

# The Genomic Basis of the Genetic Differentiation and Local Adaptive Evolution of *Pampus Echinogaster* based on SLAF-seq

**Yuan Li**

Third Institute of Oceanography, Ministry of Natural Resources

**Fangrui Lou**

School of Ocean, Yantai University

**Hai Li**

Third Institute of Oceanography, Ministry of Natural Resources

**Rui Wang**

Third Institute of Oceanography, Ministry of Natural Resources

**Zizi Cai**

Third Institute of Oceanography, Ministry of Natural Resources

**Ran Zhang**

Third Institute of Oceanography, Ministry of Natural Resources

**Youzhu Weng**

Third Institute of Oceanography, Ministry of Natural Resources

**Longshan Lin** (✉ [linlongshan1974@163.com](mailto:linlongshan1974@163.com))

Third Institute of Oceanography, Ministry of Natural Resources

**Xuan Zhuang**

Department of Biological Sciences, University of Arkansas

---

## Research Article

**Keywords:** *Pampus echinogaster*, specific-length amplified fragment sequencing (SLAF-seq), genomic variation, population-genetic differentiation, adaptive evolution

**Posted Date:** August 26th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-640424/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** Factors such as climate change (especially ocean warming) and overfishing have led to a decline in the supply of *Pampus echinogaster* and a trend of decreasing age. Exploring the genetic structure and local adaptive evolutionary mechanisms is crucial for the management of *P. echinogaster*.

**Results:** This population genomic study of nine geographical populations of *P. echinogaster* in China was conducted by specific-locus amplified fragment sequencing (SLAF-seq). A total of 935,215 SLAF tags were obtained, and the average sequencing depth of the SLAF tags was 20.80×. After filtering, a total of 46,187 high-consistency genome-wide single nucleotide polymorphisms (SNPs) were detected. Based on all SNPs, the overall genetic diversity among the nine *P. echinogaster* populations was high. The Shantou population had the lowest genetic diversity, and the Tianjin population had the highest. Meanwhile, the population genetic structure based on all SNPs revealed significant gene exchange and insignificant genetic differentiation between the nine *P. echinogaster* populations. Based on pairwise genetic differentiation ( $F_{ST}$ ), we further screened 1,852 outlier SNPs that might have been affected by habitat selection and annotated SLAF tags containing these 1,852 outlier SNPs using Blast2GO. The annotation results showed that the genomic sequences at the outlier SNPs were mainly related to material metabolism, ion transport, breeding, stress response, and inflammatory reactions, which may be related to the adaptation of *P. echinogaster* to different environmental conditions (such as water temperature and salinity) in different sea areas.

**Conclusions:** The high genetic similarity of nine *P. echinogaster* populations may have been caused by the population expansion after the last glacial period, the lack of balance between migration and genetic drift, and the long-distance diffusion of eggs and larvae. We suspected that variation of these genes associated with material metabolism, ion transfer, breeding, stress reactions, and inflammatory reactions were critical for adaptation to spatially heterogeneous temperatures in natural *P. echinogaster* populations.

## Background

Widely distributed organisms inevitably face diverse habitats, and due to habitat heterogeneity-induced differences, natural selection can improve the habitat adaptation and utilization efficiency of geographically separated populations by changing their phenotypic and genomic characteristics. Ultimately, different geographical groups can undergo adaptive differentiation, even leading to new species [1, 2]. The effective population size ( $N_e$ ) of marine organisms is usually large, and seawater can flow with minimal physical barriers, which may result in a high level of gene exchange between marine biogeographical populations and make them less susceptible to genetic drift, generally limiting their genetic differentiation [3–5]. The different life history characteristics (such as spawning type, length of the pelagic larval period, and migration mode) of marine biogeographical populations and long-term heterogeneity in environmental factors can promote adaptive differentiation of populations in different habitats [6]. At the same time, overfishing, habitat loss, and climate change have reduced some population sizes, leading to losses of genetic diversity. The resultant genetic drift and inbreeding may further reduce population sizes, thus affecting the adaptation of populations to their habitats [7, 8]. Therefore, understanding the genetic mechanism of adaptive differentiation related to habitat heterogeneity in different populations can not only reveal the evolutionary history of species but also effectively define protection units in the context of climate change in order to achieve the rational management and protection of resources.

*Pampus echinogaster* belongs to Stromateidae, Perciformes, and is widely distributed in the northwestern Pacific Ocean, including the waters of Russia, Japan, the Korean Peninsula, China's Bohai Sea, the Yellow Sea, the East China Sea, and the northern South China Sea [9–13]. The complex paleogeographic dynamics of the northwestern Pacific Ocean may affect the formation of different geographical populations of marine organisms, including *P. echinogaster*, through events such as geographical isolation caused by glaciers and postglacial recolonization, while the hydrological conditions and environmental specificity of different habitats may further cause habitat-based adaptive differentiation of different geographical populations of marine organisms [14].

*P. echinogaster* is a seasonal migratory fish species. Its inner spawning cycle of the gonad development cycle is short (about 15 days) during somewhat long reproductive periods [11], and it can produce buoyant eggs with a pelagic larval period of at least one month [15]. Thus, it can spread over long distances. The complex oceanic current system in the northwestern Pacific Ocean results in a mix of recruited populations and overwintering populations of juvenile *P. echinogaster*, and these factors may eventually shape the genetic homogeneity between different geographical populations of this species [16]. In summary, the complex life history characteristics of this species have made it difficult to study its population differentiation. In recent years, factors such as climate change (especially ocean warming) and overfishing have led to a decline in the supply of *P. echinogaster* and a trend of decreasing age [16]. In this context, it is necessary to evaluate the population differentiation of *P. echinogaster* in order to achieve the rational management of this resource.

In previous studies, researchers mainly divided *P. echinogaster* into the Yellow Sea population, Bohai Sea population, and East China Sea population based on their seasonal migratory characteristics [17] found the same population structure of *P. echinogaster* based on microsatellite molecular markers. However, based on mitochondrial control region sequences and six pairs of microsatellite loci, the results of our previous study were different from the population division of traditional fisheries; that is, we found that the seven *P. echinogaster* populations in the waters of the Yellow Sea, the Bohai Sea, and the East China Sea may belong to one free-mating group, and there was no significant genetic differentiation between the populations [16]. These discrepant research results could be explained by the lack of methods for genome-wide genotyping. In addition, since they have been limited to neutral markers, few studies have analyzed the habitat adaptation mechanisms of *P. echinogaster* populations. However, it is undeniable that environmental factors such as temperature, salinity, and pathogens vary greatly across the widely distributed range of *P. echinogaster*, which must have led to adaptive habitat differentiation between different geographical populations [18, 19].

In this context, genome sequencing by high-throughput sequencing technologies can increase the number of genetic markers, including both neutral and adaptive markers, to help achieve the fine assessment of population-genetic parameters in different *P. echinogaster* populations and to clarify the genetic mechanisms of its habitat adaptation. For specific-locus amplified fragment sequencing (SLAF-seq), the length of an effective read is  $2 \times 100$  bp, and more than 100,000 tags can be developed at one time, so genome-wide scans of genetic markers can be completed for almost any species, contributing to a better understanding of the genetic structure of the species and its population-genomic characteristics under habitat selection [20].

In this study, SLAF-seq was used for the first time to scan the genome-wide genetic markers of nine *P. echinogaster* populations off the coast of China. We used a population-genomic method to quantify the genetic variation in *P. echinogaster* populations and revealed the genetic differentiation between them. Then we predicted the sizes of different geographical *P. echinogaster* populations under climate change and human activity. Finally, we determined the regulatory mechanism related to the habitat adaptation of geographical *P. echinogaster* populations. In short, the results of this study could improve our understanding of the population structure and habitat-based adaptive differentiation of *P. echinogaster* and provide basic information for the accurate determination of its protection units. These contributions will be of great significance for maintaining the sustainability of *P. echinogaster* resources against the background of climate change and human activity.

## Results

### SLAF-seq results

After filtering, a total of 448.17 million high-quality paired-end reads were obtained from SLAF-seq of 135 *P. echinogaster* samples. The average Q30 of the reads was 95.59%, and the average GC content was 43.17%. In addition, for the control sequence of Nipponbare, which was used to evaluate the accuracy of the established library, 1.79 M reads were obtained. After clustering high-quality reads, we obtained a total of 935,215 SLAF tags, and the average sequencing depth of the SLAF

tags was 20.80×. Among the SLAF tags, 461,932 were polymorphic. With GATK and SAMTOOLS, 6,152,196 SNPs were further obtained. After the removal of low-quality SNPs, we obtained 46,187 highly consistent SNPs for subsequent genetic structure analysis.

## Population genetic diversity and population structure

The  $\pi$  and Tajima's  $D$  values of the 46,187 SNPs exhibited similar fluctuating patterns among the nine *P. echinogaster* populations (Fig. 2). The Tajima's  $D$  values of most SNPs in the nine *P. echinogaster* populations were less than 0, indicating that many rare alleles existed at a high frequency, which may have been due to the large population size.

Statistical results of genetic diversity at the 46,187 SNPs showed that the average  $H_0$ , average  $H_E$ , and  $\pi$  of the nine *P. echinogaster* populations were different ( $H_0 = 0.20470-0.23647$ ,  $H_E = 0.22220-0.23379$ , and  $\pi=0.23360-0.24251$ ), and the  $F_{IS}$  values of all nine populations were small ( $\pi=0.02228-0.08049$ ) (Table. 1). Among the nine populations, the Shantou population had the lowest genetic diversity, as it had the lowest average  $H_0$ , average  $H_E$ , and  $\pi$ . These metrics were the highest in the Tianjin population, meaning that the genetic diversity level in this population was the highest. The statistical results for genetic diversity also showed that the percentage of polymorphic SNPs in each of the nine *P. echinogaster* populations was relatively high (86.52-93.28%), suggesting that the relatively large effective population sizes caused this result. The nine *P. echinogaster* populations had significantly low values of genetic differentiation, ranging from -0.00121 to 0.00125 (Table. 2).

ADMIXTURE analysis of 46,187 SNPs revealed that an optimum  $K$  value of 2 with a minimum CV error. The clustering results from ADMIXTURE based on  $K=2$  showed that all the individuals in the nine populations were clustered into one group, and the clustering results from ADMIXTURE based on  $K = 3$  to 7 showed a similar result (Fig. 3A). The clustering pattern from ADMIXTURE was validated by analysis of NJ trees (Fig. 3B) and the PCA results (Fig. 3C) for all 46,187 SNPs. In addition, NetView P software revealed fine genetic structure among the nine *P. echinogaster* populations. In this study, the optimal KNN value was determined to be 38 based on various algorithms, and the network topological results showed that all the individuals in all nine populations were clustered together (Fig. 3D). The AMOVA results (Table. 3) also showed that the mean value representing the genetic differentiation of the nine populations was 0.03753, which was statistically significant, while statistically nonsignificant genetic differentiation existed between the three groups ( $F_{CT} = 0.00020$ ).

## Gene exchange between the nine *P. echinogaster* populations

divMigrate-Online software was used to analyze the gene exchange between the nine *P. echinogaster* populations, and the results showed that there was high-intensity gene exchange between them (Fig. 4), providing good support for the above results on genetic differentiation.

## Analysis of local adaptation

After running the Lositan program five times, a total of 1,852 outlier SNPs were screened from the 46,187 SNPs (Fig. 5), while only four outlier SNPs were screened by BayeScan software. Thus, these two  $F_{ST}$ -based methods successfully uncovered 1,852 outlier SNPs, which were used for subsequent analysis of the local adaptation of *P. echinogaster*.

Blast2GO was used to annotate the SLAF tags containing the 1,852 outlier SNPs. A total of 604 SLAF tags containing outlier SNPs were successfully annotated. GO functional classification was further applied using Blast2GO, in which 453 SLAF tags containing outlier SNPs could be assigned to 30 GO terms. The successfully annotated SLAF tags were related to multiple biological processes, molecular functions, and cellular components (Fig. 6). The biological processes mainly included cell processes, metabolic processes, and biological regulatory processes. The molecular functions mainly included

molecular binding, catalytic activity, and transport activity. The cellular components mainly included cellular components, cells, and organelles. Moreover, 194 SLAF tags containing outlier SNPs were assigned to 249 metabolic pathways (Table S1). These pathways were primarily related to material metabolism, ion transport, breeding, stress reactions, and inflammatory reactions.

## Discussion

Compared with traditional molecular markers, SNPs can be typed across the whole genome with more simplified genome sequencing technology, and they can reveal more sophisticated genetic information in evolutionary biology, especially for species with insignificant genetic differences. Previous studies on the genetic structure of *P. echinogaster* populations were mainly based on ecological characteristics [35, 36], mitochondrial DNA sequences [16], or small numbers of microsatellite loci [16, 17]. The development of population genetics research in *P. echinogaster* was restricted by the limited number of genetic markers, and therefore studies on the genetic structure of some populations have yielded different results. It is difficult to detect adaptive characteristics related to the habitats of *P. echinogaster* populations by traditional molecular markers [37]. Therefore, in this study, we used genome-wide SNPs obtained from SLAF-seq to analyze and evaluate the genetic structure and habitat adaptation characteristics in regionally represented samples of *P. echinogaster* we collected from coastal waters of China.

## Gene flow between the nine *P. echinogaster* populations

This is the first study exploring the genetic structure and the genetic diversity of different geographical populations of *P. echinogaster* at the genomic level. The results showed that in the nine *P. echinogaster* populations,  $H_D$ ,  $H_E$  and  $\pi$  were relatively high, and the average  $F_{IS}$  was relatively low. We speculate that the rapid population expansion of this species in the Pleistocene and the heterogeneous habitats in its wide distribution provided a basis for the maintenance of high genetic diversity in natural populations. Previous research has also shown that in coastal waters of China, the existing resources and effective female population sizes of *P. echinogaster* are large and stable, the male:female ratio in the spawning period is 1:1, and the species exhibits fast growth, a short gonadal development cycle, and batch spawning. These life history characteristics can bolster the size of recruit populations, thus facilitating the accumulation of more genetic mutations and rich genetic diversity between the populations.

Although *P. echinogaster* occupies a wide variety of habitats, this study confirmed that there was no significant genetic differentiation between *P. echinogaster* populations based on multiple analytical methods. Gene flow analysis also showed frequent gene exchange between *P. echinogaster* populations, which can be explained by random mating and suggests that high genetic homogeneity exists between *P. echinogaster* populations. These results are consistent with our previous results based on mitochondrial DNA sequences and microsatellite loci [16]. We speculate that population expansion of this species after the last glacial period and lacking balance between migration and genetic drift may be two main reasons for their unclear genetic structure [16, 38]. Similar genetic structure is also found in species with closed expansion times, such as the horse mackerel [39], *Sebastes schlegeli* [40], the spotted-tail goby *Synechogobius ommaturus* [41], and the small yellow croaker [42, 43]. At the same time, the open marine environment lacking obvious physical barriers; the strong diffusion ability of eggs, larvae, and adults; and the high randomness of diffusivity provided by ocean currents are key reasons for the genetic homogeneity of marine organisms [44]. *P. echinogaster* is a seasonal migratory fish species that lays eggs in batches. The spawning period can last for more than two months, and the number of eggs per mother can be 117,000-218,000 [35]. In addition, the eggs of *P. echinogaster* are buoyant, and the larvae have a pelagic period of at least 1 month. These life history characteristics may be conducive to the long-distance diffusion of eggs and larvae and can promote gene exchange between *P. echinogaster* populations, ultimately resulting in very low population genetic differentiation within a very wide distribution range [45]. In addition, we speculate that the complex ocean current system along the coast of China may further affect the population size and genetic connectivity of this species by increasing the diffusion of eggs and

larvae. Currently, *P. echinogaster* populations form a continuous distribution and present a population-wide genetic pattern [16]. In summary, the life history characteristics of *P. echinogaster* and the influence of marine environmental factors may cause the mixing of *P. echinogaster* populations, thus resulting in frequent gene exchange and high genetic homogeneity between them.

## Local adaptation of *P. echinogaster* populations

The annotation of SLAF tags containing outlier SNPs indicated the local adaptation of different *P. echinogaster* populations. *P. echinogaster* is a seasonal migratory fish with long-distance diffusion [15, 16], so aquatic environmental conditions (such as salinity and temperature) vary among geographical populations of this species. The GO annotation results showed that outlier SNPs mainly participate in metabolic processes and cellular processes, and their functions mainly involve protein binding and catalytic activity. These results suggest that driven by certain environmental factors unique to each *P. echinogaster* population, some regions of the genome corresponding to adaptive differentiation exist, leading to differences in physiological function. The KEGG results showed that multiple outlier SNPs were related to material metabolism. Fish use substances (amino acids, fat, and sugar) stored in their bodies to adapt to changes in habitat factors and for the energy required for life activities, such as migration [46–48]. On the other hand, fish also regulate the fatty acid composition of their bodies to increase membrane fluidity, thereby adapting to changes in habitat temperature [13]. Therefore, genes related to material metabolism may play an important role in the adaptation of *P. echinogaster* to the temperatures of different geographical environments. Ion (such as calcium) transport signaling pathways also play important roles in the process of local adaptive evolution in *P. echinogaster*. For example, calcium ions are involved in fish reproduction, development, learning and memory, mitochondrial function, muscle contraction, and other functions [49]. Calcium signaling is also thought to play an important role in the regulation of ion exchange and osmoregulation, which makes this pathway a reasonable target of spatially differentiated selection under osmotic pressure in *P. echinogaster* because this species feeds on a wide variety of prey. Differences in environmental factors, such as temperature and salinity, may also lead to different breeding times in different *P. echinogaster* populations. In this study, the oocyte meiosis pathway was enriched at the molecular level, also providing evidence of this phenomenon. Gene mutations related to inflammatory reactions may provide evidence for the specific resistance of different geographical groups to different habitats. In fact, a correlation between the immune response and the adaptive evolution to habitat temperature has been confirmed in populations of many other marine organisms, such as *Syngnathus scovelli* [50], *Haliotis laevigata* [19], and *Trachidermus fasciatus* [13]. It is undeniable that local adaptation of *P. echinogaster* populations may be affected by many factors, such as water quality, heavy metals, water temperature, salinity, parasites, and predators. The mechanism of local adaptation may be very complex, and various factors probably interact to drive this process.

## Conclusion

In this study, genome-wide information was obtained from nine *P. echinogaster* populations, and the genetic structure of the populations and the genomic characteristics of local adaptation were explored. Genetic structure analysis showed that the nine populations had high genetic similarity, which may be due to population expansion after the last glacial period, the lack of balance between migration and genetic drift, and the long-distance diffusion of eggs and larvae. Different habitat conditions might have caused the maintenance of many genetic mutations in the different populations. The annotation of SLAF tags containing outlier SNPs indicated that genes related to material metabolism, ion transfer, breeding, stress reactions, and inflammatory reactions were essential to the habitat adaptation of *P. echinogaster* populations.

## Methods

### Sample collection and SLAF-seq

To ensure the accuracy of the sample source, a total of 135 *P. echinogaster* samples (Fig. 1) were collected at nine different locations along the coast of China (15 individuals at each location) from October 2017 to December 2017. External morphological identification of the samples was performed mainly by referring to Nakabo [12] and Li et al. [13]. For each fresh *P. echinogaster* sample, sterile scissors and tweezers were used to obtain back muscle tissues. All muscle tissues were stored in a cryogenic freezer at -80°C for future experiments. The phenol–chloroform method was used to extract the genomic DNA from each sample. Then, 1% agarose gel electrophoresis and an Invitrogen Qubit fluorometer were used to assess the degradation degree and concentration of the genomic DNA. Qualified DNA was submitted to Biomarker Technologies Corporation (Beijing, China) for library construction and sequencing.

According to the genome size and guanine+cytosine (G+C) content of *P. argenteus*, enzymatic digestion prediction was performed. The analysis software for SLAF enzymatic digestion prediction independently developed by Biomarker Technologies Corporation [20] was used to predict the digestion of the reference genome, and the optimal digestion scheme was selected according to the following principles: (1) the percentage of enzymatic fragments located in the repeat sequences was as low as possible; (2) the enzymatic fragments were distributed as evenly as possible across the genome; (3) the length of each enzymatic fragment was consistent with that of the specific experimental system; and (4) the number of obtained enzymatic fragments (SLAF tags) agreed with the expected number of tags. With the *Hae*III restriction enzyme, the genomic DNA of each qualified sample was digested separately. Each obtained digestion fragment (SLAF tag) had a poly (A) tag added to the 3' end, the dual-index sequencing adaptors were attached [21], and this new sequence was amplified by polymerase chain reaction (PCR), purified, and mixed. The target fragments were taken out by cutting the gel. After the samples were qualified by the library, they were sequenced using the Illumina HiSeq 2500 platform. To evaluate the accuracy of enzyme digestion, *Nipponbare* was selected as the control for sequencing.

## SNP detection and screening

The raw data obtained from sequencing were characterized using the dual-index sequencing adaptor. The raw reads of each sample were obtained, and reads with a quality score lower than 30 were excluded. Based on sequence similarity, the remaining high-quality reads of each sample were clustered, and the reads with a similarity greater than 98% were considered to have clustered as a single SLAF tag. A SLAF tag with sequence differences between samples can be defined as a polymorphic SLAF tag. The sequence with the greatest sequencing depth for each SLAF tag was taken as the reference sequence. The Burrows-Wheeler alignment tool was used [22] to align the reads to the SLAF tags, both GATK [23] and SAMTOOLS [24] were used to perform single nucleotide polymorphism (SNP) calling, and the overlapping SNPs in the GATK and SAMTOOLS results were used as the final SNP dataset. The generated SNPs were saved in a variant call format (VCF) file. To ensure the accuracy of subsequent analyses, VCFtools [25] was used to screen SNPs with the following parameters: -MAF 0.01 (minimum allele frequency > 0.01); -max-missing 0.1 (filtering out the genotypes with less than 90% data); -min-meanDP 150 (minimum mean depth of coverage > 150); -min-alleles 2 -max-alleles 2 (only two alleles); -minGQ 98 (quality score > 98); -minQ 30 (retaining the loci with a quality score > 3); -remove-indels (excluding the loci containing indels); and -HWE 0.05 (excluding the loci with  $P < 0.05$  in the Hardy-Weinberg equilibrium test).

## Genetic diversity and population genetic structure

To describe the genetic diversity levels of all genome-wide SNPs of the nine *P. echinogaster* populations, with TASSEL software (version 5.2.31) [26], the nucleotide diversity ( $\pi$ ) and Tajima's  $D$  value of each SNP in each *P. echinogaster* population were estimated. With Circos software [27], the  $\pi$ , expected heterozygosity ( $H_E$ ), and Tajima's  $D$  of each SNP were visualized.

Using the “populations” module of Stacks software (version 1.34) [28], the genetic diversity levels of the nine populations, including the polymorphic loci,  $\pi$ , observed heterozygosity ( $H_O$ ),  $H_E$ , and inbreeding coefficient ( $F_{IS}$ ), were statistically

analyzed. Arlequin software (version 3.5.2.2) [29] was used to estimate the pairwise genetic differentiation ( $F_{ST}$ ) of *P. echinogaster*, and 10,000 permutations were used to analyze the significance of  $F_{ST}$ .

Based on all SNPs, four methods were used to estimate population structure and individual clustering within a population. (1) PGDSpider software (version 2.0.5.2) [30] was used to convert the VCF file of all SNPs to STR format, and then ADMIXTURE software, based on the maximum likelihood method (version 1.3.0) [31], was used to evaluate ancestors. This software uses a fast numerical optimization algorithm to achieve a fast analysis rate. During analysis, the range of genetic clusters ( $K$ ) was set to 2-7, and each analysis was repeated 10 times. Finally, based on the cross-validation (CV) error corresponding to the optimal  $K$  value, the clustering results were plotted. (2) Using TASSEL software (version 5.2.31) [26], neighbor-joining (NJ) trees of different *P. echinogaster* populations were constructed to clarify the phylogenetic relationships between all individuals in the *P. echinogaster* populations. The acquired NJ trees were visualized with iTOL software (<https://itol.embl.de/>). (3) Using PGDSpider (version 2.0.5.2) [30], the VCF file of all SNPs was converted to STR format. The adegenet package of R software [32] was used to perform principal component analysis (PCA) of all individuals and visualize the populations and interpopulation relationships. (4) NetView P software, which adopts the k-nearest neighbor (KNN) algorithm, was used to visualize the network topology among all individuals to accurately and thoroughly explore the refine population structure of *P. echinogaster*. Before running NetView P software, the range of KNN values was first set to 1-60, and the optimal KNN value was determined according to the Fast-Greedy, Infomap, and Walktrap algorithms. In other words, the genetic similarity of individuals was determined based on the optimal resolution of the genetic structure, and then the network topology of all individuals was obtained based on the optimal KNN value. (5) Arlequin (version 3.5.2.2) [29] was used for analysis of molecular variance (AMOVA) in order to detect differentiation between groups ( $F_{CT}$ ) and differentiation between populations within a group ( $F_{SC}$ ). In this study, the *P. echinogaster* populations were divided into three groups based on geographic location: Dalian, Tianjin, Qingdao, and Nantong; Zhoushan and Xiamen; and Shantou, Zhuhai, and Zhanjiang.

## Analysis of gene flow between populations

PGDSpider (version 2.0.5.2) [30] was used to convert the VCF file of all SNPs to the GENEPOP format, and the gene flow between the nine populations was analyzed using divMigrate-Online (<https://popgen.shinyapps.io/divMigrate-online/>).

## Prediction of the genomic regions and functions of *P. echinogaster* populations that are under habitat selection

Lositan [33] and BayeScan [34] were used to screen outlier SNPs based on  $F_{ST}$ . First, Lositan software was used to screen outlier SNPs by comparing their genetic differentiation and distribution of heterozygosity. The parameters in Lositan software were set as follows: 100,000 simulations, confidence interval (CI) = 0.995, false discovery rate (FDR) = 0.05, and the infinite allele model. BayeScan software was then used to screen outlier SNPs by comparing differences in allele frequency between populations. The parameters of BayeScan software were set to the default, and the FDR = 0.05. The outlier SNPs found by Lositan software after five runs and the outlier SNPs obtained by BayeScan software that overlapped were used as the final set of outlier SNPs.

To identify the functions of genomic regions under habitat selection, we first extracted genome sequences containing SNPs under environmental-driven selection. Then, Blast2GO software was used to compare SLAF tags containing outlier SNPs with the NCBI nr and Swiss-Prot protein databases, and the homologous protein sequences with the highest sequence similarity to SLAF tags were obtained (E-values < 1E-5). Using Blast2GO software, the functional properties of the obtained homologous protein sequences were classified based on GO and KEGG databases.

## Abbreviations

SLAF-seq: specific-locus amplified fragment sequencing

SNPs: single nucleotide polymorphisms

PCR: polymerase chain reaction

VCF: variant call format

MAF 0.01: minimum allele frequency > 0.01

min-meanDP 150: minimum mean depth of coverage > 150

minGQ 98: quality score > 98

minQ 30: retaining the loci with a quality score > 3

HWE 0.05: excluding the loci with  $P < 0.05$  in the Hardy-Weinberg equilibrium test

CV: cross-validation

NJ: neighbor-joining

PCA: principal component analysis

KNN: k-nearest neighbor

AMOVA: analysis of molecular variance

CI: confidence interval

FDR: false discovery rate

## Declarations

### Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. All animal experiments were approved by the Animal Care and Use Committee at the Third Institute of Oceanography, Ministry of Natural Resources, and all methods strictly obeyed the Guide for the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines 2.0 [51].

### Consent for publication

Not applicable.

### Availability of data and materials

The dataset(s) supporting the conclusions of this article is(are) included within the article (and its additional file(s)). The raw reads of 135 individuals were showed in the NCBI database under accession numbers SRR15033948 to SRR15034082 under BioProject PRJNA743405 (<http://www.ncbi.nlm.nih.gov/bioproject/PRJNA743405>) and BioSample SAMN20033072.

## Competing interests

The authors declare no conflicts of interest.

## Funding

This research was funded by the National Programme on Global Change and Air-Sea Interaction (GASI-02-SCS-YDsum), the National Key Research and Development Program of China (2018YFC1406302), Scientific Research Foundation of TIO, MNR (2019017, 2019018), Science and Technology Project of Guangdong Province, China (2019B121201001).

## Authors' contributions

Research was conceived and designed by LSL, YL and FRL. The survey was conducted by LSL and HL. RW, YL, ZZC and RZ contributed species identification. Study conception and design were provided by LSL, YZW, YL, RZ, and XZ. Material preparation, data collection was performed by ZZC, HL, and FRL. The manuscript was written and edited by LSL, YL, FRL, and XZ. All authors read and approved the final manuscript.

## Acknowledgements

The present study could not have been performed without assistance from Mr. Cheng Liu and Miss Jiali Xiang during the experimental operation and data processing. We also thank all the editors and reviewers for their constructive comments on our manuscript.

## Authors' information

<sup>1</sup> Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, Fujian, 361005, China

<sup>2</sup> School of Ocean, Yantai University, Yantai, Shandong, 264005, China

<sup>3</sup> Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701, United States

## References

1. Schluter D. Ecological character displacement in adaptive radiation. *Am Nat.* 2000;156(S4):4–16.
2. Ferchaud AL, Hansen MM. The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. *Mol Ecol.* 2016;25(1):238–259.
3. Shanks AL, Grantham BA, Carr MH. Propagule dispersal distance and the size and spacing of marine reserves. *Ecol Appl.* 2003;13:159–169.
4. Conover DO, Clarke LM, Munich SB, Wagner GN. Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *J Fish Biol.* 2006;69:21–47.
5. Cano JM, Shikano T, Kuparinen A, Merilä J. Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. *JIFS.* 2008;5:1–10.
6. Xue DX, Li YL, Liu JX. RAD genotyping reveals fine-scale population structure and provides evidence for adaptive divergence in a commercially important fish from the northwestern Pacific Ocean. *PeerJ.* 2019;7:e7242.
7. Frankham R, Briscoe DA, Ballou JD. Introduction to conservation genetics. Cambridge University Press. 2002.

8. Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R, Hedrick PW. Conservation genetics in transition to conservation genomics. *Trends Genet.* 2010;26(4):177–187.
9. Dolganov VN, Kharin VE, Zemnukhov VV. Species composition and distribution of butterflyfishes (Stromateidae) in waters of Russia. *J Ichthyol.* 2007;47(8):579–584.
10. Yamada U, Tokimura M, Hoshino K, Deng S, Zheng Y, Li S ... Kim J. Name and Illustrations of Fish from the East China Sea and the Yellow Sea–Japanese-Chinese-Korean–Tokyo. Overseas Fishery Cooperation Foundation of Japan, 2009;525–528.
11. Oh CW, Na JH, Kim JK. Population biology of Korean pomfret *Pampus echinogaster* (Basilewsky, 1855) (Perciformes: Stromateidae) on the western coast of Korea, Yellow Sea. *Anim Cells Syst.* 2009;13(1):83–89.
12. Nakabo T. Fishes of Japan with Pictorial Keys to the Species. Tokai University Press, 2013;1079–1080. In Japanese
13. Li Y, Zhou YD, Li PF, Gao TX, Lin LS. Species identification and cryptic diversity in *Pampus* species as inferred from morphological and molecular characteristics. *Mar Biodivers.* 2019;49(6):2521–2534.
14. Cheng J., Sha Z.L. Cryptic diversity in the Japanese mantis shrimp *Oratosquilla oratoria* (Crustacea: Squillidae): Allopatric diversification, secondary contact and hybridization. *Sci Rep-UK.* 2017;7:1972.
15. Yamada U, Tokimura M, Horikawa H, et al. Fishes and Fisheries of the East China and Yellow Seas. Tokai University Press, 2007;864–875. In Japanese
16. Li Y, Lin LS, Song N, Zhang Y, Gao TX. Population genetics of *Pampus echinogaster* along the Pacific coastline of China: Insights from CR and microsatellite molecular markers. *Mar Freshwater Res.* 2018;69(6):971–981.
17. Qin Y. Development of polymorphic microsatellites for *Pampus argenteus* and its analysis on population genetic structure. Zhejiang Ocean University, 2013. In Chinese
18. Savolainen O, Lascoux M, Merila J. Ecological genomics of local adaptation. *Nat Rev Genet.* 2013;14(11):807–820.
19. Sandoval-Castillo J, Robinson NA, Hart AM, Strain LWS, Beheregaray LB. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Mol Ecol.* 2018;27(7):1603–1620.
20. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H. SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One.* 2013;8(3):e58700.
21. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol.* 2013;79(17):5112–5120.
22. Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics,* 2009;25:1754–1760.
23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, ... DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297–1303.
24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis, G, Durbin R. 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics,* 2009;25:2078–2079.
25. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics,* 2011;27:2156–2158.
26. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 2007;23:2633–2635.
27. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne RD, Horsman D, Jones SJ, Marra M. CIRCOS: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639–1645.

28. Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 2013;22:3124–3140.
29. 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.
30. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 2012;28:298–299.
31. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–1664.
32. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11:94.
33. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: A workbench to detect molecular adaptation based on a  $F_{ST}$ -outlier method. *BMC Bioinformatics*, 2008;9(1):323.
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(2):977–993.
35. Jin XH, Zhao XY, Meng TX, Cui Y. Biology Resource and Environment in the Bohai Sea and Yellow Sea. Scientific Press, 2005;204–299. In Chinese
36. Zhang QH, Cheng JH, Xu HX, Shen XQ, Yu GP, Zheng YJ. Fisheries resources and their sustainable utilization in the East China Sea. Fudan University Press, 2007;183–200. In Chinese
37. Pritchard JK, Di RA. Adaptation - not by sweeps alone. *Nat Rev Genet*. 2010;11:665–667.
38. Wu RX, Liang XH, Zhuang ZM, Liu SF. Mitochondrial COI sequence variation of silver pomfret (*Pampus argenteus*) from Chinese coastal waters. *Acta Zootaxon Sin*. 2012;37:480–488. In Chinese
39. Song N, Jia N, Yanagimoto T, Lin LS, Gao TX. Genetic differentiation of *Trachurus japonicus* from the Northwestern Pacific based on the mitochondrial DNA control region. *Mitochondr DNA*. 2013;24:705–712.
40. Zhang H, Zhang Y, Zhang XM, Song N, Gao TX. Special structure of mitochondrial DNA control region and phylogenetic relationship among individuals of the black rockfish, *Sebastes schlegelii*. *Mitochondrial DNA*. 2013;24:151–157.
41. Song N, Zhang XM, Sun XF, Yanagimoto T, Gao TX. Population genetic structure and larval dispersal potential of spottedtail goby *Synechogobius ommaturus* in the north-west Pacific. *J Fish Biol*. 2010, 77(2):388–402.
42. Xiao Y, Zhang Y, Gao T, Yanagimoto T, Yabe M, Sakurai Y. Genetic diversity in the mtDNA control region and population structure in the small yellow croaker *Larimichthys polyactis*. *Environ Biol Fish*. 2009;85:303–314.
43. Xiao Y, Song N, Li J, Xiao ZZ, Gao TX. Significant population genetic structure detected in the small yellow croaker *Larimichthys polyactis* inferred from mitochondrial control region. *Mitochondrial DNA*, 2013;26(3):409.
44. Hewitt G. The genetic legacy of the Quaternary ice ages. *Nature*, 2000;405:907–913.
45. Grant WS, Bowen BW. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered*. 1998, 89:415–426.
46. Liang YG. The physiological and biochemical adaptation of long-snout catfish (*Leiocassis longirostris*) to overwintering. Huazhong Agricultural University, 2005. In Chinese
47. Pastoureaud A. Influence of starvation at low temperatures on utilization of energy reserves, appetite recovery and growth character in sea bass, *Dicentrarchus labrax*. *Aquaculture*, 1991;99(1–2):167–178.
48. Wang JQ. Advances in studies on the ecology and reproductive biology of *Trachidermus fasciatus* heckle. *Acta Hydrobiol Sin*. 1999;23(6):729–734. In Chinese
49. Berridge MJ, Lipp P, Bootman MD. The versatility and universality of calcium signalling. *Nat Rev Mol Cell Bio*. 2000;1:11–21.

50. Flanagan SP, Rose E, Jones AG. Population genomics reveals multiple drivers of population differentiation in a sex-role-reversed pipefish. *Mol Ecol.* 2016; 25:5043–5072.
51. Percie du Sert N, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson MJ. Reporting animal research: Explanation and elaboration for the ARRIVE guidelines 2.0. *PLoS Biol.* 2020; 18(7): e3000411.

## Tables

Table. 1

Statistical analysis of genetic diversity in nine *P. echinogaster* populations

Populations	Variant_Sites	Polymorphic_Loci (%)	Num_Indv	$H_O$	$H_E$	$P_i$	$F_{IS}$
DL	46,187	91.10356	13.18523	0.22708	0.22968	0.23913	0.04062
TJ	46,187	93.28382	14.14957	0.23647	0.23379	0.24251	0.02228
QD	46,187	90.89571	13.34943	0.22208	0.22806	0.23720	0.05115
NT	46,187	92.75987	14.30790	0.22198	0.22984	0.23830	0.05271
ZS	46,187	91.41317	13.61773	0.21982	0.22789	0.23680	0.05382
XM	46,187	92.93957	14.19040	0.23021	0.23215	0.24083	0.03515
ST	46,056	86.51859	12.48565	0.20470	0.22220	0.23360	0.08049
ZH	46,132	90.93254	13.81850	0.21946	0.22750	0.23668	0.05380
ZJ	46,171	92.39349	14.21602	0.22478	0.22983	0.23846	0.04376

Table. 2

Statistical analysis of  $F_{ST}$  values between the nine *P. echinogaster* populations

Populations	DL	TJ	QD	NT	ZS	XM	ST	ZH	ZJ
DL	-	0.92793	0.71171	0.36036	0.67568	0.99099	0.14414	0.77477	0.63063
TJ	0.00033	-	0.52252	0.27027	0.39640	0.98198	0.34234	0.52252	0.59459
QD	0.00135	0.00121	-	0.05405	0.55856	0.99099	0.22523	0.62162	0.32432
NT	0.00181	0.00158	0.00234	-	0.01802	0.99099	0.15315	0.19820	0.77477
ZS	0.00140	0.00137	0.00162	0.00265	-	0.98198	0.64865	0.64865	0.76577
XM	-0.00085	-0.00047	-0.00108	-0.00014	-0.00038	-	0.99099	0.99099	0.98198
ST	0.00164	0.00106	0.00160	0.00177	0.00106	-0.00114	-	0.04505	0.56757
ZH	0.00108	0.00092	0.00136	0.00183	0.00125	-0.00121	0.00182	-	0.47748
ZJ	0.00111	0.00083	0.00157	0.00107	0.00094	-0.00104	0.00092	0.00101	-

Note: The upper diagonal is the  $F_{ST}$  values based on all SNPs and below the diagonal is the  $F_{ST}$  values based on outlier SNPs.

Table 3

AMOVA results for the three *P. echinogaster* groups

Source of variation	Sum of squares	Variance Components	Percentage Variation	Fixation Index
Among three groups	5976.486	0.57729 Va	0.02	$F_{CT} = 0.00020$
Among populations within three groups	17629.269	-1.57752 Vb	-0.05	$F_{SC} = -0.00055$
Within nine populations	373639.500	2767.70000 Vc	96.25	$F_{ST} = 0.03753$

## Figures

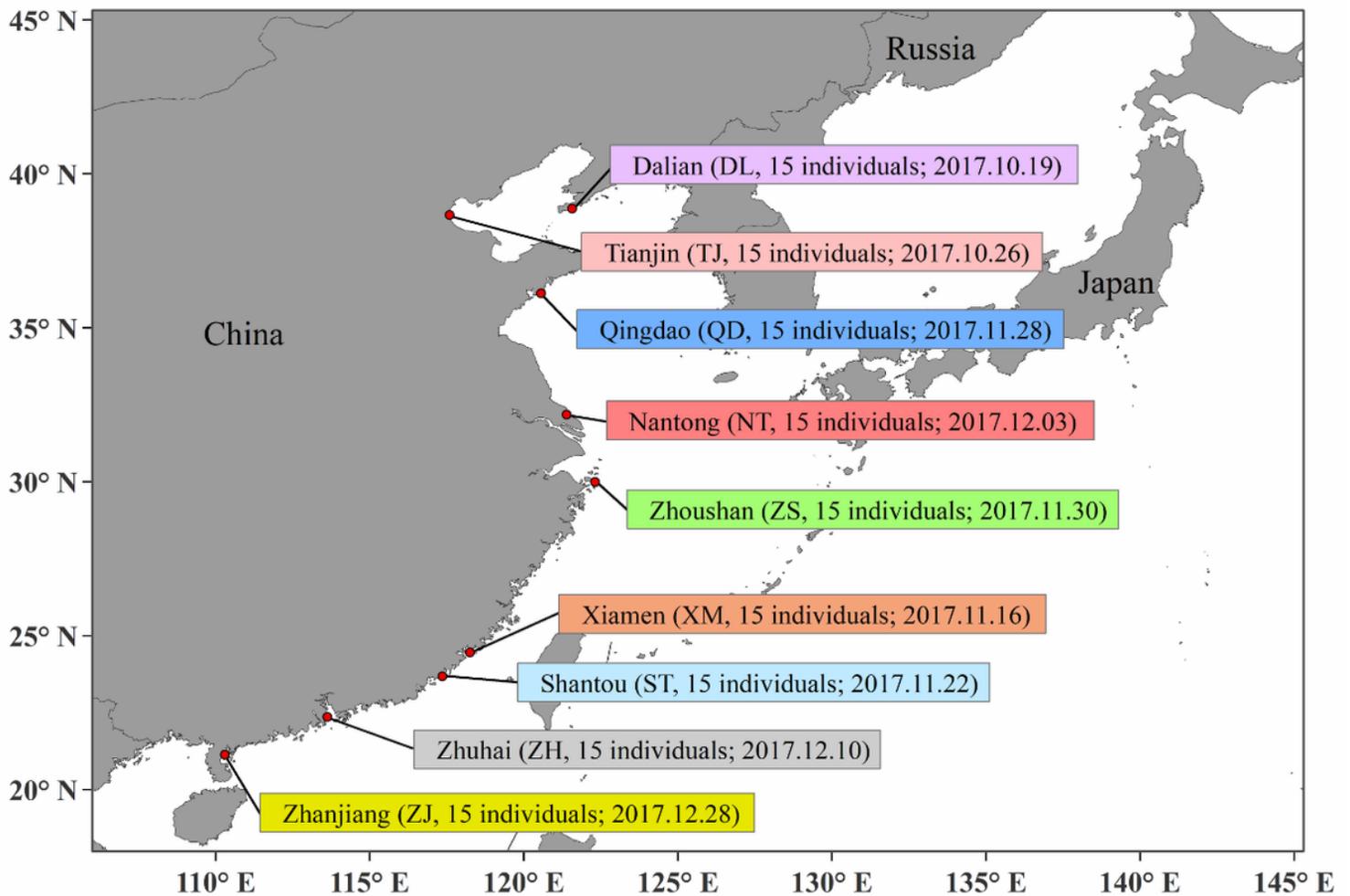


Figure 1

Sample information

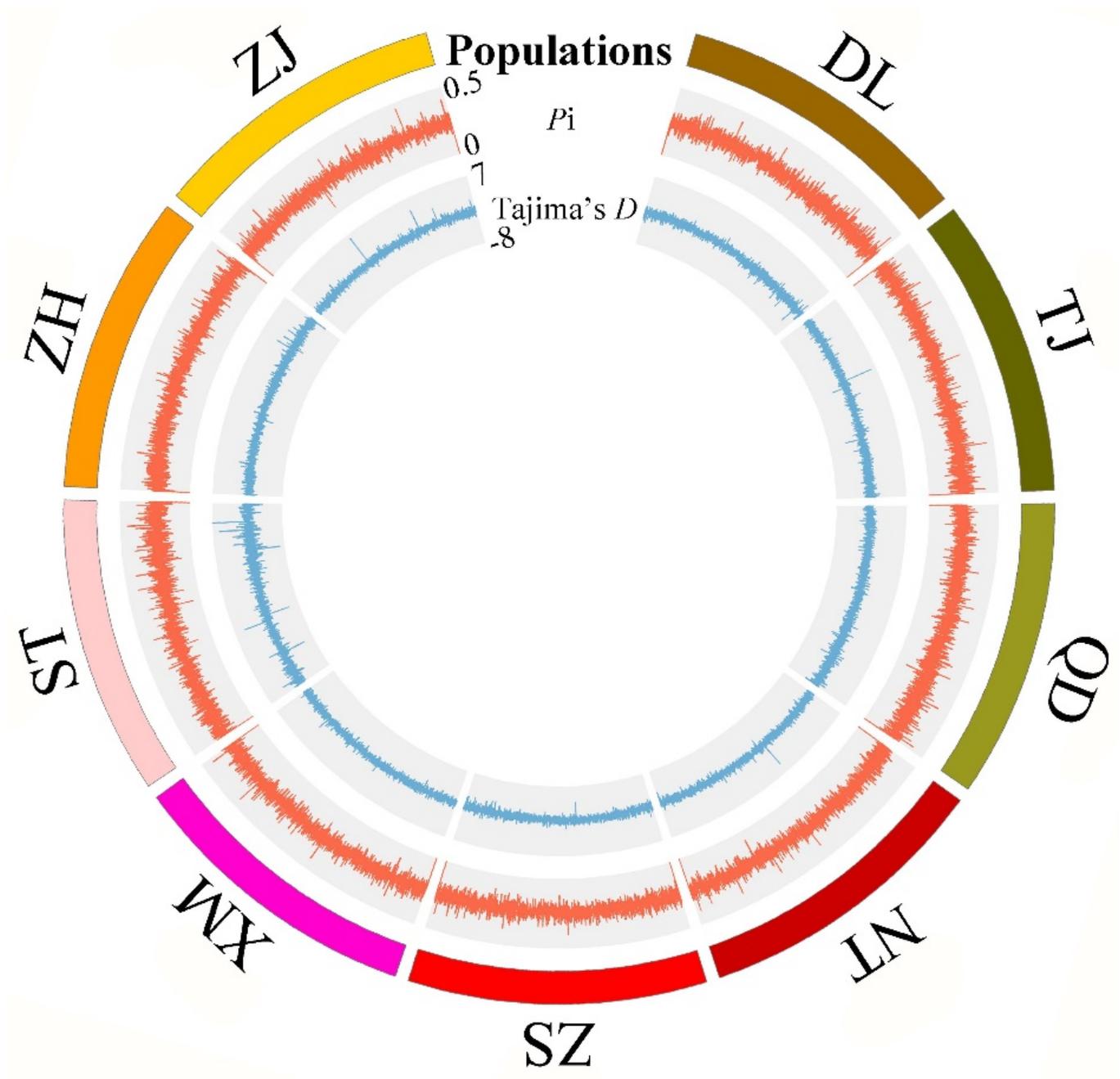
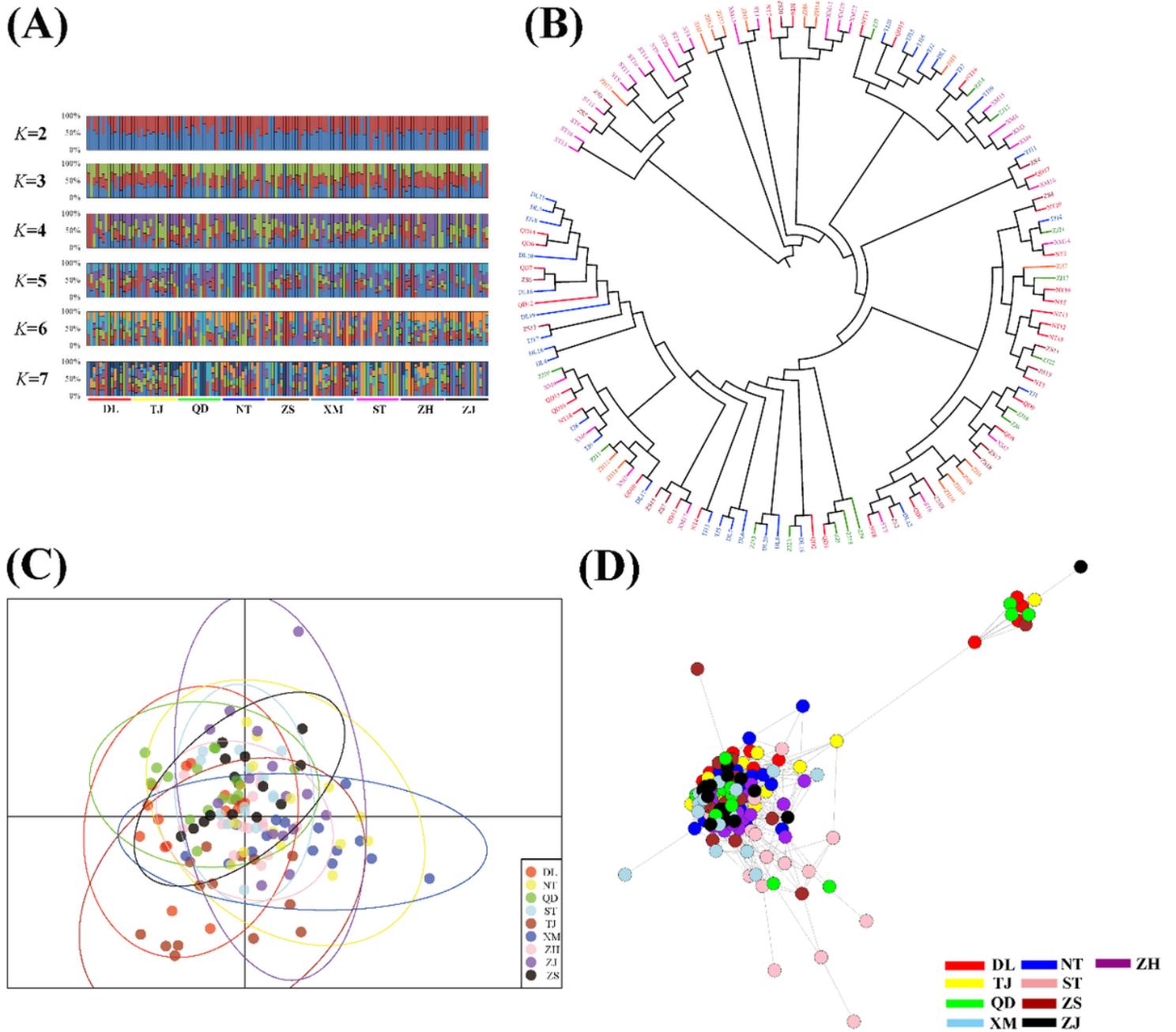


Figure 2

Statistics on the genetic diversity of a genome-wide set of SNP loci in the nine *P. echinogaster* populations



**Figure 3**

Clustering analysis results for the nine *P. echinogaster* populations based on 46,187 SNPs. (A) Clustering results from ADMIXTURE. (B) NJ tree. (C) PCA chart constructed using the first and second principal components. (D) Individual clustering from NetView P software based on a KNN = 38.



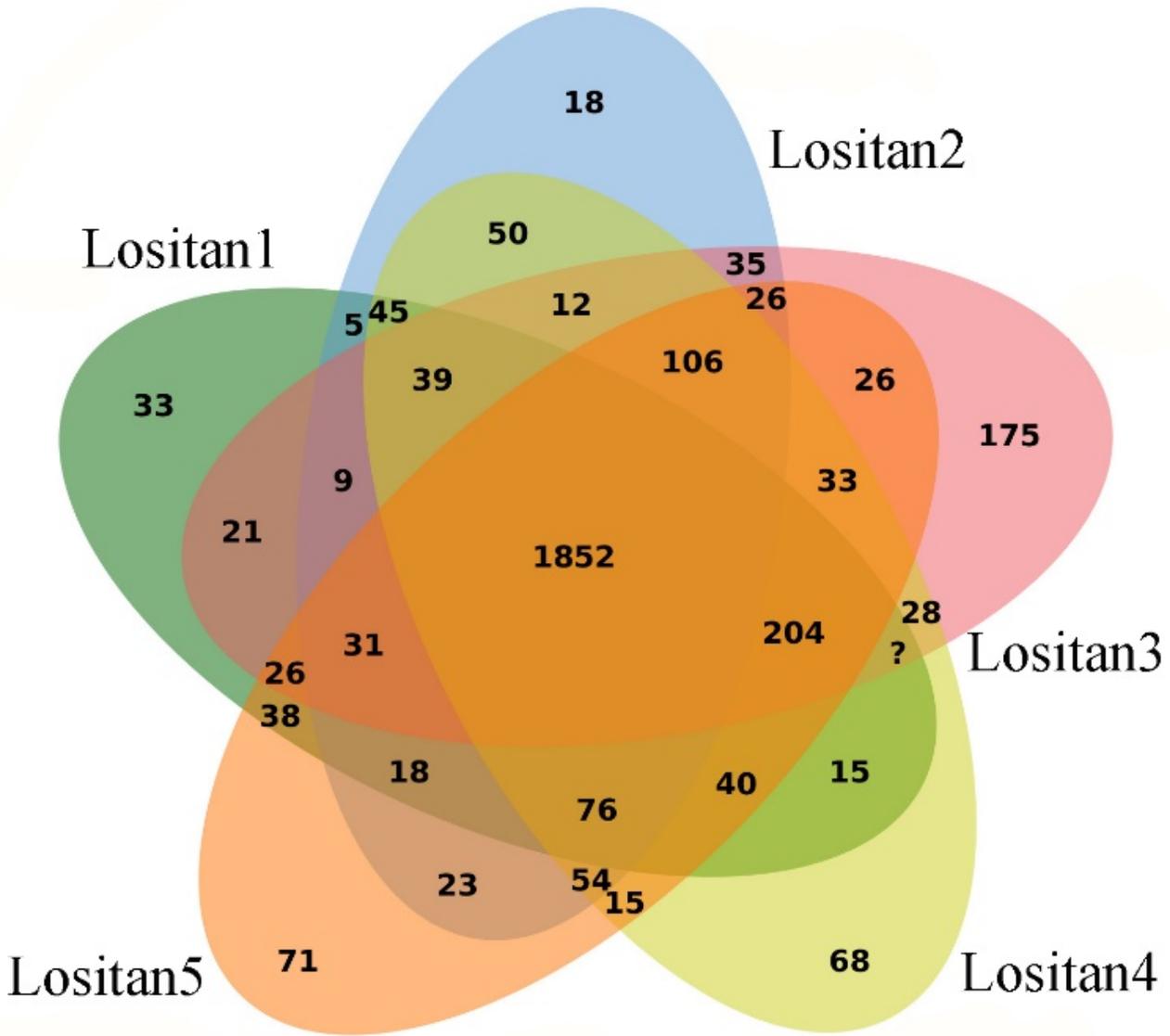
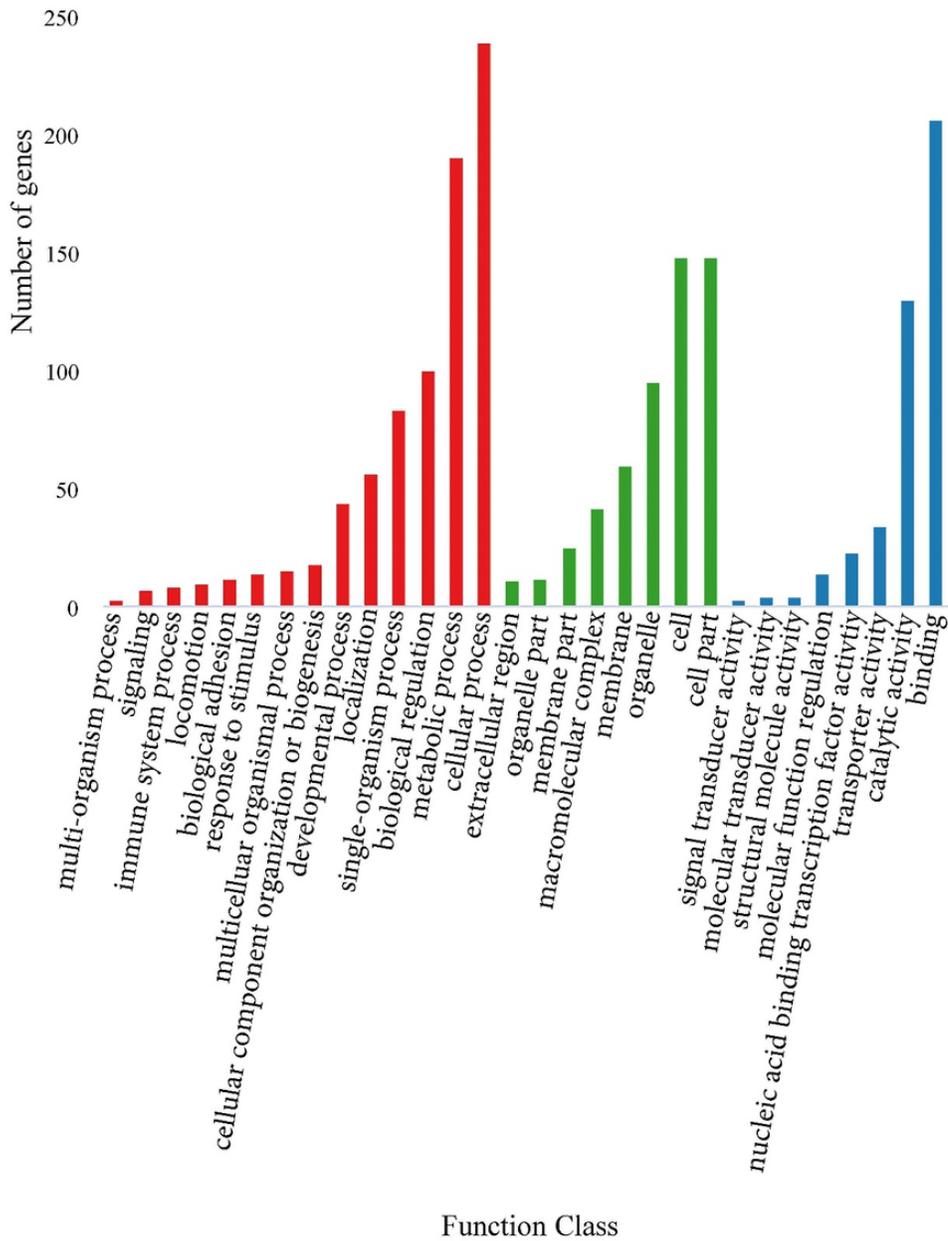


Figure 5

Results of Lositan software



**Figure 6**

GO annotation

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [4supplementaryfileThelistof194SLAftagscontainingoutlierSNPswhichassignedto249metabolicpathways0620.docx](#)