taken as one of the most convincing manifestations of the effect of selection in driving adaptive change [52, 53]. Our study provided support for the hypothesis of parallel evolution at the DNA level as one outlier SNP (CgSNP309) was overlapped in the two wild-selected population comparisons. The parallel evolution is also shown in two ways in transcription level, one by the same DNA sequence polymorphisms with direct changes in gene expression, and the other by the different DNA sequence polymorphisms with the same downstream pathways during the cultivation [54]. Flori et al. [55] detected that different DNA sequence polymorphisms but the same physiological pathways were influenced during the cultivation in dairy cattle. The similar results were also found in rainbow trout (*Oncorhynchus mykiss*) and brook charr (*Salvelinus fontinalis*) [56, 57]. However, we can not conclude that parallel evolution at functional levels may account for the result in our study as only four outliers could be annotated with GO terms.

Four different tests (hierarchical island model, fdist island model, Bayesian likelihood approach and lnRH test) were used to detect outlier SNPs potentially under selection during the selective breeding program. These approaches rely on the rationale that selection will lead to increased genetic differentiation between populations and should exhibit a reduction or elevation in genetic variation contrasted with neutral genes. Actually, selection is not the only cause of variation changes at particular loci, reduced variation or increased differentiation can also arise from genetic drift, bottlenecks or founder events [58]. The effective population size for these selected lines is likely to have been less than 100, allowing for substantial random genetic drift over six generations.

The highly consistent results from the hierarchical island model and Bayesian likelihood approach for global outlier analyses strongly suggested that the identified outlier SNPs were subject to selection. Moreover, the same three SNPs (CgSNP140, CgSNP176 and CgSNP209) were also detected as outliers using selected lines and wild populations from which the selected lines originated (CS6, RS, JS6 and MG) (data not shown). At regional level, 3 SNPs (CgSNP225, CgSNP 236 and CgSNP 176) was detected as outlier using at least two analysis methods. Those outlier SNPs found in only one test should be considered with caution. The inconsistent results obtained by the three methods (fdist island model, Bayesian likelihood approach and lnRH test) were probably due to the different measures of variability and different assumptions [59]. Although the outlier analysis provides an encouraging result, association genetics and functional studies are ultimately required to confirm the outlier loci are involved in the artificial selection during the selective breeding program in China.

In this study, SNPs has been demonstrated to be a good marker of choice for monitoring the genetic variation in wild populations and selected lines of *C. gigas*. The genetic diversity analysis showed that the selected lines tended to lose genetic diversity contrasted with the wild populations during the successive six-generation selection. Moreover, the UPGMA results suggested there was no clear division between the wild populations and selected lines. In addition, a total of 17 loci were found under selection at two levels using four outlier tests. Further functional studies are needed to confirm the role of the candidate outlier SNPs during the domestication process in *C. gigas*.

Supporting Information

S1 Fig. Population genetic structure analysis with software Structure 2.3.4. (a) Genetic clusters obtained with three groups. Each individual is represented by one vertical line with 3 segments colored proportionally according to their belonging to a genetic group. Black lines separate individuals from different populations. (b) Graph of delta K.

(TIF)

S1 Table. Summary of the statistics for 103 SNP loci in the wild and selected *Crassostrea gigas* populations.

(XLS)

Acknowledgments

We wish to thank Qi Li for his help on the project designment and comments on previous drafts of this manuscript. We also thank Hong Yu, Lingfeng Kong and Dandan Feng to help us accomplish this project. This work was supported by the National High Technology Research and Development Program, National Natural Science Foundation of China, and Seed Improvement Project of Shandong Province.

Author Contributions

Conceived and designed the experiments: QL XZ. Performed the experiments: XZ DF. Analyzed the data: XZ. Contributed reagents/materials/analysis tools: HY LK. Wrote the paper: XZ QL.

References

1. Liu YG, Liu LX. Isolation and characterization of twenty two polymorphic simple sequence repeat markers from AFLP sequences of *Crassostrea gigas*. *Conserv Genet Resour.* 2015; 7: 659–661.
2. Dundon WG, Arzul I, Omnes E, Robertb M, Magnaboscoc C, Zambonc M, et al. Detection of type 1 Ostreid Herpes variant (OsHV-1 μ var) with no associated mortality in French-origin Pacific cupped oyster *Crassostrea gigas* farmed in Italy. *Aquaculture* 2011; 314: 49–52.
3. Wang Y, Ren R, Yu Z. Bioinformatic mining of EST-SSR loci in the Pacific oyster, *Crassostrea gigas*. *Anim Genet.* 2008; 39: 287–289. doi: [10.1111/j.1365-2052.2008.01701.x](https://doi.org/10.1111/j.1365-2052.2008.01701.x) PMID: [18307582](https://pubmed.ncbi.nlm.nih.gov/18307582/)
4. Li Q, Wang QZ, Liu SK, Kong LF. Selection response and realized heritability for growth in three stocks of the Pacific oyster *Crassostrea gigas*. *Fisheries Sci.* 2011; 77: 643–648.
5. DOF. China fisheries statistic yearbook (in Chinese). China Agriculture Press; 2013.
6. Sauvage C, Boudry P, Lapegue S. Identification and characterization of 18 novel polymorphic microsatellite makers derived from expressed sequence tags in the Pacific oyster *Crassostrea gigas*. *Mol Ecol Resour.* 2009; 9: 853–855. doi: [10.1111/j.1755-0998.2009.02525.x](https://doi.org/10.1111/j.1755-0998.2009.02525.x) PMID: [21564767](https://pubmed.ncbi.nlm.nih.gov/21564767/)
7. Ward RD, English LJ, McGoldrick DJ, Maguire GB, Nell JA, Thompson PA. Genetic improvement of the Pacific oyster *Crassostrea gigas* (Thunberg) in Australia. *Aquac Res.* 2000; 31: 35–44.
8. Langdon C, Evans F, Jacobson D, Blouin M. Yields of cultured Pacific oysters *Crassostrea gigas* Thunberg improved after one generation of selection. *Aquaculture* 2003; 220: 227–244.
9. Dégremont L, Bédier E, Boudry P. Summer mortality of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). II. response to selection for survival and its influence on growth and yield. *Aquaculture* 2010; 299: 21–29.
10. Hedgecock D, Sly F. Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture* 1990; 88: 21–38.
11. Yu ZN, Guo XM. Genetic analysis of selected strains of Eastern oyster (*Crassostrea virginica* Gmelin) using AFLP and microsatellite markers. *Mar Biotechnol.* 2005; 6: 575–586.
12. Cruz P, Ibarra AM, Mejia-Ruiz H, Gaffney PM, Pérez-Enríquez R. Genetic variability assessed by microsatellites in a breeding program of Pacific white shrimp (*Litopenaeus vannamei*). *Mar Biotechnol.* 2004; 6: 157–164. PMID: [14595549](https://pubmed.ncbi.nlm.nih.gov/14595549/)

13. Kong L, Li Q. Genetic comparison of cultured and wild populations of the clam *Coelomactra antiquata* (Spengler) in China using AFLP markers. *Aquaculture* 2007; 271: 152–161.
14. Allendorf FW, Phelps SR. Loss of genetic variation in a hatchery stock of *Cutthroat trout*. *Trans Am Fish Soc.* 1980; 109: 537–543.
15. Chistiakov DA, Hellemans B, Volckaert FAM. Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* 2006; 255: 1–29.
16. Prasertlux S, Khamnamtong B, Chumtong P, Klinbunga S, Menasveta P. Expression levels of RuvBL2 during ovarian development and association between its single nucleotide polymorphism (SNP) and growth of the giant tiger shrimp *Penaeus monodon*. *Aquaculture* 2010; 308: 83–90.
17. Thanh NM, Barnes AC, Mather PB, Li Y, Lyons RE. Single nucleotide polymorphisms in the actin and crustacean hyperglycemic hormone genes and their correlation with individual growth performance in giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* 2010; 301: 7–15.
18. Cong RH, Kong LF, Yu H, Li Q. Association between polymorphism in the insulin receptor-related receptor gene and growth traits in the Pacific oyster *Crassostrea gigas*. *Biochem Syst Ecol.* 2014; 54: 144–149.
19. Jiang Q, Li Q, Yu H, Kong LF. Genetic and epigenetic variation in mass selection populations of Pacific oyster *Crassostrea gigas*. *Genes Genom.* 2013; 35: 641–647.
20. Li Q, Park C, Kijima A. Isolation and characterization of microsatellite loci in the Pacific abalone, *Haliotis discus hannai*. *J Shellfish Res.* 2002; 21: 811–815.
21. Zhong XX, Li Q, Yu H, Kong LF. Development and validation of single nucleotide polymorphism markers in the Pacific oyster, *Crassostrea gigas*, using high-resolution melting analysis. *J World Aquacult Soc.* 2013; 44: 455–465.
22. Zhong XX, Li Q, Guo X, Yu H, Kong LF. QTL mapping for glycogen content and shell pigmentation in the Pacific oyster *Crassostrea gigas* using microsatellites and SNPs. *Aquacult Int.* 2014; 22: 1877–1889.
23. Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX. POPGENE, version 1.32: the user friendly software for population genetic analysis. Molecular Biology and Biotechnology Centre, University of Alberta; 1999.
24. Sokal RR, Rohlf FJ. *Biometry: The principles and practice of statistics in biological research*, 3rd edn. W.H. Freeman; 1995.
25. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010; 10: 564–567. doi: [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x) PMID: [21565059](https://pubmed.ncbi.nlm.nih.gov/21565059/)
26. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 1965; 19: 395–420.
27. Nei M. Genetic distance between populations. *Am Nat.* 1972; 106: 283–292.
28. Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows-interface. *Mol Biol Evol.* 2010; 27: 747–752. doi: [10.1093/molbev/msp312](https://doi.org/10.1093/molbev/msp312) PMID: [20022889](https://pubmed.ncbi.nlm.nih.gov/20022889/)
29. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. *Am J Hum Genet.* 2000; 67: 170–181. PMID: [10827107](https://pubmed.ncbi.nlm.nih.gov/10827107/)
30. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; 164: 1567–1587. PMID: [12930761](https://pubmed.ncbi.nlm.nih.gov/12930761/)
31. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software Structure, a simulation study. *Mol Ecol.* 2005; 14: 2611–2620. PMID: [15969739](https://pubmed.ncbi.nlm.nih.gov/15969739/)
32. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012; 4: 359–361.
33. Excoffier L, Hofer T, Foll M. Detecting loci under selection in a hierarchically structured population. *Heredity* 2009; 103: 285–298. doi: [10.1038/hdy.2009.74](https://doi.org/10.1038/hdy.2009.74) PMID: [19623208](https://pubmed.ncbi.nlm.nih.gov/19623208/)
34. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* 2008; 180: 977–993. doi: [10.1534/genetics.108.092221](https://doi.org/10.1534/genetics.108.092221) PMID: [18780740](https://pubmed.ncbi.nlm.nih.gov/18780740/)
35. Jeffreys H. *Theory of probability*, 3rd edn. Oxford University Press; 1961.
36. Schlötterer C, Dieringer D. In *Selective sweep*: Nurminsky D (ed) A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. Plenum Publishers; 2005.
37. Kauer MO, Dieringer D, Schlötterer C. A microsatellite variability screen for positive selection associated with the “Out of Africa” habitat expansion of *Drosophila melanogaster*. *Genetics* 2003; 165: 1137–1148. PMID: [14668371](https://pubmed.ncbi.nlm.nih.gov/14668371/)

38. Li Q, Xu KF, Yu RH. Genetic variation in Chinese hatchery populations of the Japanese scallop (*Patinopecten yessoensis*) inferred from microsatellite data. *Aquaculture* 2007; 269: 211–219.
39. Fisher RA. The genetical theory of natural selection, 2nd revised edition. Dover; 1958.
40. Leberg PL. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* 1992; 46: 471–494.
41. Liu FL, Yao JT, Wang XL, Repnikova A, Galanin DA, Duan D. Genetic diversity and structure within and between wild and cultivated *Saccharina japonica* (Laminariales, Phaeophyta) revealed by SSR markers. *Aquaculture* 2012; 358: 139–145.
42. Clifford SL, McGinnity P, Ferguson A. Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. *J Fish Biol.* 1998; 52: 118–127.
43. Li Q, Park C, Endo T, Kijima A. Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (*Haliotis discus hannai*). *Aquaculture* 2004; 235: 207–222.
44. Dillon RT, Manzi JJ. Hard clam, *Mercenaria mercenaria*, broodstocks: genetic drift and loss of rare alleles without reduction in heterozygosity. *Aquaculture* 1987; 60: 99–105.
45. Wang LL, Zhang H, Song LS, Guo X. Loss of allele diversity in introduced populations of the hermaphroditic bay scallop *Argopecten irradians*. *Aquaculture* 2007; 271: 252–259.
46. Norris AT, Bradley DG, Cunningham EP. Microsatellite genetic variation between and within farmed and wild Atlantic salmon (*Salmo salar*) populations. *Aquaculture* 1999; 180: 247–264.
47. Kohlmann K, Kerten P, Flajshans M. Microsatellite-based genetic variability and differentiation of domesticated, wild and feral common carp (*Cyprinus carpio* L.) populations. *Aquaculture* 2005; 247: 253–266.
48. Thorson G. Reproductive and larval ecology of marine bottom invertebrates. *Biol Rev.* 1950; 25: 1–45. PMID: [24537188](#)
49. Suquet M, Labbé C, Puyo S, Mingant C, Quittet B, Boulais M, et al. Survival, growth and reproduction of cryopreserved larvae from a marine invertebrate, the Pacific oyster (*Crassostrea gigas*). *PLoS One* 2014; 9: e93486. doi: [10.1371/journal.pone.0093486](#) PMID: [24695576](#)
50. Xiao J, Cordes JF, Wang HY, Guo X, Reece KS. Population genetics of *Crassostrea ariakensis* in Asia inferred from microsatellite markers. *Mar Biol.* 2010; 157: 1767–1781.
51. Ni L, Li Q, Kong LF. Microsatellites reveal fine-scale genetic structure of the Chinese surf clam *Macra chinensis* (Mollusca, Bivalvia, Mactridae) in Northern China. *Mar Ecol.* 2011; 32: 488–497.
52. Harvey PH, Pagel MD. The comparative method in evolutionary biology. Oxford University Press; 1991.
53. Roberge C, Einum S, Guderley H, Bernatchez L. Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Mol Ecol.* 2006; 15: 9–20. PMID: [16367826](#)
54. Vasemägi A, Nilsson J, McGinnity P, Cross T, O'Reilly P, Glebe B, et al. Screen for footprints of selection during domestication/captive breeding of Atlantic Salmon. *Comp Funct Genomic.* 2012; 2012: 628204.
55. Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, et al. The genome response to artificial selection: a case study in dairy cattle. *PLoS One* 2009; 4: e6595. doi: [10.1371/journal.pone.0006595](#) PMID: [19672461](#)
56. Tymchuk W, Sakhrani D, Devlin RH. Domestication causes large-scale effects on gene expression in rainbow trout: analysis of muscle, liver and brain transcriptomes. *Gen Comp Endocr.* 2009; 164: 175–183. doi: [10.1016/j.ygcen.2009.05.015](#) PMID: [19481085](#)
57. Sauvage C, Derôme N, Normandeau E, Cyr JS, Audet C, Bernatchez L. Fast transcriptional responses to domestication in the brook charr *Salvelinus fontinalis*. *Genetics* 2010; 185: 105–112. doi: [10.1534/genetics.110.115071](#) PMID: [20194962](#)
58. Kane NC, Rieseberg LH. Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus*. *Genetics* 2007; 175: 1823–1834. PMID: [17237516](#)
59. Li MH, Iso-Touru T, Laurén H, Kantanen J. A microsatellite-based analysis for the detection of selection on BTA1 and BTA20 in northern Eurasian cattle (*Bos taurus*) populations. *Genet Sel Evol.* 2010; 42: 32. doi: [10.1186/1297-9686-42-32](#) PMID: [20691068](#)