Population Genomics Provides Insights into the Population Structure and Climate-driven Adaptation of Collichthys Lucidus

Linlin Zhao
First Institute of Oceanography, Ministry of Natural Resources
Fangyuan Qu
First Institute of Oceanography, Ministry of Natural Resources
Na Song
Ocean University of China
Zhiqiang Han
Zhejiang Ocean University
Tianxiang Gao
Zhejiang Ocean University
Zhaohui Zhang (zhang@fio.org.cn)
First Institute of Oceanography, Ministry of Natural Resources

Research Article

Keywords: Collichthys lucidus, genetic diversity, population structure, local adaptive, population genomics

Posted Date: February 1st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-151802/v1

Population genomics provides insights into the population structure and climate-driven adaptation of *Collichthys lucidus*

Linlin Zhao[1], Fangyuan Qu[1], Na Song[2], Zhiqiang Han[3], Tianxiang Gao[3], Zhaohui Zhang[1*]

*Corresponding author: zhang@fio.org.cn

Abstract

Background: Understanding the genetic structure and local adaptive evolutionary mechanisms of marine organisms is crucial for the conservation and management of biological resources. *Collichthys lucidus* is an ideal candidate for investigating population differentiation and local adaptation under heterogeneous environmental pressure.

Results: To elucidate the fine-scale genetic structure and local thermal adaptation of *C. lucidus*, we performed restriction site-associated DNA tag sequencing (RAD-seq) of 177 individuals from 8 populations, and a total of 184,708 high-quality single nucleotide polymorphisms (SNPs) were identified. All the results revealed significant population structure with high support for two distinct genetic clusters, namely, the northern group (populations DL, TJ, LYG, NT, ZS, and WZ) and southern group (populations XM and ZH). The genetic diversity of the southern group was evidently lower than that of the northern group, which indicated that the southern group was possibly under climate-driven natural selection. In addition, a total of 314 SNPs were found to be significantly associated with temperature variation. Annotations of temperature-related SNPs suggested that genes involved in material (protein, lipid, and carbohydrate) metabolism and immune responses were critical for adaptation to spatially heterogeneous temperatures in natural *C. lucidus* populations.

Conclusion: In the context of anthropogenic activities and environmental change, the results of the present population genomic work could make important contributions to the understanding of genetic differentiation and adaptation to changing environments.
Keywords: *Collichthys lucidus,* genetic diversity, population structure, local adaptive, population genomics

1. Background

Inferring the genetic diversity, population structure and changing patterns of marine species is critical to the successful management of exploited populations, allowing conservation units to be identified and individuals to be assigned to geographical areas [1, 2]. However, assessing the current genetic structure and population connectivity of marine species remains a major challenge [3] because marine species usually have large population sizes and long planktonic larval stages. Moreover, the degree of connectivity between populations is often high due to less obvious geographic barriers in the oceans [4]. These biological characteristics may lead to high levels of genetic diversity and low levels of genetic differentiation between populations, even in marine species with a large distribution range [5]. Additionally, many marine species may have had insufficient time for divergence since the colonization of postglacial habitats [6]. In conclusion, it may be difficult to evaluate the genetic diversity and population structure of marine species, which may affect the reasonable management of these species.

Population genetics has a great advantage in correctly explaining the population genetic structure of marine species and exploring its influencing factors. Although previous studies have detected genetic differentiation patterns of marine species at a small spatial scale based on a small number of genetic markers [7], it is undeniable that these limited genetic markers may lead to large deviations in the results. In fact, the deficiency of hierarchical

52 analysis is extremely obvious when it is applied to the genetics of large populations with
53 limited genetic markers, even for species with weak migration abilities [8]. Therefore, it is
54 essential to increase the number of genetic markers to help us understand the population
55 structure of marine species in detail.

56 *Collichthys lucidus* is an ideal candidate for studying the influence of complex
57 geographic features of the Northwestern Pacific (NWP) on the population differentiation of
58 marine species. This species is a short-migratory shallow-sea fish that prefers brackish water
59 in estuaries and has a life cycle including pelagic eggs [9, 10]. Previous population studies
60 based on mitochondrial DNA segments suggested that changes in sea level during the
61 Pleistocene limited the spread of *C. lucidus* and promoted the emergence of isolated
62 populations, which ultimately had a significant impact on the systematic geographic pattern
63 of this species [9, 11]. However, previous studies restricted to neutral genetic markers provide
64 limited insights into the mechanisms of population structure and local adaptation. Population
65 genomics provides powerful genome-wide genotyping methods and holds great promise for
66 population genetic studies, as it can allow the detection of local adaptation under
67 climate-driven pressure [12]. This is the case because population genomics may increase the
68 power and resolution of traditional genetic approaches by increasing the number of variable
69 genome-wide genetic markers. Population genomics can also reveal genetic variation in
70 adaptive traits [13].

71 In this study, eight *C. lucidus* populations were collected from the China Sea, and
72 RAD-seq was used to identify genome-wide single nucleotide polymorphisms (SNPs) in the
73 species. The genome-wide SNPs were further used to explore the high-resolution population
74 genetic structure and local adaptation mechanism of *C. lucidus*. This study provides insights
75 into the evolutionary history and genetic diversity of *C. lucidus*. The results also provide
76 fundamental information for the management and conservation of *C. lucidus* resources under

fishing pressure.

2. Results

2.1 RAD sequencing and genotyping

RAD sequencing of 177 *C. lucidus* individuals resulted in 2,773,840,982 read pairs, 2,707,124,768 of which were retained after quality filtering (Table S1). A total of 2,561,795,216 read pairs were properly mapped to the genome for SNP calling (Table S1). After filtering out the low-quality SNPs, 184,708 SNPs were retained for subsequent population structure analysis.

2.2 Population genetic diversity and structure

These 184,708 SNPs were applied to calculate the genetic diversity of eight *C. lucidus* populations (Table 1). Estimates of H_O, H_E, and *P*i averages over the 184,708 SNPs varied across the eight *C. lucidus* populations (H_O = 0.2353 ~ 0.4147, H_E= 0.2279 ~ 0.3486, and *P*i = 0.2348 ~ 0.3560). Among the eight *C. lucidus* populations, the TJ and ZH populations showed the highest and lowest H_O, H_E, and *P*i values, indicating the highest and lowest population genetic diversity, respectively. All eight *C. lucidus* populations generally had a high percentage of polymorphic loci (73.4023 ~ 99.4640%). However, the genetic diversity in the XM and ZH populations was relatively low according to all indexes. We also found that the F_{IS} values of the eight *C. lucidus* populations were low (-0.1477 ~ 0.0209), suggesting that each population contained a large number of individuals. Additionally, most of the pairwise F_{ST} values between the eight *C. lucidus* populations were significant (Table 2), ranging from 0.00087 to 0.16222. Across all eight *C. lucidus* populations except the XM and ZH populations, the *P*i and Tajima's *D* values of genome-wide SNPs showed similar

fluctuation trends (Figure 1). The negative Tajima's *D* values indicated natural selection pressure in the XM and ZH populations, while balancing selection may have resulted in a large proportion of SNPs with positive Tajima's *D* values in the other populations.

Table 1. Genetic diversity statistics of eight *C. lucidus* populations.

Populations	Variant Sites	% Polymorphic Loci	Num Indiv	H_O	H_E	*P*i	F_{IS}
DL	184,708	97.4544	20.0000	0.3460	0.3282	0.3366	-0.0221
TJ	184,708	99.1809	24.0000	0.4147	0.3486	0.3560	-0.1419
LYG	184,708	99.1820	24.0000	0.3353	0.3360	0.3432	0.0209
NT	184,708	99.3254	24.0000	0.3360	0.3358	0.3430	0.0183
ZS	184,708	99.1554	20.0000	0.3440	0.3339	0.3425	-0.0019
WZ	184,708	99.4640	24.0000	0.4118	0.3437	0.3510	-0.1477
XM	184,708	73.4023	24.0000	0.3086	0.2414	0.2465	-0.1442
ZH	184,708	73.6021	17.0000	0.2353	0.2279	0.2348	0.0013

Table 2. Pairwise F_{ST} values between eight *C. lucidus* populations.

Populations	DL	TJ	LYG	NT	ZS	WZ	XM	ZH
DL	-	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
TJ	0.01696	-	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
LYG	0.01449	0.00268	-	**0.12613**	**0.00000**	**0.00000**	**0.00000**	**0.00000**
NT	0.01693	0.00262	0.00087	-	**0.00000**	**0.00000**	**0.00000**	**0.00000**
ZS	0.02554	0.00628	0.00598	0.00583	-	**0.00000**	**0.00000**	**0.00000**
WZ	0.02717	0.00601	0.00681	0.00477	0.00434	-	**0.00000**	**0.00000**
XM	0.16222	0.14225	0.14439	0.14089	0.14124	0.13591	-	**0.00000**
ZH	0.16315	0.14313	0.14534	0.14209	0.14195	0.13694	0.00362	-

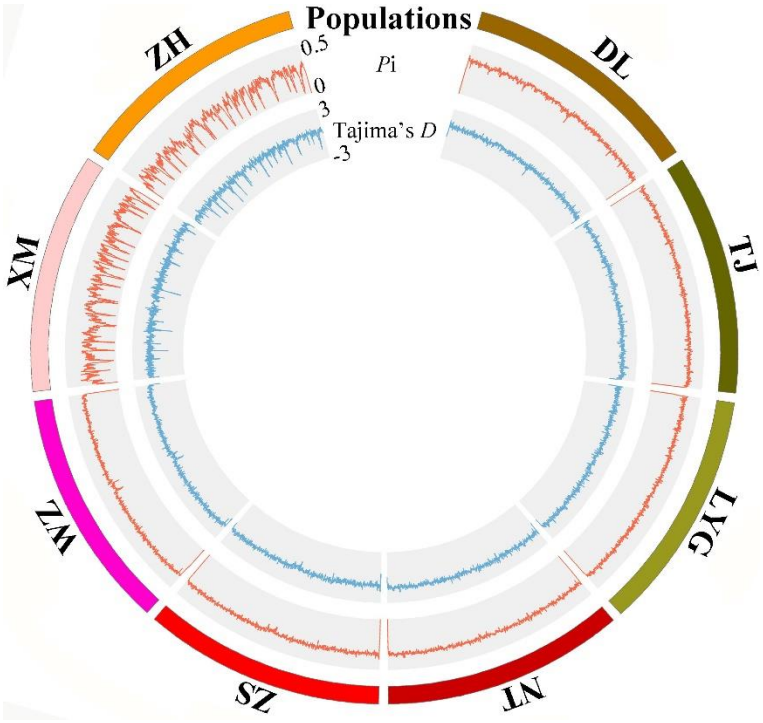

Figure 1. Genome-wide distribution of *P*i and Tajima's *D* values across eight *C. lucidus* populations based on 184,708 SNPs.

ADMIXTURE software was first used for clustering analysis of the eight *C. lucidus* populations (Figure 2). With *K* values of 2 and 3, the DL, TJ, LYG, NT, ZS, and WZ populations formed an ancestral cluster, and the Xiamen and Zhuhai populations formed another cluster. Meanwhile, the results of the PCA (Figure 3) were consistent with the ADMIXTURE results, which indicated that all populations formed two distinct clusters. NetView P with kNN = 20 was applied to reveal the clustering relationships of all *C. lucidus* individuals at a fine scale, and the results further supported the previous ADMIXTURE clustering pattern with $K = 2$ and 3, showing that individuals were grouped into two different clusters, with all individuals from Xiamen and Zhuhai clustered together (Figure 4). Additionally, the hierarchical AMOVA (Table 3) showed that the F_{ST} across the eight populations was 0. 07084, and there was significant genetic differentiation between the two groups ("Dalian, Tianjin, Lianyungang, Nantong, Zhoushan, and Wenzhou" and "Xiamen and

Zhuhai", F_{CT} = 0.13, P = 0.0004).

Table 3. Analysis of molecular variance (AMOVA) performed for two *C. lucidus* groups

Source of Variation	Sum of Squares	Variance Components	Percentage Variation	Fixation Index
Between two groups	590224.76	4353.42 Va	12.68	F_{CT} = 0.13
Among populations within two groups	252582.39	331.23 Vb	0.96	F_{SC} = 0.01
Within eight populations	5647712.50	31907.98 Vc	93.13	F_{ST} = 0. 07

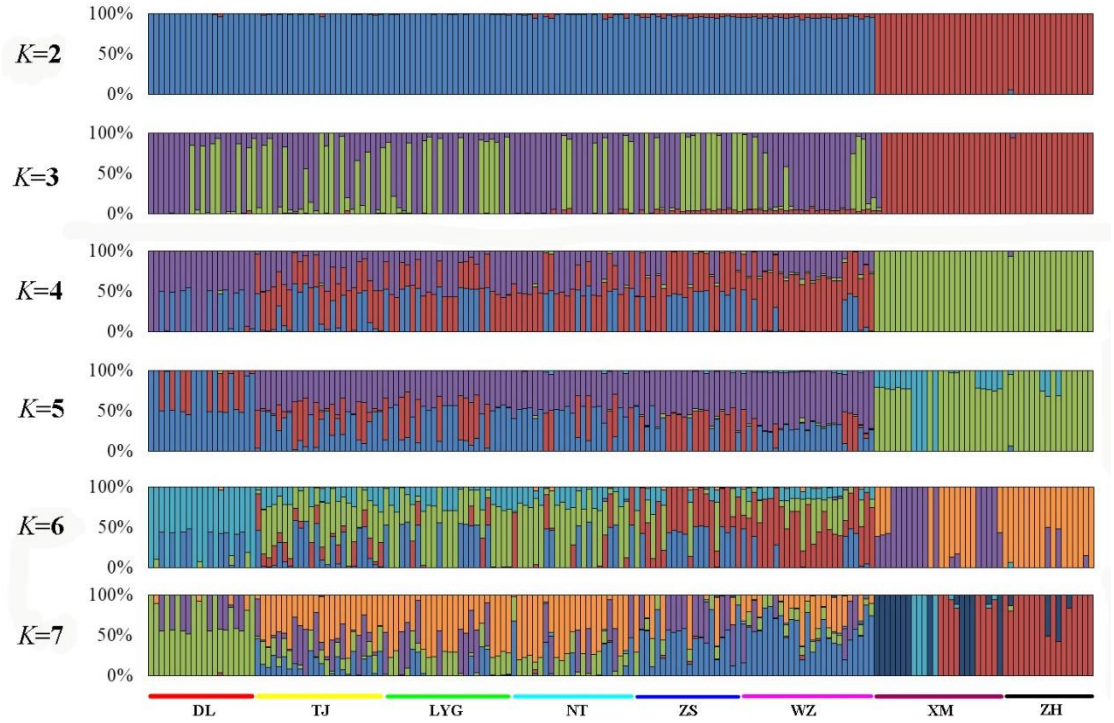

Figure 2. Plots of *C. lucidus* individual ancestry inference for K=2 to 7 based on 184,708 SNPs.

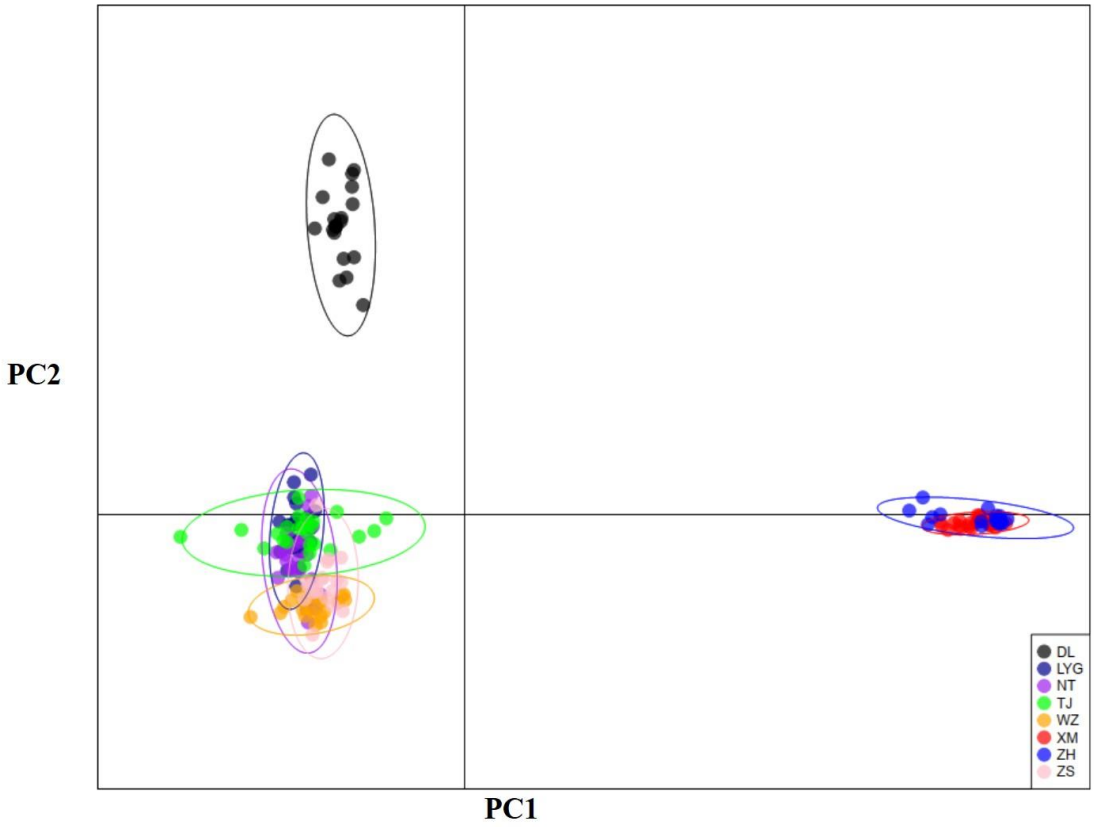

Figure 3. PCA scatter plots with prior population information using first and second components.

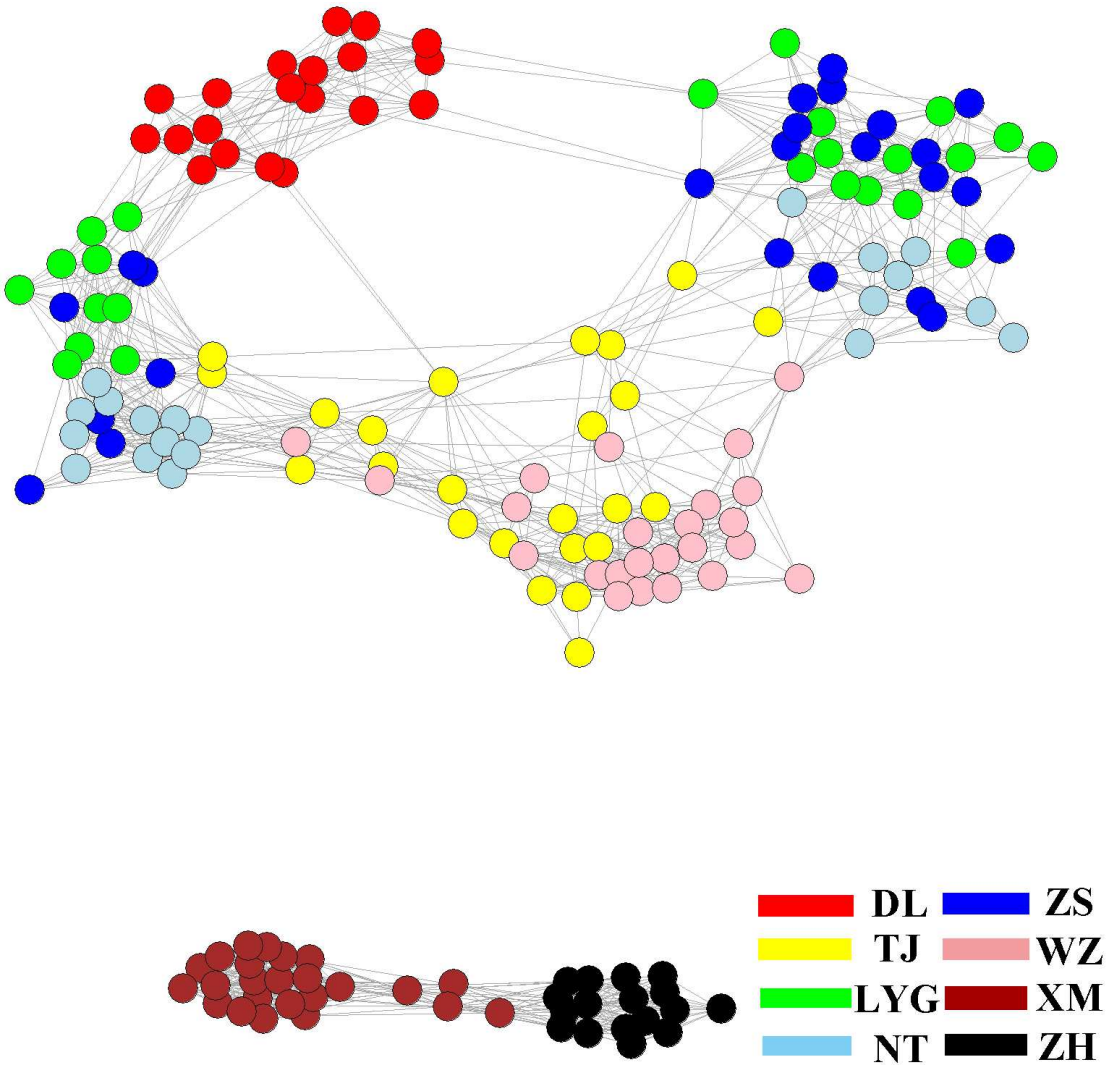

Figure 4. Individual clustering plot based on NetView P with KNN = 20.

2.3 Candidate genomic regions under temperature-driven selection

In the present study, we first calculated the average ASST, LSST and HSST of eight sea areas over 68 years (Table S2). Then, Bayenv software revealed a total of 314 SNPs associated with temperature variables. Of these SNPs, 255 were associated with ASST, 56 were associated with LSST, and 4 were associated with HSST. There was little overlap among SNPs associated with ASST, LSST and HSST. Thereafter, we used the 314 overlapping ASST-related, LSST-related, and HSST-related SNPs as the candidate temperature-selected SNPs. Whole-genome sequences containing 314 SNPs were then used

for further annotations, and the results showed that 105 sequences containing temperature-selected SNPs matched homologous protein sequences in the nonredundant protein sequences (Nr) database (Table S3). Next, the enrichment of sequences containing temperature-selected SNPs in Gene Ontology (GO) categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways was tested, and 30 significantly enriched GO terms (Figure 5) and 16 significantly enriched KEGG pathways (Table S4) were identified.

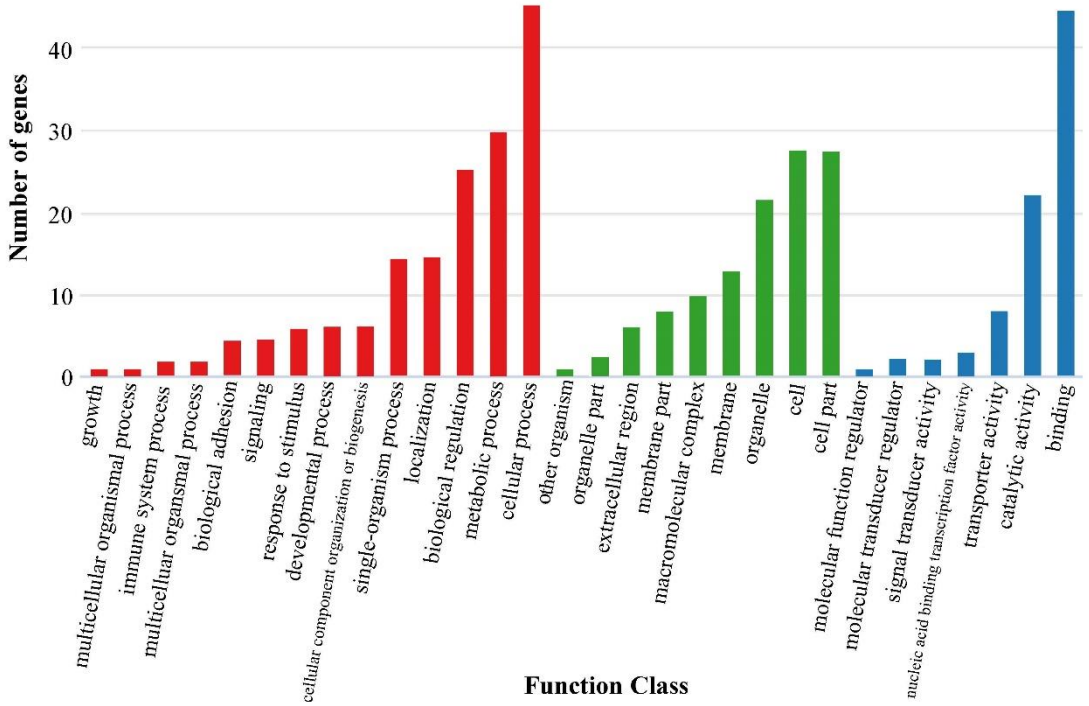

Figure 5. GO annotation information for whole-genome temperature-selected SNPs.

3. Discussion

Habitat heterogeneity has profound effects on the population genetic diversity of most marine species. Genome-level analysis not only provides detailed information on the structure, dynamics, and environmental adaptation processes of different populations but also helps us predict how populations will respond to future climate change. We first used RAD-seq to obtain genome-wide SNPs in *C. lucidus* and then delineated fine population genetic characteristics and local adaptation characteristics of the eight *C. lucidus* populations

166 at the genomic level. The present study is helpful in providing a reference for population
167 genomics analysis of other marine species and contributing to the conservation and
168 management of *C. lucidus*, especially in the context of ocean climate change.
169
170 ***3.1 Genome-wide SNPs delineating the fine population structure of the eight C. lucidus***
171 ***populations***
172 Reports of the genetic differentiation of *C. lucidus* are quite limited. The most recent
173 population genetic study was based on the mitochondrial control region [11], which identified
174 *C. lucidus* in offshore China as being divided into a southern group and a northern group,
175 with Zhoushan as the boundary. Song et al. [9] obtained a contradictory conclusion based on
176 a mitochondrial sequence and suggested that the *C. lucidus* in offshore China could be
177 divided into the East China Sea group and the South China Sea group. Herein, we used
178 184,708 SNPs to provide a higher-resolution analysis of population structure than the
179 abovementioned studies, and the results supported the conclusion of Song et al. [9].
180 All the results from ADMIXTURE analysis with $K = 2$ and 3, PCA and NetView P
181 showed that the populations investigated in this study were divided into two clusters, in
182 which the DL, TJ, LYG, NT, ZS, and WZ populations formed a cluster and the XM and ZH
183 populations formed another cluster. Notably, *C. lucidus* is a euryhaline and eurythermal
184 marine fish, and its eggs and adults have a strong diffusion ability. Therefore, this distribution
185 pattern is interesting because there are no obvious barriers between Xiamen and Wenzhou. In
186 fact, the same distribution pattern was also found in *Chelon haematocheilus* [14]. We
187 speculate that the decline in sea levels during the glacial maximum may have led to long-term
188 geographic isolation between the two *C. lucidus* clusters [15] and eventually intensified the
189 development of limited dispersal potential, reproductive isolation and local adaptive
190 heterogeneity between the two clusters [9]. During interglacial periods, although rising sea

191 levels enhanced the dispersal of the two clusters [16, 17], the individuals in the two clusters
192 may not have been able to reproduce, or the diffused individuals may have been unable to
193 adapt to the heterogeneous environment, in turn eventually dying.
194 We also calculated low pairwise F_{ST} values within *C. lucidus* clusters, although the
195 F_{ST} values between the two *C. lucidus* clusters were high. Insufficient time to attain
196 migration-drift equilibrium may have resulted in this pattern. Additionally, we speculate that
197 the strong dispersal potential of eggs and adults may be a key driver of the low pairwise F_{ST}
198 values within *C. lucidus* clusters. In fact, ocean current transport can enhance the diffusion
199 ability of *C. lucidus* and therefore significantly increase gene flow within *C. lucidus* clusters,
200 ultimately contributing to the low pairwise F_{ST} values within *C. lucidus* clusters [18].
201
202 ***3.2 Genomic regions of temperature selection in eight C. lucidus populations***
203 With the advantages of identifying adaptive SNPs, RAD-seq can facilitate insights into
204 the genetics of local adaptation in natural populations [19]. The *C. lucidus* samples in this
205 study were collected from eight different geographical locations that differed in temperature.
206 Therefore, temperature may be an important selective force affecting the genotypic and
207 phenotypic compositions of local populations. In the present study, we detected a number of
208 important candidate SNPs that appeared to be affected by temperature heterogeneity. Notably,
209 different numbers of candidate SNPs were obtained using different datasets (ASST, LSST,
210 and HSST). We speculate that the LSST is more likely than the HSST to affect the spatial
211 distribution of *C. lucidus*.
212 The GO annotation results showed that the sequences containing the
213 temperature-selected SNPs were mainly involved in metabolic and cellular processes, and
214 their functions were mainly in binding and catalytic activity. This suggests that temperature
215 differences drove adaptive differentiation in parts of the genome of *C. lucidus* populations,

ultimately leading to differences in physiological regulation. The KEGG annotation results showed that the sequences containing temperature-selected SNPs were mainly associated with material (protein, lipid, and carbohydrate) metabolism and immune responses. A previous study revealed that the metabolic capacity of organisms was more susceptible to selection under selective force [20]. Moreover, proteins, lipids, and carbohydrates are critical energy materials [21]. Additionally, Lou et al. [22] hypothesized that temperature can affect the priority of metabolic patterns, such as anaerobic carbohydrate metabolism, which is preferred by organisms at low temperatures. Therefore, we speculate that mutations in material metabolism-related genes may cause different *C. lucidus* populations to change their metabolic patterns in order to maintain a maximum metabolic capacity at different temperatures. Additionally, environmental heterogeneity will inevitably cause variation in environmental pressure among *C. lucidus* populations. Mutations in genes associated with the immune response may provide evidence of resistance specificity to temperature stress in different geographic populations. For example, the immune response was shown to be related to local adaptation to different water temperatures in *Tylosurus crocodilus crocodilus* [23], *Trachidermus fasciatus* [13], and *Larimichthys polyactis* [24].

4. Conclusion

We revealed the population genetic structure and genomic regions under temperature-driven selection based on genome-wide SNPs in eight *C. lucidus* populations. Genetic structure analysis revealed significant population structure, with high support for two distinct clusters among the eight populations. We speculate that long-term geographic isolation during the glacial maximum may have intensified the development of limited dispersal potential, reproductive isolation and local adaptive heterogeneity between the two *C. lucidus* clusters, eventually leading to strong genetic differentiation. Despite its limited ability

to detect temperature-selected SNPs, RAD-seq remains an effective method for genetic studies of local adaptation to temperature in natural *C. lucidus* populations. Unsurprisingly, many genetic variants appeared to be selected by temperature differences. Annotations of these temperature-selected SNPs suggested that genes involved in material (protein, lipid, and carbohydrate) metabolism and immune responses were critical for the adaptation of different *C. lucidus* populations to temperature. This information about the functions of temperature-selected SNPs could help us determine how different *C. lucidus* populations respond to local temperatures.

5. Methods

5.1 Specimen collection and RAD sequencing

All *C. lucidus* samples were collected from eight locations in China, namely, Dalian (DL), Tianjin (TJ), Lianyungang (LYG), Nantong (NT), Wenzhou (WZ), Zhoushan (ZS), Xiamen (XM), and Zhuhai (ZH) (Figure 6 and Table S5). Muscles were extracted from each individual using sterilized scissors and forceps. All muscle tissues were separately preserved in 95% ethanol and stored at -80°C prior to the subsequent experiments. Genomic DNA was extracted following the standard phenol-chloroform extraction method. After assessing the quality of the genomic DNA, we constructed a paired-end library with high-quality genomic DNA following the protocol described by Etter et al. [25] and then sequenced the library on the Illumina HiSeq 2500 sequencing platform.

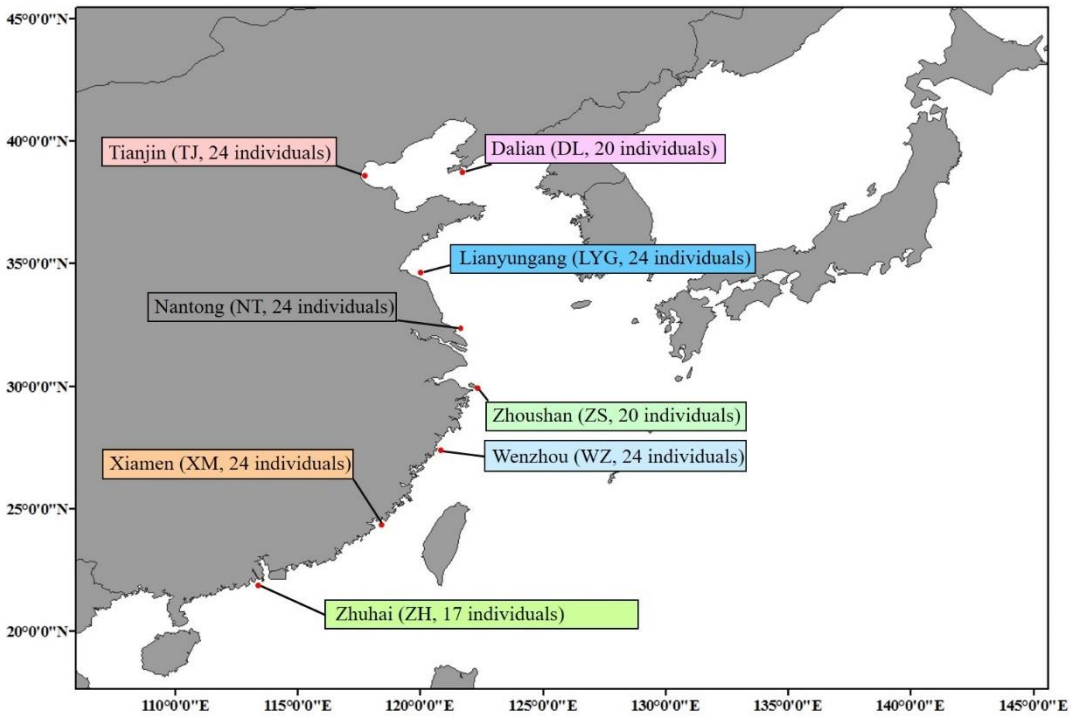

Figure 6 Map of the sampling locations and sample numbers for eight *C. lucidus* populations.

5.2 RAD data processing and SNP filtering

All raw reads in FASTQ format were filtered using Trimmomatic software (version 0.36; [26]) based on the following criteria: (I) raw reads with sequencing adaptors; (II) a ratio of unidentified nucleotides in the raw reads ≥8%; and (III) raw reads that had more than 50% of base calls with a low quality score (Q<30). After filtering, we downloaded the whole-genome sequence of *C. lucidus* [27] and used it as a reference sequence for subsequent SNP filtering. The whole-genome sequence was first constructed into an index file using BWA software (version 0.7.12; [28]). The clean reads of each sample were then aligned to the whole-genome sequence using the "bwa-mem" algorithm in BWA software (version 0.7.12; [28]) with default parameters. SNP calling was subsequently performed using SAMtools software (version 1.3.1; [29]) with the following parameters: -q 1 -C 50 -t AD, ADF, ADR, DP, SP -m 2 -F 0.002. The generated SNPs were sorted in a variant call format (VCF) file. Furthermore,

277 we removed the low-quality SNPs using SAMtools (version 1.3.1; [29]) with the following
278 parameters: --maf 0.01 --max-missing 0.1 --min-meanDP 150 --min-alleles 2 --max-alleles 2
279 --minGQ 98 --minQ 30 --remove-indels --hwe 0.05.
280
281 ***5.3 Population genetic diversity and differentiation***
282
283 The genome-wide patterns of genetic variation, including nucleotide diversity (*P*i) and
284 Tajima's *D* at each SNP, were estimated using TASSEL software (version 5.2.31; [30]). The
285 results were visualized using Circos software [31].To calculate the genetic diversity within
286 populations, nucleotide diversity (*P*i), observed heterozygosity (H_O), expected heterozygosity
287 (H_E), and the inbreeding coefficient (F_{IS}) were calculated using the "populations" module in
288 Stacks software (version 1.34; [32]. Pairwise genetic differentiation (F_{ST}) values between
289 populations and their significance were calculated using Arlequin software (version 3.5; [33]
290 with 10,000 permutations.
291 Population genetic structure based on all SNPs was analyzed using five methods. (I)
292 The Bayesian model-based clustering program ADMIXTURE (version 1.3.0; [34]) was used
293 to investigate individual ancestries, with five replicates of coancestry cluster (*K*) values
294 ranging from 2 to 7. The optimal *K* value corresponded to the lowest cross-validation error.
295 (II) We conducted a principal component analysis (PCA) using the "*adegenet*" package [35]
296 implemented in R software to infer population structure. (III) We calculated the allele-sharing
297 distance using PLINK software [36]. The NetView pipeline (version 0.7.1; [37]) with a
298 K-nearest neighbor (KNN) step ranging from 1 to 45 was then used to construct the fine-scale
299 relationships between all individuals, and the networks were visualized using Cytoscape
300 software [38]. (IV) On the basis of the optimal value of *K*, we divided all the groups into two
301 groups (group 1: DL, TJ, LYG, NT, ZS, and WZ; group 2: XM and ZH) and then performed

an analysis of molecular variance (AMOVA) in Arlequin software (version 3.5; [33]) to estimate the differentiation among groups (F_{CT}) and the differentiation among populations within groups (F_{SC}).

5.4 Outlier SNP detection and annotation

The genotype-environment association method implemented in Bayenv (version 2.0; [39]) was applied to detect putative SNPs correlated with temperature variations. First, we obtained the high-resolution mean lowest sea surface temperature (LSST), mean annual sea surface temperature (ASST) and mean highest sea surface temperature (HSST) data of eight sea areas over 68 years (from 1950 to 2017) by combining data from the Japan Meteorological Agency (JMA; http://www.jma.go.jp/jma/index.html), Advanced Very High Resolution Radiometer (AVHRR; http://oceanwatch.pifsc.noaa.gov/thredds/catalog.html) and Geostationary Operational Environmental Satellites (GOES; http://oceanwatch.pifsc.noaa.gov/thredds/catalog.html). Then, Bayenv tests were applied to identify putative SNPs correlated with temperature variations, and a Bayes factor (BF) value higher than 10 was set as the filtering condition for putative SNPs. We repeated the Bayenv analysis four times to avoid false positives, and only the SNPs that were continuously screened were used for subsequent analysis. Thereafter, we used the overlapping ASST-related, LSST-related, and HSST-related SNPs as the candidate temperature-selected SNPs. To determine the genetic mechanisms underlying temperature-related adaptive differentiation between *C. lucidus* populations, gene sequences containing these SNPs were then annotated using Blast2GO software [40].

Supplementary Information

Table S1. Sequence information for all individuals.

Table S2. The average ASST, LSST and HSST of eight sea areas.

Table S3. Nr annotation information for whole genome containing temperature-selective SNPs.

Table S4. KEGG annotation information for whole genome containing temperature-selective SNPs.

Table S5. Sampling information for eight sites of *C. lucidus*.

Abbreviations

H_O: Observe heterozygosity, H_E: Expected heterozygosity; *P*i: Polymorphism information; F_{IS}: the inbreeding coefficient; F_{ST}: Fixation index; PCA: Principle component analysis; kNN: K-nearest neighbor; ASST: Average sea surface temperature; LSST: Low sea surface temperature; HSST: High sea surface temperature.

Acknowledgements

We would like to thank Fangrui Lou for his technical support and Binbin Shan for collecting samples.

Authors' contribution

Z.Z., Z.L. and G.T. conceived and designed the experiments. Q.F. and S.N. performed the sample DNA extraction. Z.L. and H.Z. performed the computational analyses. Z.L. drafted the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (41706187).

Availability of data and materials

Sequences are available from GenBank with the Bioproject accession numbers PRJNA679902.

Declarations

The authors declare that they have no competing interests.

Ethics approval and consent to participate

C. lucidus is not protected species in China. Tissues used in this study were collected from wild according to the regulations of Fishery Administration of the Ministry of Agriculture. All experimental protocols and procedures were approved by Institutional Animal Care and Use Committee of First Institute of Oceanography, Ministry of Natural Resources.

Consent for publication

Not applicable

Author details

[1] First Institute of Oceanography, Ministry of Natural Resources, Qingdao, Shandong, 266100, China.

[2] Fishery College, Ocean University of China, Qingdao, Shandong, 266061, China.

[3] Fishery College, Zhejiang Ocean University, Zhoushan, Zhejiang, 316022, China.

References

1. Funk WC, McKay JK, Hohenlohe PA, Allendorf FW. Harnessing genomics for delineating conservation units. Trends Ecol Evol. 2012;27:489-96.
2. Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R, et al. Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nat Commun. 2012;3:851.
3. Miller AD, van Rooyen A, Rašić G, Ierodiaconou DA, Gorfine HK, Day R, et al. Contrasting patterns of population connectivity between regions in a commercially important mollusc Haliotis rubra: integrating population genetics, genomics and marine LiDAR data. Mol Ecol. 2016;25:3845-64.
4. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697-709.
5. DeWoody JA, Avise JC. Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. J Fish Biol. 2000;56:461-73.
6. Hauser L, Carvalho GR. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish Fish. 2008;9:333-62.
7. Teske PR, Sandoval-Castillo J, van Sebille E, Waters J, Beheregaray LB. On-shelf larval retention limits population connectivity in a coastal broadcast spawner. Mar Ecol Prog Ser. 2015;532:1-12.
8. Xu S, Song N, Zhao L, Cai S, Han Z, Gao T. Genomic evidence for local adaptation in the ovoviviparous marine fish Sebastiscus marmoratus with a background of population homogeneity. Sci Rep. 2017;7:1562.
9. Song N, Ma G, Zhang X, Gao T, Sun D. Genetic structure and historical demography of Collichthys lucidus inferred from mtDNA sequence analysis. Environ Biol Fishes. 2014;97:69-77.
10. Song W, Jiang KJ, Zhang FY, Zhao M, Ma LB. Molecular cloning and gene

394 expression analysis of cystatin C-like proteins in spinyhead croaker Collichthys
395 lucidus. Genet Mol Res. 2016;15. doi: 10.4238/gmr.15017417.
396 11. Liang SZ, Song W, Ma CY, Jiang KJ, Zhang FY, Zhao M, et al. Genetic structure of
397 Collichthys lucidus populations from China coastal waters based on mtDNA control
398 region. Mar Fish. 2019;41:138-48.
399 12. Lozier JD, Zayed A. Bee conservation in the age of genomics. Conserv Genet.
400 2017;18:713-29.
401 13. Li YL, Xue DX, Zhang BD, Liu JX. Population genomic signatures of genetic
402 structure and environmental selection in the catadromous roughskin sculpin
403 trachidermus fasciatus. Genome Biol Evol. 2019;11:1751-64.
404 14. Liu JX, Gao TX, Wu SF, Zhang YP. Pleistocene isolation in the Northwestern Pacific
405 marginal seas and limited dispersal in a marine fish, Chelon haematocheilus
406 (Temminck & Schlegel, 1845). Mol Ecol. 2007;16:275-88.
407 15. Palumbi SR. Genetic divergence, reproductive isolation, and marine speciation. Annu
408 Rev Ecol Syst. 1994;25:547-72.
409 16. Shen KN, Jamandre BW, Hsu CC, Tzeng WN, Durand JD. Plio-Pleistocene sea level
410 and temperature fluctuations in the northwestern Pacific promoted speciation in the
411 globally-distributed flathead mullet Mugil cephalus. BMC Evol Biol. 2011;11:83.
412 17. Xu J, Chan TY, Tsang LM, Chu KH. Phylogeography of the mitten crab Eriocheir
413 sensu stricto in East Asia: pleistocene isolation, population expansion and secondary
414 contact. Mol Phylogenet Evol. 2009;52:45-56.
415 18. Hu ZM, Zhang J, Lopez-Bautista J, Duan DL. Asymmetric genetic exchange in the
416 brown seaweed Sargassum fusiforme (Phaeophyceae) driven by oceanic currents. Mar
417 Biol. 2013;160:1407-14.
418 19. Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW.

Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol Ecol Resour. 2017;17:362-5.

20. Priede IG. Metabolic scope in fishes. In: Tytler P and Calow P, editor. Fish Energetics: New Perspective. London: Croom-Helm; 1985.

21. Wang J. Advances in studies on the ecology and reproductive biology of Trachidermus fasciatus Heckel. Acta Hydrobiol Sin. 1999;23:729-34.

22. Lou F, Han Z, Gao T. Transcriptomic responses of two ecologically divergent populations of Japanese Mantis Shrimp (Oratosquilla oratoria) under thermal stress. Animals (Basel). 2019;9:399.

23. Flanagan SP, Rose E, Jones AG. Population genomics reveals multiple drivers of population differentiation in a sex-role-reversed pipefish. Mol Ecol. 2016;25:5043-72.

24. Zhang BD, 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.

25. Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA. Local de novo assembly of RAD paired-end contigs using short sequencing reads. PLoS One. 2011;6:e18561.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.

27. Cai M, Zou Y, Xiao S, Li W, Han Z, Han F, et al. Chromosome assembly of Collichthys lucidus, a fish of Sciaenidae with a multiple sex chromosome system. Sci Data. 2019;6:132.

28. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078-9.

30. 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;23:2633-5.
31. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639-45.
32. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013;22:3124-40.
33. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564-7.
34. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655-64.
35. 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.
36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559-75.
37. Steinig EJ, Neuditschko M, Khatkar MS, Raadsma HW, Zenger KR. netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. Mol Ecol Resour. 2016;16:216-27.
38. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol. 2011;696:291-303.
39. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010;185:1411-23.

40. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674-6.

Figures

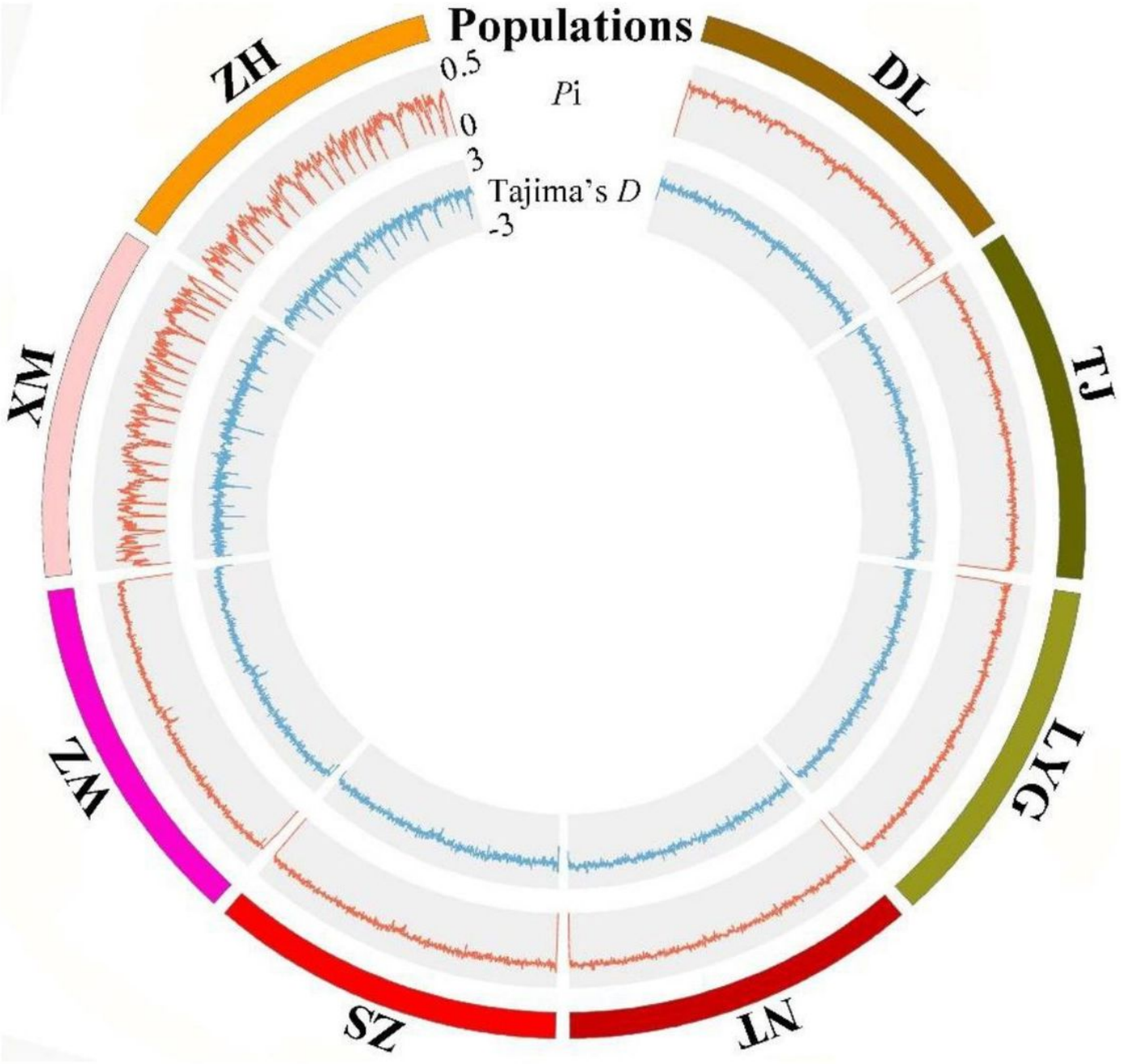

Figure 1

Genome-wide distribution of Pi and Tajima's D values across eight C. lucidus populations based on 184,708 SNPs.

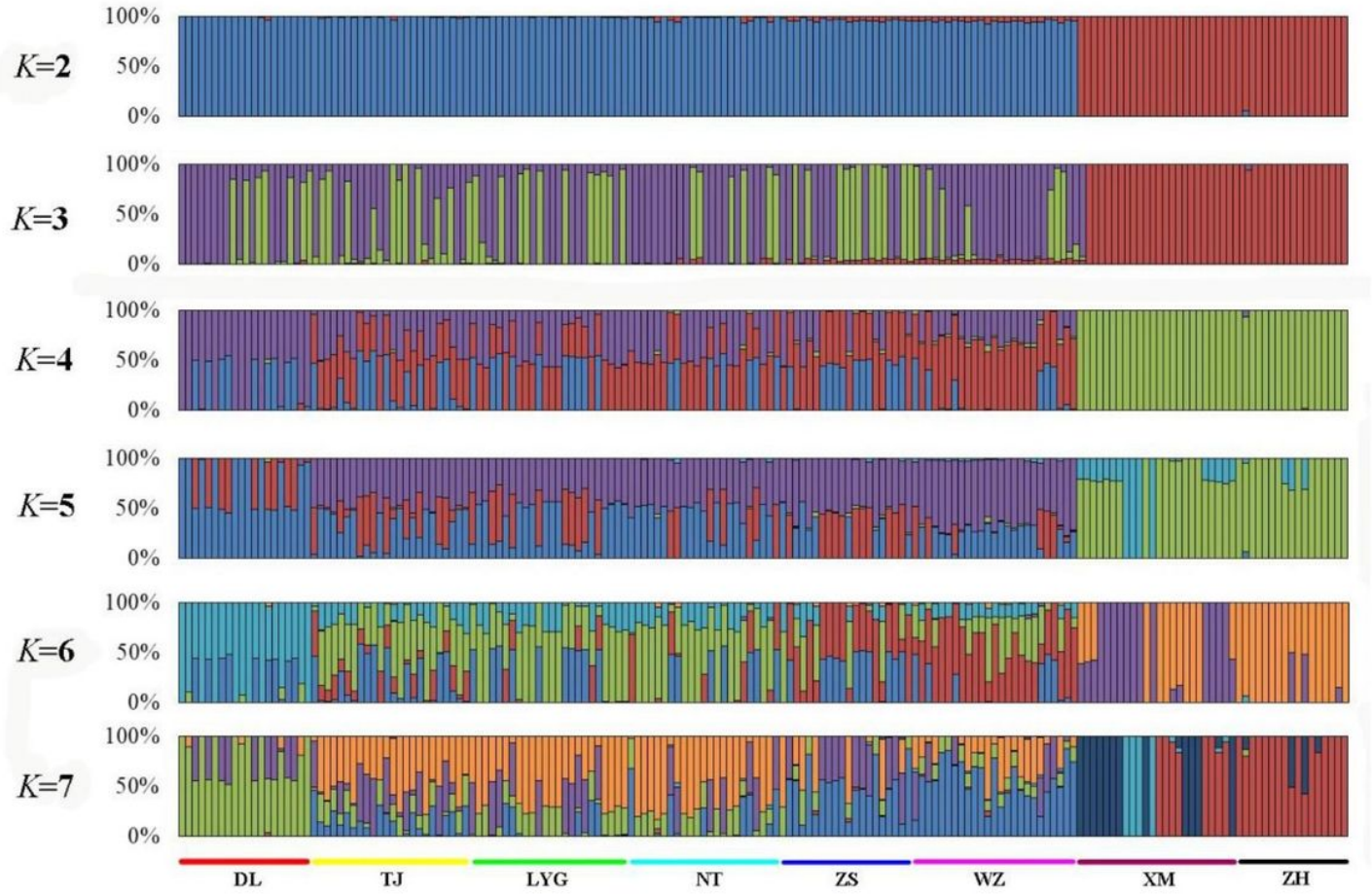

Figure 2

Plots of C. lucidus individual ancestry inference for K=2 to 7 based on 184,708 SNPs.

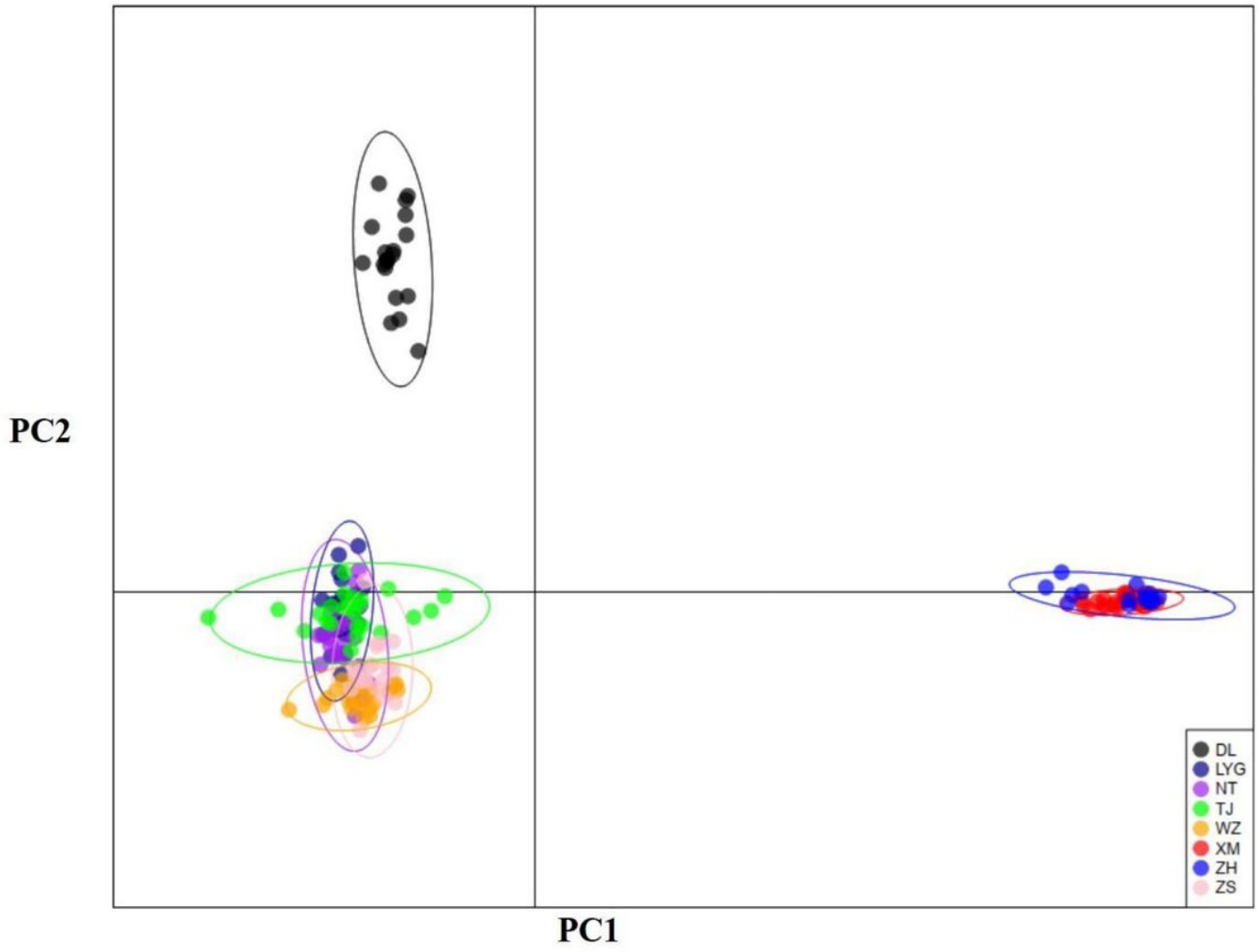

Figure 3

PCA scatter plots with prior population information using first and second components.

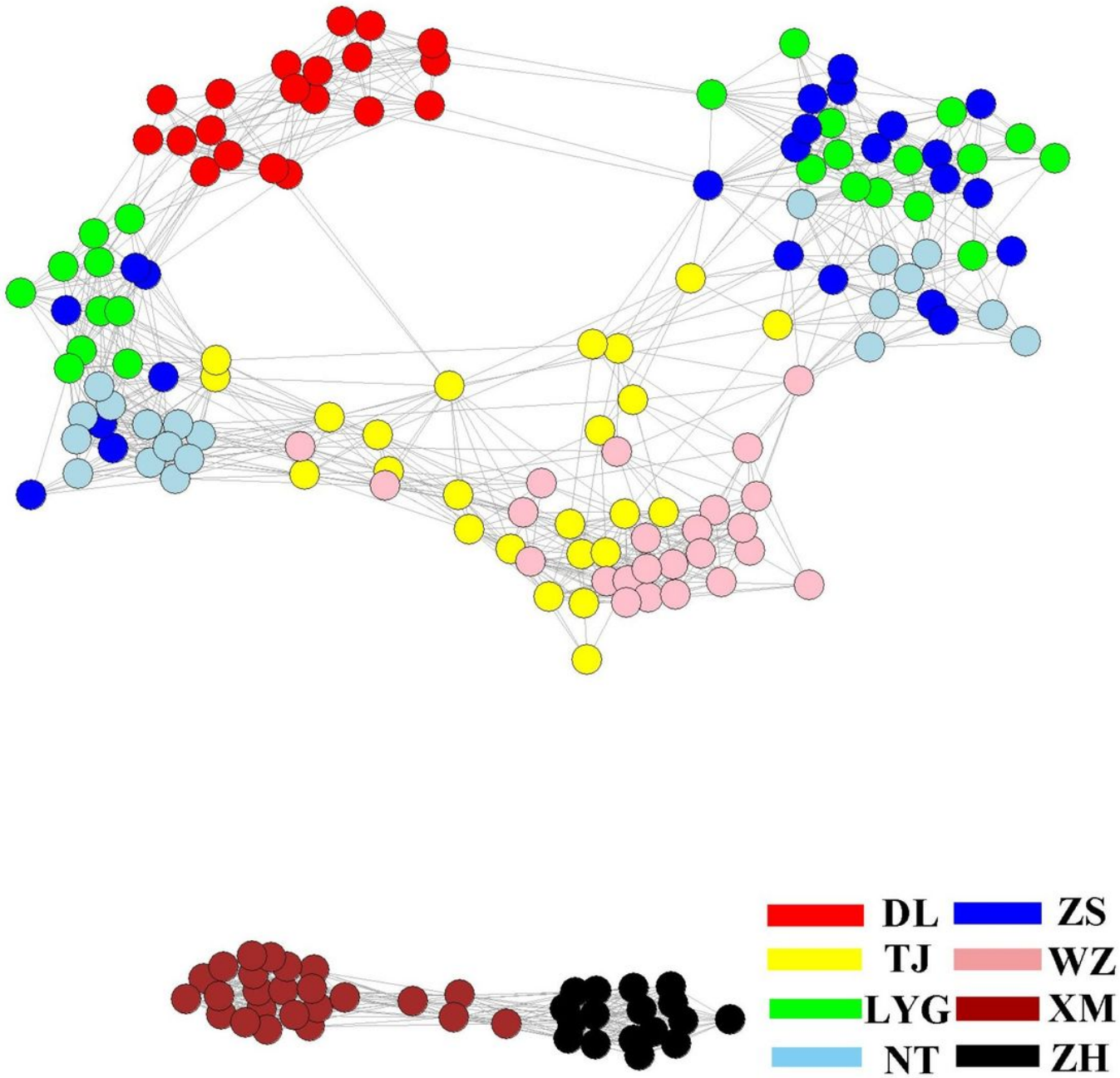

Figure 4

Individual clustering plot based on NetView P with KNN = 20.

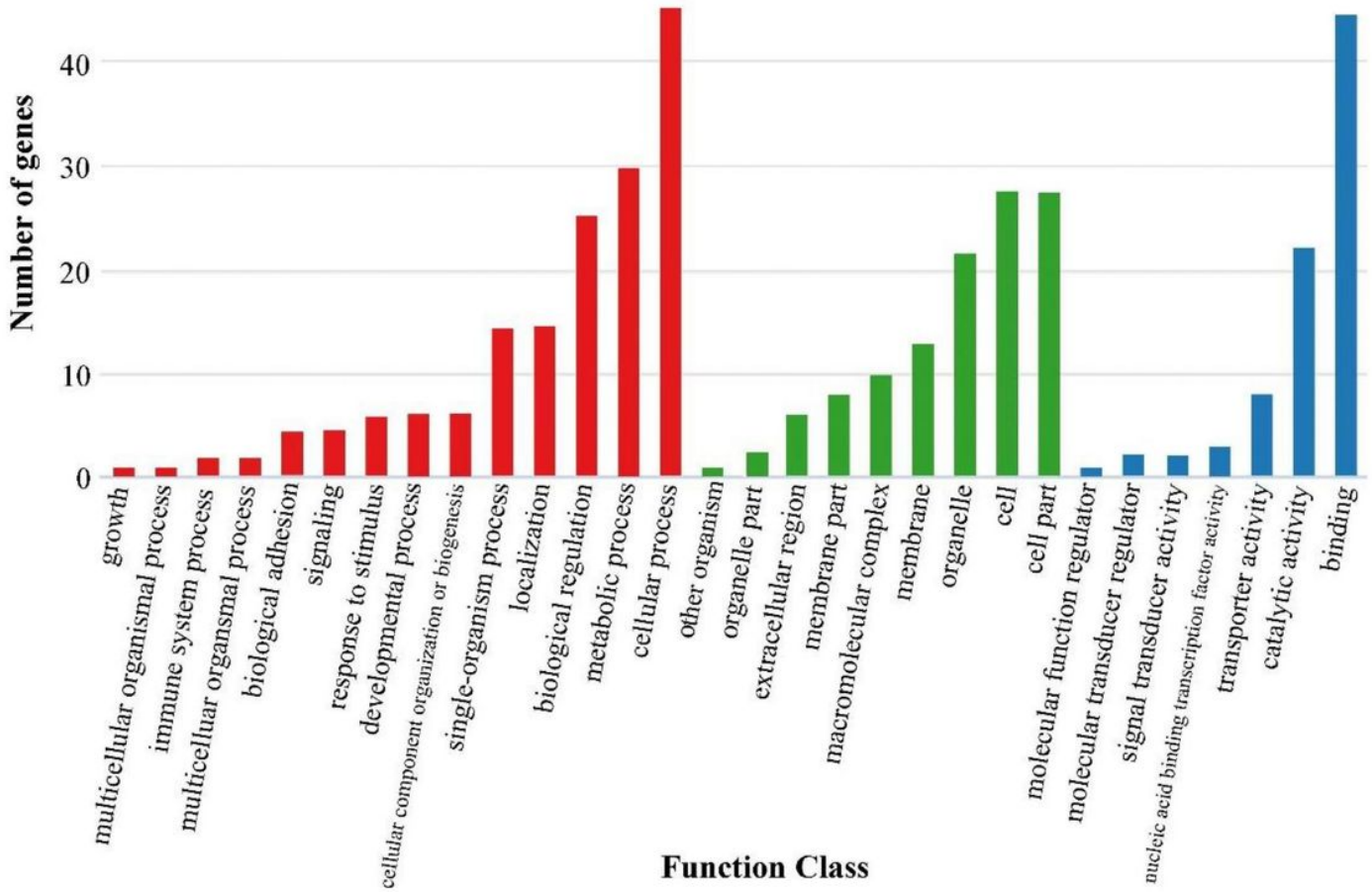

Figure 5

GO annotation information for whole-genome temperature-selected SNPs.

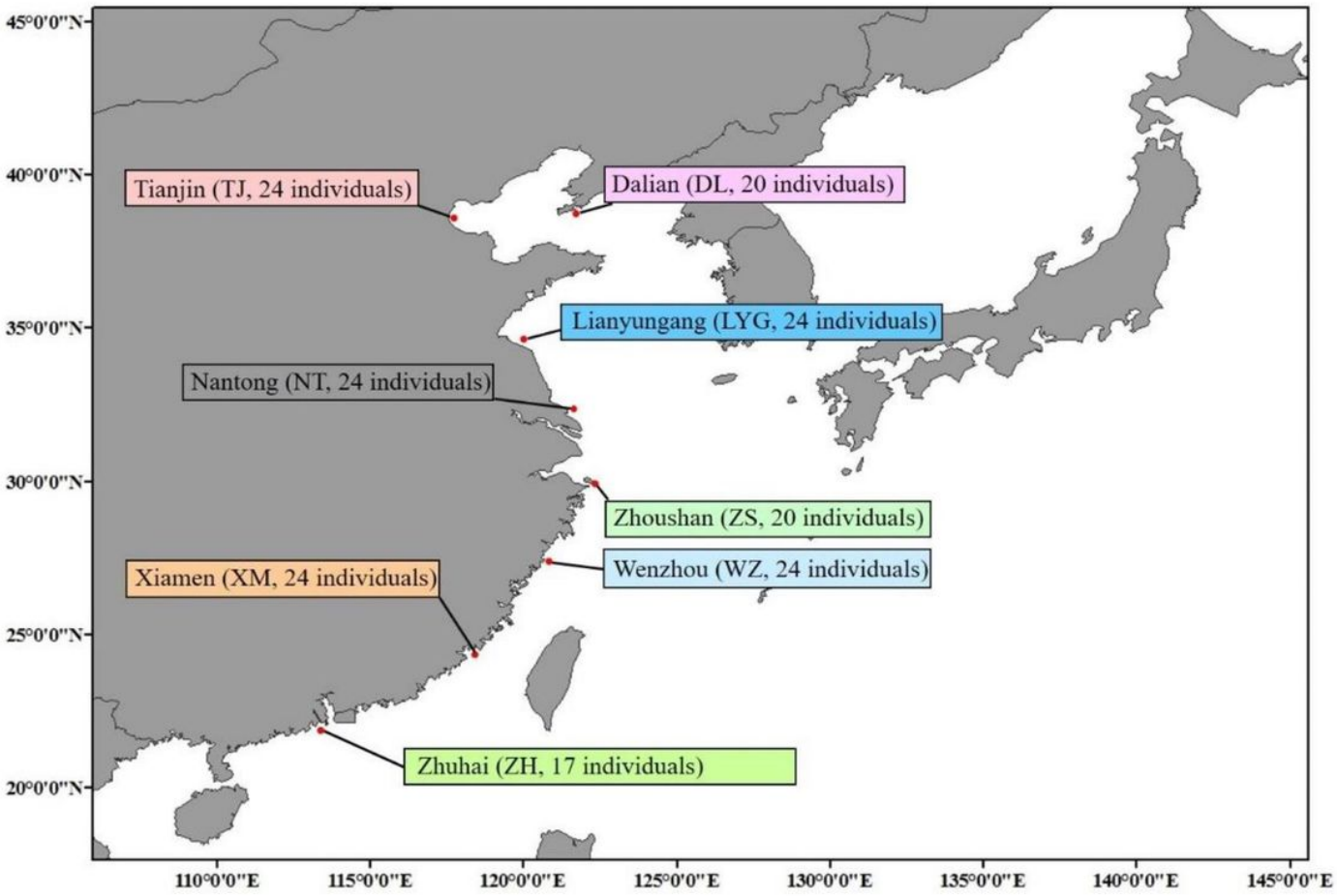

Figure 6

Map of the sampling locations and sample numbers for eight C. lucidus populations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

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

- supplementarymaterial.pdf