

RESEARCH ARTICLE

Genotyping-by-Sequencing of the regional Pacific abalone (*Haliotis discus*) genomes reveals population structures and patterns of gene flow

Bo-Hye Nam^{1‡}, Hyaekang Kim^{2‡}, Donghyeok Seol^{2,3}, Heebal Kim^{2,3}, Eun Soo Noh¹, Eun Mi Kim¹, Jae Koo Noh¹, Young-Ok Kim¹, Jung Youn Park¹, Woori Kwak^{3*}

1 Biotechnology Research Division, National Institute of Fisheries Science, Busan, Republic of Korea, **2** Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea, **3** Genome, Inc, Seoul, Republic of Korea

‡ These authors equally contributed and should be regarded as co-first authors.

* woori@hoonygen.com



OPEN ACCESS

Citation: Nam B-H, Kim H, Seol D, Kim H, Noh ES, Kim EM, et al. (2021) Genotyping-by-Sequencing of the regional Pacific abalone (*Haliotis discus*) genomes reveals population structures and patterns of gene flow. PLoS ONE 16(4): e0247815. <https://doi.org/10.1371/journal.pone.0247815>

Editor: Geerat J. Vermeij, University of California, UNITED STATES

Received: June 2, 2020

Accepted: February 12, 2021

Published: April 7, 2021

Copyright: © 2021 Nam 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 sequencing data generated and used in this study are available on the NCBI BioProject accession PRJNA665342.

Funding: This research was supported by the National Institute of Fisheries Science (No. R2021041 to BHN) and the Collaborative Genome Program of the Korea Institute of Marine Science and Technology Promotion funded by the Ministry of Oceans and Fisheries (No. 20180430 to BHN), Korea. WK is affiliated with eGenome, a commercial company. eGenome had no role in study design,

Abstract

Continuous monitoring of the present genetic status is essential to preserve the genetic resource of wild populations. In this study, we sequenced regional Pacific abalone *Haliotis discus* samples from three different locations around the Korean peninsula to assess population structure, utilizing Genotyping-by-Sequencing (GBS) method. Using *Pst*I enzyme for genome reduction, we demonstrated the resultant library represented the whole genome region with even spacing, and as a result 16,603 single nucleotide variants (SNVs) were produced. Genetic diversity and population structure were investigated using several methods, and a strong genetic heterogeneity was observed in the Korean abalone populations. Additionally, by comparison of the variant sets among population groups, we were able to discover 26 Korean abalone population-specific SNVs, potentially associated with phenotype differences. This is the first study demonstrating the feasibility of GBS for population genetic study on *H. discus*. Our results will provide valuable data for the genetic conservation and management of wild abalone populations in Korea and help future GBS studies on the marine mollusks.

Introduction

Abalones are a type of marine gastropods belonging to the family of reef-dwelling snails, Haliotidae. There are about 70 different species distributed across tropical and temperate coastal areas, such as Australia, the United States, or East Asia. Abalones have been a valuable food source for humans worldwide, and they are also an important research resource for ecological and evolutionary studies due to their diverse phenotypic appearances and global distribution [1]. Among abalone species, the Pacific abalone *Haliotis discus* is known as the most popular and highly valued species for commercial fisheries resources owing to the quality of meat and traditional consumer preferences [2, 3]. The *H. discus* species has a wide geographic

data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: WK is affiliated with eGnome, a commercial company. This does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

distribution in coastal waters of East Asia, and two subspecies exist: *H. discus hannai* inhabiting throughout Korea and northeastern Japan, and *H. discus discus* is distributed in the southern coastal areas of Korea and southwestern Japan [4].

For appropriate conservation of wild animals, continuous monitoring of the population structure and genetic status is essential because the maintenance of genetic diversity is important for whole species protection by providing disease- or stress-resistance genes to the gene pool [5]. Therefore, through genetic monitoring and biodiversity studies of wild populations, understanding the unique characteristics of regional populations and applying proper management strategies are suggested to preserve the genetic variation [6]. Previously, most studies on the Pacific abalones were, however, mainly focused on the development of artificial breeding programs and culture techniques such as seed production, spawning, or juvenile nursing, and there exist a relatively low number of researches carried out the genetic analysis of the wild populations [7, 8]. Moreover, these population genetics studies have been only performed using molecular markers such as microsatellites or mitochondrial DNA markers [9–11]. Most of them were carried out to mainly compare genetic diversity between wild and hatchery populations [12, 13]. So far, only two studies, utilizing microsatellite markers, explored genetic variation among regional wild abalone populations around the Korean peninsula. Traditionally, these molecular markers have been widely used for genetic studies of natural populations, even though it relied on small numbers of loci to make inferences [14]. In a few studies, those markers displayed lacked power to detect differentiation because they were limited to the narrow regions of the genome [15, 16]. For example, in a study using wild *Esox lucius* (Pike), restriction site-associated DNA sequencing (RAD-seq) not only resolved population genetic structure with better resolution than microsatellites but also were able to identify candidate loci under selection [17], and in another study, microsatellites had insufficient power to reveal the admixed status of one of the chicken populations in comparison with SNP dataset [18]. Thus, it could lead to insufficient and inconsistent results in abalones as well, making it difficult to simplify population characteristics and their evolutionary history [15]. Of the two population studies on regional Korean abalones mentioned above, one indicated genetic differentiation among populations collected from east, west, and southern coast [4], the other observed genetic separation between the eastern and pooled western and southern populations [19].

The recent advances in sequencing technology have led to enable sampling of the genome more densely and efficient production of more accurate and sufficient genetic information [15, 20]. One of the major recent advances has been the development of Genotyping-by-Sequencing (GBS). GBS uses restriction enzymes (REs) to reduce genome complexity and performs next-generation sequencing. The reduced representation approaches offer the ability to not only produce data covering the whole genome range with reduced cost but also provide high-resolution genetic information [21]. Furthermore, GBS can be employed for organisms with little or no previous genomic information [22]. With these advantages, it can provide major benefits especially in ecological and conservation genomics, since it is usually desirable to have a large sample size for studying wild populations, and sequencing of these samples with large genomes can cost a lot, particularly where the reference sequence information is absent [22]. GBS was originally designed for SNP genotyping in plant species, but recently the protocol has been quickly adopted to a wide variety of species, including marine animals [15]. For mollusks, GBS method has been applied in the construction of genetic linkage map of *Ruditapes philippinarum* (manila clam), the genomic best linear unbiased prediction (GBLUP) study of *Perna canaliculus* (greenshell mussel), and the population studies of *Haliotis fulgens* and *Haliotis laevigata* along the Western Australian coast [23–26]. Yet, it had not been previously utilized for research on *H. discus*.

In this present study, we performed GBS in a total of 102 regional abalones consisting of three regional Pacific abalone populations (*H. discus hannai*) collected from each of three seas (East, West, and South) in Korea and two outgroup populations, one Japan population (*H. discus discus*) and one red abalone population (*H. rufescens*). Using the variant information, we evaluated the genetic diversity and the population structure of *H. discus* populations inhabiting around the Korean peninsula and elucidated the impacts of geographical features and ocean currents in the gene flow among marine mollusks. Furthermore, by comparison of variant sets against red abalones, morphologically very different from *H. discus*, Korean Pacific abalone population-specific variants were discovered, which may be implicated with phenotypically different traits between groups. The results of the present study will provide an assessment of the utility of GBS for genetic analysis of abalone species and contribute to the management and genetic conservation of the wild Pacific abalone populations.

Materials and methods

Sample collection

Five abalone populations (102 samples) were used in this study (Table 1 and Fig 1). Of these, 24 Goseong samples were collected in Goseong, S. Korea, 30 Yeosu samples were collected in Geomundo, Yeosu, S. Korea, and 33 Taean sample were collected in Anmyeondo, Taean, S. Korea. Abalones were obtained by shell fish divers from each regional fishing community, and those released from farms were excluded, based on the green marks that farmed abalones possess on the shell. Five Japan samples from Mie, Japan were provided by Prof. Yamakawa of TUMST, and 11 Red samples collected in California coast, USA were provided by Prof. Lim of Mokpo University. Goseong, Yeosu, Taean, and Mie samples were used for characterization of Korean abalone population, and to identify Korean *H. discus hannai*-specific single nucleotide variants (SNVs), Goseong, Yeosu, Taean, and Red samples were used. No permissions were required for sample collection in this study because they were obtained during the season in which regional fishing communities allow harvesting.

GBS library construction and sequencing

Total genomic DNA was extracted from muscle tissue using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instruction. The amount of DNA was quantified using the standard procedure of Quant-iT PicoGreen dsDNA Assay Kit (Molecular Probes, Eugene, OR, USA) with Synergy HTX Multi-Mode Reader (Biotek, Winooski, VT, USA) and normalized to 20 ng/μl. DNA (200ng) was digested with 8U of High-fidelity *Pst*I (New England BioLabs, Ipswich, MA, USA) at 37 C for 2 hours and heated to 65°C for 20 minutes to inactivate the enzyme. DNA libraries for genotyping-by-sequencing (GBS) were constructed according to the protocols as described previously [27, 28] with minor modifications. The restriction digestion of DNA with *Pst*I was followed by ligation of adapters with specific barcode for each sample. The sets of 102 ligations were purified using QIAquick PCR Purification Kit (Qiagen). Ligation samples were pooled and 5μl were amplified in 50μl

Table 1. Species, sampling location, sample size, shell length, and sampling date for each abalone population used in this study.

Population	Abalone species	Location	Sample size	Shell length (cm)	Sampling date
Goseong	<i>H. discus hannai</i>	Goseong, S. Korea	24	9±2	July, 2013
Yeosu	<i>H. discus hannai</i>	Geomundo Island, Yeosu, S. Korea	30	10±1	August, 2014
Taeon	<i>H. discus hannai</i>	Anmyeondo Island, Taeon, S. Korea	33	10±1	August, 2014
Japan	<i>H. discus discus</i>	Mie, Japan	5	11±1	March, 2015
Red	<i>H. rufescens</i>	California, USA	11	-	October, 2015

<https://doi.org/10.1371/journal.pone.0247815.t001>



Fig 1. Approximate sampling locations of the abalone populations and trajectories of the major ocean currents. The name, location, and sample size of each population are presented on the map. The major ocean currents around the Korean peninsula are also shown with their names, locations, and the trajectories. Red lines represent warm currents and the blue line represents cold currents.

<https://doi.org/10.1371/journal.pone.0247815.g001>

reaction by PCR using AccuPower Pfu PCR Premix (Bioneer, Daejeon, South Korea) and 25 pmol of Illumina adaptors:

5' – AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT–3'

and

5' – CAAGCAGAAGACGGCATAACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT–3'.

PCR cycles consisted of 98°C for 5 min followed by 18 cycles of 98°C for 10 s, 65°C for 5 s, and 72°C for 5 s, with a final extension step at 72°C for 5 min. The PCR product was also purified using QIAquick PCR Purification Kit (Qiagen) and then evaluated the distribution of fragment sizes with BioAnalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The GBS library was sequenced in the Illumina NextSeq500 (Illumina, San Diego, CA, USA) with the length of 150 bp single-end reads following the manufacturer instruction. Read count distribution of the GBS library is shown in S1 Fig.

Variant calling

Before conducting variant calling analysis, we conducted de-multiplexing of GBS data using GBSXtoolkit [29]. Reads of each sample were mapped to the draft genome of *H. discus hannai*

constructed in our previous study [1], using Bowtie2 [30] with the default options. We then used SAMtools [31] to create index files for reference and bam files. Of the aligned reads, in order to remove possible PCR duplicates, regions covered by abnormally excessive number of identical reads were filtered (in this study, reads with more than 10x of average read depth), since applying PCR duplicates filter is not possible for GBS data utilizing a single-end sequencing approach. Then, genome Analysis Toolkit (GATK) [32] was used for downstream processing and variant calling. Local realignment was conducted using GATK to correct misalignments due to the presence of InDels. The “Haplotypecaller” of GATK was used for calling candidate single nucleotide variants (SNVs) and InDels. To avoid possible false positive variants, argument “VariantFiltration” of the same software was adopted with the following options: 1) SNVs with a phred-scaled quality score of less than 30 were filtered; 2) SNVs with QD (unfiltered depth of non-reference samples; low scores are indicative of false positives and artifacts) <5 were filtered; 3) SNVs with FS (phred-scaled P value using Fisher’s exact test) >200 were filtered, as FS represents variation on either the forward or the reverse strand, which are indicative of false-positive calls; 4) SNVs with MQ0 (the number of reads which have mapping quality zero) >4 or MQ0/DP (proportion of mapping quality zero reads over total depth) >0.1 were filtered to remove uncertain calls; 5) more than 3 SNVs within 10bp window were filtered. Likewise, we also excluded variants that were predicted to overlap with InDel mutation, or overlap with zero coverage regions in more than one sample. SNVs were then further filtered using VCFtools 4.0 [33] for minor allele frequencies < 5%, missing rates > 30%, Hardy-Weinberg equilibrium *P*-values < 0.05.

PCA, population structure analysis, and population parameters

Principal component analysis (PCA), based on the SNV information of Korea and Japan samples (91 samples), was conducted using SNPRelate R package [34]. For population admixture analysis, VCF file was converted to STRUCTURE input file format using PGDSpider [35], and population admixture was estimated using STRUCTURE [36] under admixture model for $K = 1-6$ using burn-in length of 5,000 and 50,000 MCMC replicates with 10 iterations. The most likely number of clusters (K) was determined by inferring $L(K)$ and ΔK statistic [37] in STRUCTURE HARVEST [38]. $L(K)$ is an average of replicated values of $\ln P(D)$ at each K . The most likely K is traditionally identified using the highest value of $L(K)$, but in many cases, $L(K)$ continues increasing slightly when the true K is reached. Therefore, we additionally computed ΔK based on Evanno et al [37]. Using ΔK , clearer peak can be shown at the most likely value of K . We utilized pixy [39] for estimating nucleotide diversity in 50Kb sliding windows to take into account invariant sites as well, and pairwise F_{ST} values with a 50Kb sliding window for each pair of populations was calculated using VCFtools 4.0 [33]. Lastly, to understand the effects of geographical distance on the genetic distance of populations, we performed Mantel test using GenAlex 6.5 [40] with 10,000 permutations. The genetic distance matrix was generated from 7,000 SNPs randomly selected by data thinning in PLINK 2.0 [41] since GenAlex could not handle whole SNP set. The geographical distance matrix was calculated from minimum distance between each regional group pair.

SNV annotation and identification of *H. discus hannai*-specific SNVs

SnEff (version 4.2) [42] was used to assign the expected impacts of all the filtered SNVs and their functional annotation. Since there is no database for abalone among the pre-built databases for SnEff, we used *Haliotis discus hannai* draft genome, constructed in our previous study, and its gene annotation to build database. Korean abalone-specific SNVs were then extracted based on the existence of non-synonymous (missense and nonsense) SNVs present only in Korean abalone populations (a total of 86 individuals) at orthologous positions compared to *H. rufescens* species, and in this step, only fixed SNVs were considered. Further, these

sets were subjected to enrichment analysis to identify significant SNVs using SnpSift (version 4.2) CaseControl [43]. CaseControl analysis counted the number of genotypes present in case-control groups, and then p-value calculation was performed using Fisher's exact test to identify SNVs that were significantly associated with case vs. control groups [43]. For Fisher's exact test, 2 by 2 contingency table was created using two variables, dominant model (Ref (A/A) and Alt (A/a + a/a)) and population information (case and control groups) [43]. Korean abalones were used as the case group, while red abalones were used as the control group, and cutoff p-value < 0.01 was applied.

Results

Summary of sequencing reads mapping and SNVs coverage statistics

The resultant *Pst*I GBS library was sequenced using Illumina technology (NextSeq500), and it generated approximately 204.8 million reads from 102 samples which equated to a total of 17.43 Gbp of sequence data (S1 Table). On average, the dataset contained 1,987,914 reads per sample (SD±849,601.2; median: 1,804,068). After quality filtering, the reads were aligned, using Bowtie2, to the draft genome sequence of *H. discus hannai* which was constructed in our previous study [1], and the average alignment rate was 72.17% (S1 and S3 Figs). Since digestion by RE is not a random process, GBS can generate a non-uniform distribution of sequenced reads thereby producing biased variant information. Therefore, it is important to evaluate whether data produced by the selected RE can represent the whole genomic range with even spacing. Because not every reads harbor SNV sites, we examined the genome coverage by computing distances between neighboring regions covered by sequenced reads across the reference genome, rather than obtaining distances between SNVs. In each sample, approximately 48,488 covered regions were produced on average, and it means that ideally, the regions were expected to be spaced approximately at 38,749 bp intervals throughout the genome, considering the total length of the reference genome was 1,878,915,344 bp. According to our alignment results, the average distance between the regions was ~33,043bp, which was similar to the ideal distance (S2 Fig). Therefore, we concluded that the desired genome coverage could be obtained on the abalone genome with the GBS library generated by *Pst*I.

After correcting possible misalignments and removal of possible PCR duplicates, SNV calling using GATK pipeline initially detected 529,488 SNVs, and stringent filtering steps finally remained 16,603 SNVs for a set of *H. discus* samples (Korea and Japan). The average number of SNVs in each population varied from $3,407 \pm 381.8$ to $4,186 \pm 348.8$; the least number of SNV set was observed in the Yeosu abalone population, while the highest value was found in the Japan abalone population (S1 Table). Transitions are the most common type of nucleotide substitutions, and in our final sets of SNVs, 55% of the base changes were transitions and 45% were transversions, with an observed Ti/Tv ratio of 1.19, 1.19, 1.19, and 1.20 for Goseong, Yeosu, Taean, and Japan, respectively (Table 2). For a set of Korea and Red samples, 23,123 SNVs remained after the filtering steps, and all the detected SNVs were functionally annotated and classified into different types of SNV effects using SnpEff (genetic variant annotation and effect prediction tool) (Table 2). Approximately 65.20% and 3.08% of detected SNVs in all five populations were found in intergenic and intron regions, respectively. These SNVs were classified by SnpEff to have modifier impacts, and the prediction of their effects on phenotype was difficult [44]. 1,201 (5.17%) had impacts in exon regions, and among them, 525 were identified to be non-synonymous SNVs. They were annotated as producing moderate to high functional impacts on gene functions by leading to changes in protein effectiveness via altering the coded amino acid sequence and could be associated with the phenotypic differences among the population.

Table 2. Summary of all detected SNVs identified from 5 abalone populations.

	Goseong	Yeosu	Taeon	Japan	All	Korea	Red	All
Sample counts	23	30	33	5	91	86	11	97
SNV	16,598	16,603	16,583	15,362	16,603	23,123	14,764	23,123
Transition	9,034	9,037	9,026	8,396	9,037	12,470	7,958	12,470
Transversion	7,564	7,566	7,557	6,966	7,566	10,653	6,806	10,653
Ti/Tv	1.1943	1.1944	1.1944	1.2053	1.1944	1.1706	1.16923	1.1706
SNV categories								
Intergenic								
Intergenic region	14,892	14,897	14,880	13,790	14,897	20,546	12,981	20,546
Upstream gene variant	2,214	2,214	2,206	2,070	2,214	3,011	1,946	3,011
Downstream gene variant	2,652	2,652	2,650	2,402	2,652	3,571	2,212	3,571
Intron	715	715	712	657	715	973	567	973
Exon								
Synonymous variant	353	353	353	330	353	677	576	677
Missense variant	315	315	315	292	315	511	388	511
Start lost	1	1	1	1	1	2	1	2
Stop lost	0	0	0	0	0	1	1	1
Stop gained	10	10	10	8	10	11	3	11
Stop retained variant	0	0	0	0	0	1	0	1
Non coding exon variant	1,395	1,394	1,391	1,288	1,394	2,176	1,536	2,176
Splice site acceptor	1	1	1	1	1	1	0	1
Splice site donor	0	0	0	0	0	0	0	0
Splice site region	15	15	15	13	15	30	23	30

Functional annotation and locational classification of SNVs are shown. SNV categories are categorized by the types of SNV effects, and each number represents a total number of variants having the effect type.

<https://doi.org/10.1371/journal.pone.0247815.t002>

Characterization of Korean Pacific abalone populations using SNVs

Several parameters in population genetics were used to characterize the Korean Pacific abalone populations. First, genome-wide measures of nucleotide diversity (π) on a per-site basis were estimated from the SNV data. Average nucleotide diversity (π) of three Korean *H. discus* populations showed a similar level of genetic diversity (Goseong: 9.906×10^{-4} , Yeosu: 1.076×10^{-3} , Taeon: 9.686×10^{-4}), with a relatively low value in Japan population (9.126×10^{-4}). To investigate the relationships among populations, PCA was performed. Eigenvector 1 clearly separated samples into Korea and Japan group, and eigenvector 2 distinguished Goseong samples (Korea) from Yeosu and Taeon samples (Korea) (Fig 2A). To further resolve and separate the individuals from three geographical populations in Korea, additional PCA was performed with those samples. In the resultant scatter plot, the abalone samples of Taeon and Yeosu, located on the west and the south coasts of the Korean peninsula, respectively, formed clusters close to each other, and the Goseong abalones were clearly separated from them (Fig 2B). These findings were consistent with STRUCTURE analysis results. The results revealed the presence of three sub-population clusters in our data set as the highest value of $L(K)$ and ΔK appeared for $K = 3$ (S4 Fig). With $K = 3$, though the genetic structures of all Korean abalones shared some similarities, strong heterogeneity among geographical groups was shown (Fig 3). Among the groups, it was observed that Taeon group exhibited relatively high population homogeneity, with the majority of individuals assigned to a single cluster (cluster 3), in agreement with the fact that Taeon is geographically located in a most isolated region, and Yeosu

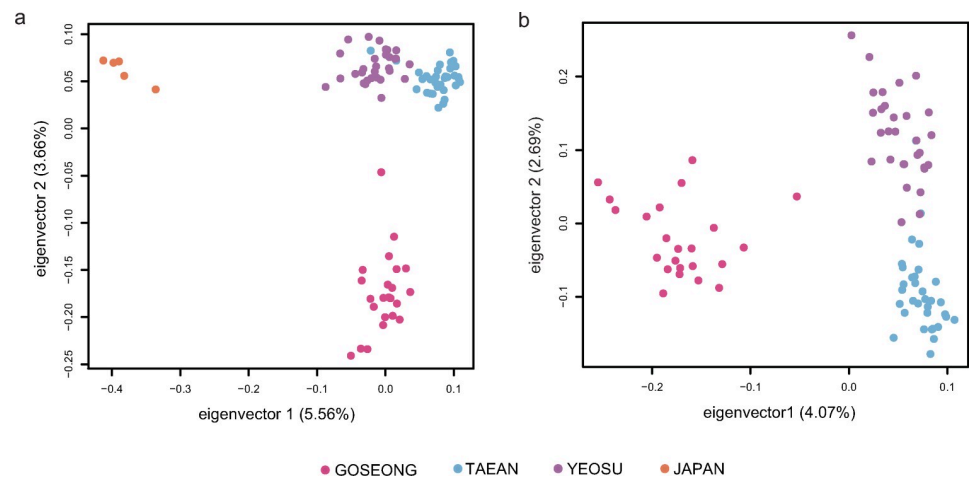


Fig 2. Relationships among Korean *H. discus hannai* populations determined by Principal Component Analysis (PCA). PCA of SNV data for 91 abalone samples. 16,603 SNVs with genotypes were used in this analysis. Each dot represents an individual, and each geographic population is represented by different colors. (A) All 91 individuals (Korea and Japan), and (B) 86 individuals (Korea). Korean populations were clustered separately from Japan samples, and among Korean populations, individuals from the Goseong population were clearly separated from the Tae'an and Yeosu abalones which were clustered close to each other.

<https://doi.org/10.1371/journal.pone.0247815.g002>

and Goseong populations showed admixtures of cluster 1 and 2 and cluster 2 and 3, respectively. Yeosu population exhibited the highest levels of Japan ancestry, reflecting the possible gene flow. On the other hand, the Goseong samples are the sole constituents of cluster 3, consistent with their distinctive genetic structures identified by PCA. Similar patterns of differentiation among the populations were also presented by the average pairwise F_{ST} values. It was identified that the Japan samples were closest to the Yeosu samples (0.1046), followed by Goseong (0.1617) and Tae'an (0.1888). Among Korean individuals, comparison between Yeosu and Tae'an population had lower overall F_{ST} value compared to the Goseong to Yeosu and Goseong to Tae'an pairs (Table 3). Similarly, the results of Mantel test detected a significant and slightly positive correlation ($R_{xy} = 0.215$, $P(R_{xy}\text{-rand} \geq R_{xy}\text{-data}) = 0.001$, $R^2 = 0.046$) between the genetic and geographical distribution (S5 Fig).

Identification of Korean Pacific abalone-specific SNVs

Additionally, in order to search candidate SNVs accounting for phenotypic differences of Korean abalone populations against red abalones, we further extracted fixed non-synonymous SNVs that were significantly enriched, from 23,123 SNVs for 86 Korean abalone samples and 11 red abalone samples, based on SNV annotation and CaseControl analysis. We identified a total of 809 Korean abalone (*H. discus hannai*)-specific SNVs. Among them, 26 SNVs were found within coding regions, from which 23 genes were identified by BLAST search as genes encoding hypothetical protein LOTGIDRAFT_163461, hypothetical protein CGI_10016004, adenosine 3'-phospho 5'-phosphosulfate transporter 1-like isoform X2 exhibited multiple protein mutations (S2 Table). All of the 26 SNVs were missense mutations and include 1 start loss, with no nonsense mutation discovered. We also looked into each SNV on a protein level using InterPro database [45] and investigated whether it caused any physiological and chemical changes (S3 Table). Of the 26 missense mutations, 15 mutations resulted in alteration of chemical properties of amino acids, and five mutations were located in, or in close proximity to, functional domains. Each of the four mutations existed within functional domains were harbored in genes encoding monocarboxylate transporter 12 (major facilitator superfamily domain), zinc finger protein 850-like (zinc finger C_2H_2 -type), carbonic anhydrase (alpha

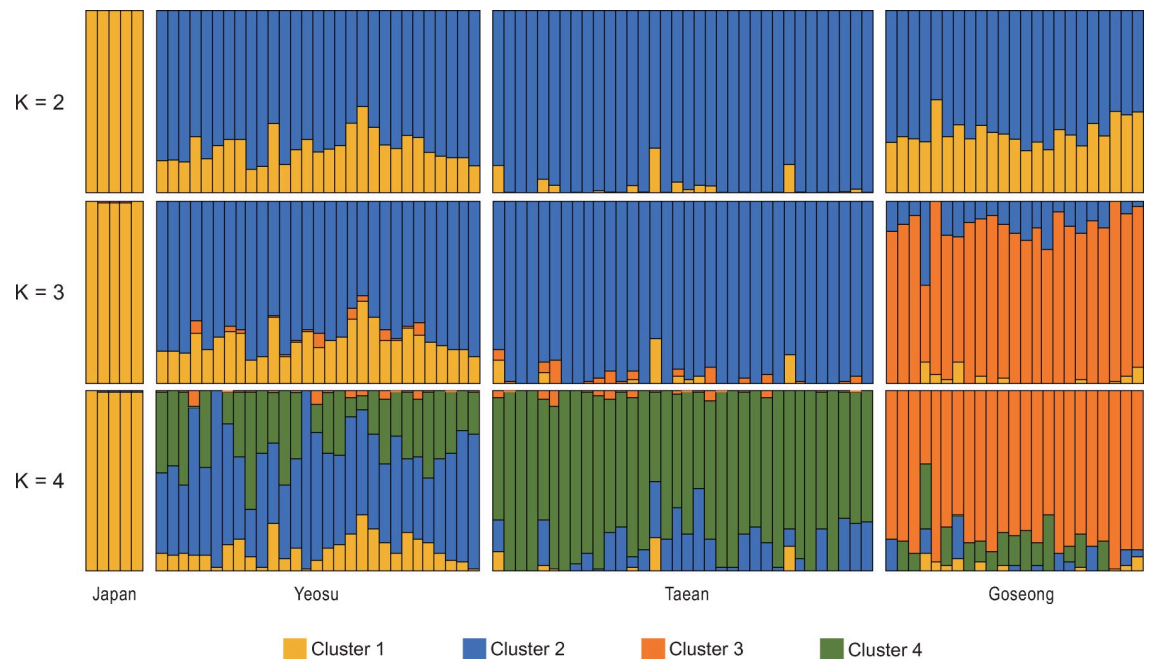


Fig 3. Genetic structures of Korean *H. discus hannai* populations. Population structure using the STRUCTURE analysis. Results for $K = 2$ to 4 are shown. Each vertical line represents an individual, and individuals were grouped by 4 geographic populations. Colors represent the inferred ancestry from K ancestral populations. Populations are labeled below the figure. With a most likely number of $K = 3$, the Taean population exhibited relatively high population homogeneity, and the Yeosu population showed an admixture of cluster 1 and 2. The Goseong population was the most genetically distinct, with a high proportion of cluster 3.

<https://doi.org/10.1371/journal.pone.0247815.g003>

carbonic anhydrase domain), and chitin binding domain-containing protein (chitin binding domain), and some of them were previously reported to be associated with shell formation of abalones [46–48]. One variant found in close proximity to death domain (DED) was harbored in a gene producing caspase-8 which was involved in innate immune responses [49–51]. We also found the start loss variant was located in von Willebrand factor A (vWA) domain-containing protein 7, which was known to be involved in shell formation as well [47, 52–54].

Discussion

Existence of a strong genetic heterogeneity in the Korean Pacific abalones

In contrast to terrestrial species, marine organisms generally show reduced intraspecific genetic differentiation among geographic regions [55–57]. This mainly attributes to the absence of physical barriers to movement in open oceans and their biological characteristics such as large population size and higher dispersal potential during the diverse life-history

Table 3. Matrix of pairwise F_{ST} values among four abalone populations.

	Japan	Yeosu	Taean
Yeosu	0.104555		
Taean	0.188840	0.032859	
Goseong	0.161680	0.063484	0.072502

Pairwise F_{ST} values between each population pair are shown, based on the variant information derived from GBS analysis. The Japan samples were closest to the Yeosu samples, followed by Goseong and Taean. Among Korean individuals, comparison between Yeosu and Taean population had lower overall F_{ST} value compared to the Goseong to Yeosu and Goseong to Taean pairs.

<https://doi.org/10.1371/journal.pone.0247815.t003>

stages (i.e., planktonic egg, larvae, or adult stages) [58–60]. Marine fishes are generally extensive dispersal, resulting in wide-ranging gene flow due to the active migration of adults and passive dispersal of planktonic larvae [61]. However, marine invertebrates differ from them in which adults have a sedentary lifestyle, and gene flows largely dependents on the passive dispersion facilitated by ocean currents during the planktonic larval phase [4, 60, 62, 63]. Like many sessile mollusks, abalones have strictly benthic habit with poor migratory ability after they first settle, so the primary mode of dispersal is a planktonic larval phase which lasts 4–5 days (in *H. discus hannai*) [64–67]. Also, the Korean peninsula is surrounded by three marine ecosystems (East, South, and Yellow sea), each with different biological and oceanographic characteristics [68]. The unique life history of abalones and the characteristic differences of each coastal water around the Korean peninsula render the Pacific abalones as an interesting subject in studying influences of the geographical features and the ocean currents on the population dynamics of marine mollusks.

To assess population differentiation, we investigated patterns of the genetic structure of four regional Pacific abalone populations using 23,123 SNVs, collected from each of the three seas surrounding the Korean peninsula and from coastal waters of Japan. Our results suggested the existence of genetic heterogeneity among Korean groups which was strongly affected by the geographical features of the peninsula and major ocean currents congregating around it. First of all, regression analysis from Mantel test showed that the genetic distance among individuals slightly increased with geographical distance ($R_{xy} = 0.215$, $P(R_{xy\text{-rand}} \geq R_{xy\text{-data}}) = 0.001$, $R^2 = 0.046$), indicating that the geographical distribution contributed partly to the genetic differentiation among the populations (S5 Fig). Also, F_{ST} , PCA, and STRUCTURE analysis revealed high levels of geographical differentiation and relatively admixed patterns within *H. discus hannai* populations around the peninsula compared to Japan samples, and the genetic variation among populations clearly indicated three distinct structures in abalone populations in Korea (Figs 2 and 3). Among them, it was shown that individuals collected from Goseong (East coast), where two different currents meet, were genetically most separated from the other two and were sole constituents of cluster 3 (Figs 1 and 3). Previous studies on *Paralichthys olivaceus* (olive flounder) and *Gadus macrocephalus* (Pacific cod) have presented significant differentiation in the East sea areas as well [69, 70], and a similar phenomenon was also observed in admixed population structure of *Donax deltooides* (saltwater clam) at the point of different currents converge [71]. Similar results were also reported in two preceding studies assessed the population genetic structure of Korean wild Pacific abalones. An et al. [19] found different population structures existed between the east populations and the pooled western and southern populations. Park, C.-J. et al. [4] detected the same pattern of genetic separation of the East Sea populations with the existence of similar but distinguishable structures between the South and the West populations, which was observed in this study, too.

At the Korean peninsula, the Kuroshio Current flowing northeastward splits into the Yellow Sea Warm Current and the East Korea Warm Current [68]. The East Korea Warm Current transports marine organisms and a large quantity of heat to the East Sea [72]. This warm current merges with the North Korea Cold Current flowing southward along the Korean coast [68] (Fig 1). The confluence of these currents forms the subpolar front, generally described as located south of 40° N. Therefore, one possible explanation for the origin of cluster 3 in the Goseong population could be the inflow of abalone larvae inhabiting the northern part of the Korean peninsula, along the North Korea Cold Current. No genetic studies have yet been carried out on the abalones of the North Korea region. However, evidence from a report by the Korea Institute for International Economic Policy stated that North Korea exported abalones to China, indicating that the species was distributed north of Goseong as well [73]. In this work, the Goseong population solely represented the East coast population. Additional

sampling along the coast will be appreciated to determine the origin of the cluster 3 and better understand the influence of the two currents on the genetic structuring of the East coast abalones. The similar genetic structure between the Taean and Yeosu populations suggested relatively active gene flows among these regions. It was suggested in a previous study that a clockwise gyre consists of a portion of warm current branched off the Kuroshio and then flowing into the Yellow Sea and southward inflow of the West Korea Coastal Current was responsible for the similarity between the two localities, which was responsible for the larval transport [19] (Fig 1). Likewise, it may facilitate the Korean Pacific abalones to mix with larvae dispersed from other geographical populations as the ocean currents congregate around the Korean peninsula from various directions. Also, the existence of at least three distinct genetic structures in Korean Pacific abalones proposed in our study suggests the need for separate management strategies to preserve the genetic diversity of the species.

Candidate SNVs involved in phenotypically different traits

There are differences in morphological and physiological characteristics between *H. discus* and *H. rufescens*. The shell length of *H. rufescens* can reach up to 31cm, making it the largest abalone species in the world, and the shell length of *H. discus* varies between 10cm to 15cm [74]. Also, *H. rufescens*'s shell is thicker and stronger, so that it is more resistant to shell-boring ectoparasite infection such as *Polydora* spp. than *H. discus* species [3, 75]. Due to these evident differences, we utilized the SNV sets of pooled Korean *H. discus* populations and *H. rufescens* population to search for candidate SNVs potentially associated with the phenotype differences. For this analysis, we used only fixed non-synonymous SNVs with moderate to high impacts on gene function, and we were able to discover 26 SNVs existing in coding regions which were significantly enriched in the case-control analysis. Some of these SNVs were located in shell formation related genes such as vWA domain-containing protein 7, Kunitz-like protease inhibitor, carbonic anhydrase alpha, and chitin binding domain-containing gene [46–48, 52–54, 76], and for latter two genes, the SNVs were located within functional domains. It also enabled us to test the capability of GBS SNV datasets to detect population-specific SNVs which were potentially implicated with specific traits.

The formation of the molluscan shell is regulated by a matrix of extracellular macromolecules that are secreted by the shell-forming tissue, mantle [76]. This matrix is a complex mixture of CaCO_3 , shell matrix proteins (SMP), pigments, lipids, polysaccharides, and glycoproteins [46, 76]. The molecular mechanisms of this calcifying shell formation are only beginning to be elucidated in Mollusks, but it was reported in many studies that the SMPs play important roles in shell formation and forming distinct morphology [46, 77]. vWA, carbonic anhydrase and chitin-binding domain-containing proteins are well known SMPs in many marine mollusk species. vWA domains in extracellular proteins mediate adhesion via metal ion-dependent adhesion sites, implying its role in protein-protein interaction between layers [47]. They are found in various marine mollusk species, including *Crassostrea gigas*, *Mytilus coruscus*, *Lottia gigantea*, *Pteriida fucata*, etc. [47, 52–54]. Carbonic anhydrases are responsible for controlling pH by converting CO_2 to HCO_3^- in Nacrein [47]. And chitin-binding domain plays significant roles in the formation of biominerals because chitin is the major framework in which CaCO_3 forms. It is usually found together with vWA in SMPs, and both are typically domains of collagen, the fundamental component of ECM [46, 48]. Kunitz-like protease inhibitor was previously observed in the shell matrix of *Haliotis asinina* and also found from the mantle cells of diverse mollusks [76]. The most likely function of these inhibitors was suggested as the protection against degradation by exopeptidases which were produced by marine microorganisms [76] such as tissue inhibitors of metalloproteinase identified in *P. fucata*

species [47]. However, no studies have identified genes directly responsible for molluscan shell size or thickness yet, and the 26 Korean *H. discus hannai*-specific SNVs presented here are the first to be reported in this study. Moreover, since the present study had a limitation in a relatively small sample size of *H. rufescens* population, and thus they were hard to represent the genetic characteristics that covered species' range, future studies including a higher number of samples are needed to further validate these preliminary findings. Also, most population-specific genetic differences are likely to be neutral, and it is above the capability of this study to determine how each SNV functionally affects the gene. Thus, we cautiously suggest these identified SNV sets as promising candidates to guide further investigation.

Conclusions

This study evaluated the genetic diversity and the population structure of *H. discus* populations around the Korean peninsula and elucidated the impacts of the geographical features and the ocean currents in the gene flow among marine mollusk populations. Based on the patterns of population genetic structuring in the F_{ST} , PCA, and STRUCTURE, the ocean currents congregating around the Korean peninsula were key factors of genetic heterogeneity among groups as the directionalities of ocean currents worked as a directional bias in gene flow. Besides, by comparison of variant sets, we identified Korean Pacific abalone-specific variants, which might be associated with phenotypically different traits between groups. Our results will provide valuable data for the genetic conservation and management of wild abalone populations in Korea and help the future GBS studies on the marine mollusks.

Supporting information

S1 Fig. De-multiplexed read count distribution of 102 GBS libraries of Haliotidae population.

(TIF)

S2 Fig. Distribution of distances between regions covered by sequenced reads across the abalone genome.

(TIF)

S3 Fig. Regional distribution of whole genome variants of 102 *Haliotis* samples using SnpEff.

(TIF)

S4 Fig. Diagram of Mean L(K) (\pm SD) and Delta K (Δ K) in STRUCTURE analysis. (a) Mean L(K) (\pm SD) over seven runs from K = 1 to K = 6; (b) Delta K (Δ K); (c) Estimation following Evanno et al.

(TIF)

S5 Fig. Correlation between geographical distances and genetic distances among 86 Korean *H. discus* samples based on Mantel test.

(TIF)

S1 Table. Summary of sequencing results and SNVs from the 102 abalone samples.

(XLSX)

S2 Table. Functional description of genes harboring 26 *H. discus*-specific SNVs identified by BLAST.

(XLSX)

S3 Table. List of 26 *H. discus*-specific SNVs including coding effect, codon/amino acid change, and property change information.

(XLSX)

S1 Appendix. Final set of SNVs after filtering steps in VCF (variant call format) file.

(GZ)

Author Contributions

Conceptualization: Bo-Hye Nam, Woori Kwak.

Data curation: Hyaekang Kim, Donghyeok Seol, Eun Mi Kim.

Formal analysis: Bo-Hye Nam, Hyaekang Kim, Eun Mi Kim.

Funding acquisition: Bo-Hye Nam.

Investigation: Jung Youn Park.

Methodology: Bo-Hye Nam, Eun Soo Noh, Young-Ok Kim, Jung Youn Park.

Project administration: Bo-Hye Nam, Eun Soo Noh, Woori Kwak.

Resources: Eun Mi Kim, Jae Koo Noh.

Software: Heebal Kim.

Supervision: Woori Kwak.

Validation: Donghyeok Seol.

Visualization: Hyaekang Kim, Donghyeok Seol.

Writing – original draft: Bo-Hye Nam, Hyaekang Kim, Donghyeok Seol, Woori Kwak.

Writing – review & editing: Bo-Hye Nam, Heebal Kim, Jae Koo Noh, Young-Ok Kim, Woori Kwak.

References

1. Nam B-H, Kwak W, Kim Y-O, Kim D-G, Kong HJ, Kim W-J, et al. Genome sequence of pacific abalone (*Haliotis discus hannai*): the first draft genome in family Haliotidae. *GigaScience*. 2017; 6(5):gix014. <https://doi.org/10.1093/gigascience/gix014> PMID: 28327967
2. Cook PA, Roy Gordon H. World abalone supply, markets, and pricing. *Journal of Shellfish Research*. 2010; 29(3):569–71.
3. Lafarga-De la Cruz F, Núñez-Acuña G, Gallardo-Escárate C. Hybridization between *Haliotis rufescens* and *Haliotis discus hannai*: evaluation of fertilization, larval development, growth and thermal tolerance. *Aquaculture Research*. 2013; 44(8):1206–20.
4. Park C-J, Hara M, Lee J-H, Noh JK, Kim HC, Park JW, et al. Genetic population structure of the wild Pacific abalone (*Haliotis discus*) in Korea and Japan based on microsatellite DNA markers. *Biochemical systematics & ecology*. 2012; 44:86–95.
5. Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend MJEI. Ecological consequences of genetic diversity. 2008; 11(6):609–23.
6. Carroll EL, Bruford MW, DeWoody JA, Leroy G, Strand A, Waits L, et al. Genetic and genomic monitoring with minimally invasive sampling methods. 2018; 11(7):1094–119.
7. Yuewen D, Xiao L, Guofan Z, Hongen Z. Genetic parameter estimates for growth traits at early stage of Pacific abalone, *Haliotis discus hannai* Ino. *Acta Oceanologica Sinica*. 2007.
8. Li J, Wang M, Fang J, Liu X, Mao Y, Liu G, et al. Reproductive performance of one-year-old Pacific abalone (*Haliotis discus hannai*) and its crossbreeding effect on offspring growth and survival. *Aquaculture*. 2017; 473:110–4.

9. Hara M, Sekino M. Genetic differences between hatchery stocks and natural populations in Pacific Abalone (*Haliotis discus*) estimated using microsatellite DNA markers. *Marine biotechnology*. 2007; 9(1):74–81. <https://doi.org/10.1007/s10126-006-6060-2> PMID: 17294314
10. Park C-J, Nam WS, Lee J-H, Noh JK, Kim HC, Park JW, et al. Analysis of genetic divergence according to each mitochondrial DNA region of *Haliotis discus hannai*. *The Korean Journal of Malacology*. 2013; 29(4):335–41.
11. Hara M, Fujio Y. Geographic distribution of isozyme genes in natural abalone. *Bulletin of Tohoku National Fisheries Research Institute*. 1992.
12. An HS, Lee J-H, Dong CM, Noh JK, Kim HC, Park CJ, et al. New polymorphic microsatellite markers in Pacific abalone *Haliotis discus hannai* and their application to genetic characterization of wild and aquaculture populations. 2010; 32(5):413–8.
13. An HS, Hong SW, Kim EM, Lee J-H, Noh JK, Kim HC, et al. Comparative genetic diversity of wild and released populations of Pacific abalone *Haliotis discus discus* in Jeju, Korea, based on cross-species microsatellite markers including two novel loci. 2010; 14(4):305–13.
14. Schilling MP, Wolf PG, Duffy AM, Rai HS, Rowe CA, Richardson BA, et al. Genotyping-by-sequencing for *Populus* population genomics: an assessment of genome sampling patterns and filtering approaches. *PLoS One*. 2014; 9(4):e95292. <https://doi.org/10.1371/journal.pone.0095292> PMID: 24748384
15. Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular ecology*. 2013; 22(11):2841–7. <https://doi.org/10.1111/mec.12350> PMID: 23711105
16. Waples RS, Gaggiotti O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology*. 2006; 15. <https://doi.org/10.1111/j.1365-294X.2006.02890.x> PMID: 16629801
17. Sunde J, Yıldırım Y, Tibblin P, Forsman AJFig. Comparing the Performance of Microsatellites and RAD-seq in Population Genetic Studies: Analysis of Data for Pike (*Esox lucius*) and a Synthesis of Previous Studies. 2020; 11:218.
18. Gärke C, Ytournal F, Bed'hom B, Gut I, Lathrop M, Weigend S, et al. Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. 2012; 43(4):419–28. <https://doi.org/10.1111/j.1365-2052.2011.02284.x> PMID: 22497629
19. An H, Lee J, Park J. Population genetics of the Pacific abalone (*Haliotis discus hannai*) in Korea inferred from microsatellite marker analysis. *Genet Mol Res*. 2012; 11(4):3904–22. <https://doi.org/10.4238/2012.November.12.8> PMID: 23212330
20. Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, et al. An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS one*. 2013; 8(1):e54603. <https://doi.org/10.1371/journal.pone.0054603> PMID: 23372741
21. Scheben A, Batley J, Edwards DJPbj. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. 2017; 15(2):149–61.
22. Fuentes-Pardo AP, Ruzzante DEJMe. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. 2017; 26(20):5369–406.
23. Nie H, Yan X, Huo Z, Jiang L, Chen P, Liu H, et al. Construction of a high-density genetic map and quantitative trait locus mapping in the manila clam *Ruditapes philippinarum*. *Scientific reports*. 2017; 7(1):229. <https://doi.org/10.1038/s41598-017-00246-0> PMID: 28331182
24. Ashby R, Hess A, Baird H, Roberts R, King N, Symonds J, et al., editors. Genotyping-by-sequencing for the Greenshell™ mussel industry. *Proceedings of the World Congress on Genomics Applied to Live-stock Production*; 2018.
25. Gruenthal K, Witting D, Ford T, Neuman M, Williams J, Pondella D, et al. Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics*. 2014; 15(1):109–21.
26. Sandoval-Castillo J, Robinson NA, Hart AM, Strain LW, Beheregaray LBJME. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. 2018; 27(7):1603–20.
27. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS one*. 2011; 6(5):e19379. <https://doi.org/10.1371/journal.pone.0019379> PMID: 21573248
28. De Donato M, Peters SO, Mitchell SE, Hussain T, Imumorin IG. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. *PLoS One*. 2013; 8(5):e62137. <https://doi.org/10.1371/journal.pone.0062137> PMID: 23690931

29. Herten K, Hestand MS, Vermeesch JR, Van Houdt JK. GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC bioinformatics*. 2015; 16(1):73. <https://doi.org/10.1186/s12859-015-0514-3> PMID: 25887893
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9(4):357. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
32. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
33. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011; 27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522
34. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*. 2012; 28(24):3326–8. <https://doi.org/10.1093/bioinformatics/bts606> PMID: 23060615
35. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2011; 28(2):298–9. <https://doi.org/10.1093/bioinformatics/btr642> PMID: 22110245
36. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources*. 2009; 9(5):1322–32. <https://doi.org/10.1111/j.1755-0998.2009.02591.x> PMID: 21564903
37. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. 2005; 14(8):2611–20.
38. Earl DA, J. C. Gr. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. 2012; 4(2):359–61.
39. Korunes KL, Samuk K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. 2020.
40. PE PRSJB. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. 2012; 28(19):2537.
41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. 2015; 4(1):s13742–015-0047-8.
42. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*. 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672
43. Ruden DM, Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, et al. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Frontiers in genetics*. 2012; 3:35. <https://doi.org/10.3389/fgene.2012.00035> PMID: 22435069
44. Khalkhali-Evrigh R, Hafezian SH, Hedayat-Evrigh N, Farhadi A, Bakhtiarzadeh MR. Genetic variants analysis of three dromedary camels using whole genome sequencing data. *PloS one*. 2018; 13(9): e0204028. <https://doi.org/10.1371/journal.pone.0204028> PMID: 30235280
45. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research*. 2018; 47(D1): D351–D60.
46. Feng D, Li Q, Yu H, Kong L, Du S. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Scientific reports*. 2017; 7:45754. <https://doi.org/10.1038/srep45754> PMID: 28374770
47. Liu C, Li S, Kong J, Liu Y, Wang T, Xie L, et al. In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Scientific reports*. 2015; 5:17269. <https://doi.org/10.1038/srep17269> PMID: 26608573
48. Addadi L, Joester D, Nudelman F, Weiner S. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. *Chemistry—A European Journal*. 2006; 12(4):980–7. <https://doi.org/10.1002/chem.200500980> PMID: 16315200
49. Lee Y, De Zoysa M, Whang I, Lee S, Kim Y, Oh C, et al. Molluscan death effector domain (DED)-containing caspase-8 gene from disk abalone (*Haliotis discus discus*): molecular characterization and expression analysis. *Fish & Shellfish Immunology*. 2011; 30(2):480–7. <https://doi.org/10.1016/j.fsi.2010.11.014> PMID: 21130887

50. Fan T-J, Han L-H, Cong R-S, Liang J. Caspase family proteases and apoptosis. *Acta biochimica et biophysica Sinica*. 2005; 37(11):719–27. <https://doi.org/10.1111/j.1745-7270.2005.00108.x> PMID: 16270150
51. Valmiki MG, Ramos J. Death effector domain-containing proteins. *Cellular Molecular Life Sciences*. 2009; 66(5):814–30. <https://doi.org/10.1007/s00018-008-8489-0> PMID: 18989622
52. Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cléon I, Guichard N, Marin F. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *The FEBS journal*. 2013; 280(1):214–32. <https://doi.org/10.1111/febs.12062> PMID: 23145877
53. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*. 2012; 490(7418):49. <https://doi.org/10.1038/nature11413> PMID: 22992520
54. Liao Z, Bao L-f, Fan M-h, Gao P, Wang X-x, Qin C-l, et al. In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *Journal of proteomics*. 2015; 122:26–40. <https://doi.org/10.1016/j.jprot.2015.03.027> PMID: 25857279
55. Castro A, Stewart B, Wilson S, Hueter R, Meekan M, Motta P, et al. Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*. 2007; 16(24):5183–92. <https://doi.org/10.1111/j.1365-294X.2007.03597.x> PMID: 18092992
56. Theisen T, Bowen B, Lanier W, Baldwin J. High connectivity on a global scale in the pelagic wahoo, *Acanthocybium solandri* (tuna family Scombridae). *Molecular Ecology*. 2008; 17(19):4233–47. <https://doi.org/10.1111/j.1365-294x.2008.03913.x> PMID: 19378403
57. Palumbi SR. Marine speciation on a small planet. *Trends in Ecology & Evolution*. 1992; 7(4):114–8. [https://doi.org/10.1016/0169-5347\(92\)90144-Z](https://doi.org/10.1016/0169-5347(92)90144-Z) PMID: 21235975
58. Zhan A, Hu J, Hu X, Zhou Z, Hui M, Wang S, et al. Fine-scale population genetic structure of Zhikong scallop (*Chlamys farreri*): do local marine currents drive geographical differentiation? *Marine Biotechnology*. 2009; 11(2):223–35. <https://doi.org/10.1007/s10126-008-9138-1> PMID: 18766401
59. Kenchington E, Patwary M, Zouros E, Bird C. Genetic differentiation in relation to marine landscape in a broadcast-spawning bivalve mollusc (*Placcopecten magellanicus*). *Molecular Ecology*. 2006; 15(7):1781–96. <https://doi.org/10.1111/j.1365-294X.2006.02915.x> PMID: 16689898
60. Baus E, Darrock D, Bruford M. Gene-flow patterns in Atlantic and Mediterranean populations of the Lusitanian sea star *Asterina gibbosa*. *Molecular Ecology*. 2005; 14(11):3373–82. <https://doi.org/10.1111/j.1365-294X.2005.02681.x> PMID: 16156809
61. Henriques R, Potts WM, Santos CV, Sauer WH, Shaw PW. Population connectivity and phylogeography of a coastal fish, *Atractoscion aequidens* (Sciaenidae), across the Benguela current region: evidence of an ancient vicariant event. *PloS one*. 2014; 9(2):e87907. <https://doi.org/10.1371/journal.pone.0087907> PMID: 24586296
62. Azuma N, Kunihiro Y, Sasaki J, Mihara E, Mihara Y, Yasunaga T, et al. Genetic variation and population structure of hair crab (*Erimacrus isenbeckii*) in Japan inferred from mitochondrial DNA sequence analysis. *Marine Biotechnology*. 2008; 10(1):39–48. <https://doi.org/10.1007/s10126-007-9033-1> PMID: 17955293
63. Waters J, Roy M. Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? *Molecular Ecology*. 2004; 13(9):2797–806. <https://doi.org/10.1111/j.1365-294X.2004.02282.x> PMID: 15315690
64. Ward R, Elliott N. Genetic population structure of species in the South East Fishery of Australia. *Marine Freshwater Research*. 2001; 52(4):563–73.
65. Ayre D, Minchinton T, Perrin C. Does life history predict past and current connectivity for rocky intertidal invertebrates across a marine biogeographic barrier? *Molecular ecology*. 2009; 18(9):1887–903. <https://doi.org/10.1111/j.1365-294x.2009.04127.x> PMID: 19434808
66. Hsu T-H, Gwo J-C. Genetic diversity and stock identification of small abalone (*Haliotis diversicolor*) in Taiwan and Japan. *PloS one*. 2017; 12(6):e0179818. <https://doi.org/10.1371/journal.pone.0179818> PMID: 28662122
67. Sekino M, Saido T, Fujita T, Kobayashi T, Takami H. Microsatellite DNA markers of Ezo abalone (*Haliotis discus hannai*): a preliminary assessment of natural populations sampled from heavily stocked areas. *Aquaculture*. 2005; 243(1–4):33–47.
68. Rebstock GA, Kang YS. A comparison of three marine ecosystems surrounding the Korean peninsula: responses to climate change. *Progress in Oceanography*. 2003; 59(4):357–79.
69. Kim WJ, Kim KK, Han HS, Nam BH, Kim YO, Kong HJ, et al. Population structure of the olive flounder (*Paralichthys olivaceus*) in Korea inferred from microsatellite marker analysis. *Journal of fish biology*. 2010; 76(8):1958–71. <https://doi.org/10.1111/j.1095-8649.2010.02638.x> PMID: 20557649

70. Gwak WS, Nakayama K. Genetic variation and population structure of the Pacific cod *Gadus macrocephalus* in Korean waters revealed by mtDNA and msDNA markers. *Fisheries Science*. 2011; 77(6):945–52.
71. Miller AD, Versace VL, Matthews TG, Montgomery S, Bowie KC. Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia—implications for predicting the movement of passive dispersers across a marine biogeographic barrier. *Ecology evolution*. 2013; 3(5):1248–61. <https://doi.org/10.1002/ece3.535> PMID: 23762511
72. Kitamura A, Takano O, Takata H, Omote H. Late Pliocene—early Pleistocene paleoceanographic evolution of the Sea of Japan. *Palaeogeography, Palaeoclimatology, Palaeoecology*. 2001; 172(1–2):81–98.
73. Choi J. Evaluation of trade between North Korea and China in 2017 and the future prospect. 2018.
74. Mardones A, Augsburger A, Vega R, de Los Ríos-Escalante P. Growth rates of *Haliotis rufescens* and *Haliotis discus hannai* in tank culture systems in southern Chile (41.5 oS). *Latin American Journal of Aquatic Research*. 2013; 41(5).
75. González R, Lohrmann KB, Pizarro J, Brokordt K. Differential susceptibility to the Withering Syndrome agent and renal coccidia in juvenile *Haliotis rufescens*, *Haliotis discus hannai* and the interspecific hybrid. *Journal of invertebrate pathology*. 2014; 116:13–7. <https://doi.org/10.1016/j.jip.2013.12.002> PMID: 24333918
76. Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C, et al. Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Proteome Science*. 2010; 8(1):54. <https://doi.org/10.1186/1477-5956-8-54> PMID: 21050442
77. Falini G, Albeck S, Weiner S, Addadi L. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. *Science*. 1996; 271(5245):67–9.