Genomic analysis reveals the influence of climate change and currents on adaptation in an estuarine species

Ao Li[1,2,3,4,9], He Dai[5,9], Ximing Guo[8,9], Ziyan Zhang[1,6], Kexin Zhang[1,6], Chaogang Wang[1,6], Wei Wang[1,3,4,7], Hongju Chen[5], Xumin Li[5], Hongkun Zheng[5], Guofan Zhang[1,3,4,7*] and Li Li[1,2,3,4*]

[1]CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. [2]Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao, China. [3]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China. [4]National and Local Joint Engineering Key Laboratory of Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. [5]Biomarker Technologies Corporation, Beijing, China. [6]University of Chinese Academy of Sciences, Beijing, China. [7]Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology, Qingdao, China. [8]Haskin Shellfish Research Laboratory, Department of Marine and Coastal Sciences, Rutgers University, Port Norris, NJ, USA. [9]These authors contributed equally: A. Li, H. Dai and X. Guo. [*]e-mail: gzhang@qdio.ac.cn and lili@qdio.ac.cn

Abstract

Understanding the evolutionary forces driving adaptive divergence and identifying the genomic variations, especially those mediating the plastic responses are critical to evaluate the adaptive capacity of species upon rapidly changing climate. Here we report a high-quality genome assembly for an estuarine oyster (*Crassostrea ariakensis*) and 264 resequenced wild individuals from 11 estuaries along Chinese coastline. This estuarine oyster evolved decreased polymorphism and a clear population structure than that of marine species. Historical glaciations, ocean currents and environmental selection play important role in shaping and maintaining their divergence patterns. We identified genes, especially for expanded genes *solute carrier family*, showing strong selective signals and most of them responded to temperature and salinity challenges, suggesting their significance in environmental adaptation. Higher genetic divergence of environment-responsive genes especially in upstream intergenic regions potentially regulate their higher plastic changes, providing genomic basis of plasticity upon climatic selection. Our findings contribute to assess species' vulnerabilities to climate-driven decline or extinction.

Rapidly changing climate threatens the global biodiversity and geographic distribution of organisms. Accompanying with the rising of global temperature over the past decades, the dry regions is getting drier, while wet regions wetter[1]. The salinity difference is getting greater in estuaries for increasing of salinity in the north China, while decreasing in the south[2]. Except for environmental selection, species' distribution patterns are also influenced by stochastic forces including gene flow and genetic drift. For marine species living in an open area, gene flow is restricted compared to previous perspectives, and historical climate (like glaciations) and tectonic events can shift their geographic distributions, as well as the role of the natural barrier of diluted water[3-7]. Fine-scale local adaptation among different environmental gradients has been revealed in many marine species, even those with higher dispersal capacity[8-13]. Evolutionary adaptation to various environments evolves plastic changes and accumulation of genetic mutations to alter phenotypes reaching the fitness optimum over generations[14-16]. Most of recent studies independently revealed the significance of genomic variations[12,17-19] and plasticity[13,20-22] especially for environmentally responsive genes of marine species in adaptation to changing ocean environments such as challenges from temperature and salinity disturbances. Their interaction, genomic variations of environmentally responsive genes/traits, is a critical predictor of organismal adaptive potential to climate-driven challenges[23,24]. however, genetic basis of plasticity to environmental heterogeneity remains poorly understood.

Oysters are keystone species in intertidal and estuarine ecology and one of the most important aquaculture species worldwide. Oyster is sessile species thriving in the coastal zone with highly environmental variation, and both genetic divergence and plasticity contribute to its adaptive evolution[11,25-27], which suggest that the oyster is excellent model for studying genomic basis of plastic responses to changing climates. Estuarine oysters (*Crassostrea ariakensis*, Fig. 1a) are broadly distributed in the estuaries of eastern Asia and experience high extent of environmental gradients in temperature and salinity[28-30]. A clear genetic structure among different geographic

67 populations by limited neutral markers[28,31-33], and differentiation in plastic responses
68 to temperature and salinity stresses between northern and southern populations has
69 been revealed[2,34]. Understanding whole-genome selective signals and its contribution
70 to plasticity requires exploring genomic variations and plastic changes together at the
71 same gene level. Long-read sequencing platforms provide a chromosome-level
72 assembly and a more informative genome for exploring variations under natural
73 selection[35-37]. Integrating comparative genomics and genome-wide gene expression
74 profiles can not only shed light on selective signals by identifying genomic variations
75 adapted to variable ecotypes, but also provide genomic mechanisms that mediate gene
76 expression plasticity by identifying genomic variations that located around the
77 environmental responsive genes[11,12,19].
78 Here, we sequenced the genome of estuarine oyster *C. ariakensis* and re-sequenced
79 264 wild oysters collected from 11 sites across most of Chinese estuaries, to reveal its
80 genetic structure, the potential driven forces and identify genomic regions with
81 selective signals. Transcriptomes under acute thermal and high-salt stresses, and
82 reciprocal transplantation were conducted to explore expression patterns of
83 environmentally responsive genes and further characterized the contribution of
84 genomic variations at different regions to its gene expression plasticity.

Results and Discussion

Genome assembly and annotation. To generate a high-quality reference genome, we sequenced the genome of estuarine oyster *C. ariakensis* using a combination of short-read and long-read sequencing platforms to generated a contig-level assembly. Hi-C libraries were also constructed and sequenced to organize them into a scaffold-level genome assembly. The estimated genome size based on the *k*-mer distribution analysis was 583.41 Mb (Supplementary Fig. 1, Supplementary Table 1). The polymorphism level was 0.58%, which was less than half as that of wild marine oyster species[38] (Pacific oyster *Crassostrea gigas*). and this may be resulted from limited gene flow of the oyster that are specifically adapted to estuarine environments. We applied a hierarchical assembly approach using 183.70 Gb of long-reads (299.24-fold coverage), 64.39 Gb of paired-end reads (104.89-fold coverage) and 106.34 Gb (173.22-fold coverage) of Hi-C data. The final, polished and high-contiguity genome assembly spans 613.89 Mb comprising 630 contigs with a contig N50 of 6.97 Mb. Approximately 99.6% of the genome across 416 scaffolds with a scaffold N50 of 62.26 Mb is presented on 10 linkage groups corresponding to 10 chromosomes of the estuarine oysters (Fig. 1b, Supplementary Table 2). To our knowledge, this is the most contiguous assembly among bivalve genomes using long-reads sequencing published to date[35-37]. We mapped pair-end reads to the assembled genome to assess assembly accuracy, resulting in a 97.93% mapping rate. The genome assembly captured 92.54% of the Benchmarking Universal Single Copy Orthologs (BUSCO) datasets (Fig. 1c), indicating a high level of gene region completeness in the genome assembly. Transcripts from the RNA-seq data were used to assess gene coverage rate, and 97.15% of the transcripts could be aligned to the assembly, indicating most of the gene sequences were contained (Supplementary Table 3). The accuracy of genome sequencing data was 98.32% as determined with Sanger sequencing (Supplementary Table 4). These results verified that the accuracy and completeness of the estuarine oyster genome assembly are high and of high quality at the chromosome scale.

We predicted 29,631 protein-coding genes in the estuarine oyster genome combining gene evidence from homology annotation, *de novo* annotation and transcripts of mRNA sequencing, and 96.13% of which were functionally annotated (Supplementary Table 5). Various non-coding RNA sequences were also identified and annotated in the genome, including 1,077 transfer RNAs, 20 micro RNAs and 131 ribosomal RNAs. A total of 332.40 Mb (54.14%) of repeats and transposable elements were identified (Supplementary Table 6). Generally, the gene density was inversely related to the content of repeat elements across all chromosomes.

Genome resequencing, variation calling and population structure. We resequenced 264 wild oysters of *C. ariakensis* from 11 estuarine areas, representing the major distribution range across north, middle and south geographical regions of Chinese coastlines[28,29] (Fig. 2a), and obtained 3.81 Tb clean data. The mapping rate averaged 95.33% varied from 86.46% to 96.66%, and the effective mapped read depth averaged 19.89× by aligning reads to the *C. ariakensis* reference genome. We generated 145,271,754 SNPs (range from 487,881 to 640,962 per individual) and 103,080,822 indels (range from 342,486 to 443,381 per individual) acrossing genes ranging from 11,383 to 13,321. In addition, we found a considerably lower polymorphism that averaged 0.47 heterozygous SNPs per Kb per individual (Supplementary Table 7), which was more than 35-fold lower than that in Pacific oyster populations[11]. The number of genomic variations including SNPs and indels (insertions and deletions) was gradually increased from north to south estuarine oysters, and more than half of variations were SNPs (58.45%) that mainly located at intergenic regions (57.38%, Supplementary Fig. 2).

Genetic structure analysis based on genome-wide SNPs supported previous findings that differentiations occurred among different geographic populations of *C. ariakensis* in China, using fitness-related traits, neutral markers and transcriptomic analysis[28,32-34]. The optimal number of population clusters was identified as $k = 3$ (Supplementary Fig. 3), exactly representing three main geographic sea areas as north estuaries of China

(NEC, 5 sites), middle estuaries of China (MEC, 2 sites) and south estuaries of China (SEC, 4 sites) (Supplementary Fig. 4). Principle component analysis (PCA), explaining 15.98% of genetic variance by the two PCs, consistently reveals three distinct populations of oysters corresponding to NEC, MEC and SEC. A fine-scale subpopulation was further detected that oysters from Qingdao (QD) site and farther southern subpopulation (SEC-b) including Taishan (TS) and Qinzhou (QZh) sites were separated from NEC and SEC, respectively (Fig. 2b). Moreover, Phylogenetic tree using neighbor-joining (NJ) method supports the above clustering, which first distinguish the southern population from others and further identified another subpopulation (NEC-a) within NEC distributed along estuaries of southern Bohai Sea including oysters from Binzhou (BZ) and Dongying (DY) sites (Fig. 2c). A total of six subpopulations were identified for 11 wild estuarine sites in China. We calculated the pairwise F_{ST} for all polymorphic positions among three populations and QD oysters, and revealed strong divergence between SEC and other populations, ranged from 0.143 to 0.225 (Fig. 2d), which also detected among oysters from different geographical sites (Supplementary Table 8). Oysters from MEC and NEC were clustered together in phylogenetic tree with a lower genetic divergence ($F_{ST} < 0.05$), which is comparable with the divergence among populations of marine oyster species in north China[11]. Linkage disequilibrium (LD, measured as r^2) decreased to half of its maximum value range from 2.54 kb to 3.00 kb among three populations of estuarine oysters (Supplementary Fig. 5), which is substantially slower than marine oyster species[11]. Our findings provided insights into the fine-scale population structure of estuarine oysters along the coast of China and revealed the stronger genetic divergence of southern oysters from others.

Limited gene flow, genetic drift and positive selection shape the distribution patterns. Population structure of estuarine oysters was largely concordant with the direction of ocean currents especially during the summer, where the northern or southern coastal currents are not cross over the middle area nearby the Changjiang

estuary (Fig. 2a and Supplementary Fig. 6). To investigate the effects of ocean currents on gene flow, we examined nucleotide diversity (π) for each population and found that higher genetic diversity was observed in the middle population (3.56×10^{-4}) and QD oysters (3.59×10^{-4}) where is the joint of more than one ocean currents, than that of northern (3.33×10^{-4}) and southern (3.43×10^{-4}) populations (Fig. 2d) where only adjoin one ocean current (Fig. 2a and Supplementary Fig. 6). The same pattern of nucleotide diversity was also confirmed by oysters from individual geographical sites that oysters from middle sites showed higher π values, while oysters from northern and southern sites have lower diversity (Fig. 3a), which is contrast to the terrestrial vertebrates where wildlife harbored higher genetic diversity in southern China[39]. These findings indicate that ocean currents play important role in shaping and maintaining genetic structure of *C. ariakensis* in China[11,40], which contribute to the homogenization within oyster population by facilitating the mixture of larvae from sites adjoining the currents with the same direction, but the divergence among populations adjoining the currents with the different directions such as the NEC and SEC. However, the MEC oysters showed very limited mixed ancestries from northern and southern populations (Supplementary Fig. 4), as well as the fine-scale subpopulations within SEC and NEC even in the same direction of ocean currents (Fig. 2a), indicating the role of ocean currents in mixing gene pools is restricted as also found in marine oyster populations[11]. In addition, Changjiang dilute water (CDW) is considered as an natural barrier that facilitates divergence between north and south populations of many marine species in China[5-7,41,42]. Here, however, we found that oysters sampled from the north (NT) and south (SH) part of Changjiang River were clustered together as MEC, indicating CDW is not the barrier and has limited influence on distribution and divergence patterns for estuarine oysters where organisms experience wide-range of salinity disturbance.

We employed the pairwise sequentially Markovian coalescent (PSMC) to reconstruct the demographic history and assess fluctuations in effective population size (*Ne*) of ancestral estuarine oysters *C. ariakensis* in response to Quaternary climatic change

using deep-coverage (25-28×) oyster genomes of two or three individuals from each of three genetic/geographic populations, as well as three individuals of marine oyster species from resequencing data with the average coverage of 20×[11]. Both of the marine and estuarine oysters were severely influenced by glaciation events during the past million years that the *Ne* had a peak at ~0.90 mya before the Mindel glaciation (MG, 0.68~0.80 mya) and then was substantially decreased when subjected to subsequent three times of glaciations including MG, Riss glaciation (RG, 0.24~0.37 mya) and Würm glaciation (WG, 10,000~120,000 years ago), and a relatively slower decreasing was found at inter-glaciation period between RG and WG. However, the *Ne* of estuarine oyster species was massively lower before historical glaciations and consistently lower than marine species, and the latter was earlier and more rapidly increased the population size (Fig. 3b). Moreover, we compared nucleotide diversity among three populations of estuarine oysters and marine oysters (n = 26), and found that marine species have more than 25-fold higher of π values than that of estuarine oysters ($\pi_{\_C.\ gigas}$ = 9.27 × 10^{-3}, Fig. 3c). Similarly, stickleback fishes adapted to freshwater areas exhibited lower π and *Ne* values than counterparts dwelling in brackish areas[43]. These findings addressed that the specified life history of adaptation to restricted estuarine areas affected the *Ne* and nucleotide diversity of *C. ariakensis*.

Generally, all three oyster populations exhibited similar demographic trajectories until about 0.2 mya (Fig. 3b). The *Ne* curves of southern population was the earliest to split from others that occurred ~0.18 mya corresponding to the isolation period by the land bridge between Taiwan and Mainland of China from 0.2 mya to 25,000 years ago[44]. Accordantly, we calculated the putative range of divergence time was 0.14~0.63 mya between northern and southern populations of *C. ariakensis* using previously reported pairwise sequence divergence of *COI* gene[28], based on the sequence divergence of *COI* gene[41] and divergence time[45] between *C. gigas* and *C. angulata*. In addition, the split of *Ne* curves between middle and northern populations occurred ~90,000 years ago, corresponding to the grisly fall of the sea level at the sub-glaciation of WG that the majority of the Bohai Sea was land[46]. Lower *Ne* and nucleotide diversity of

northern population (Fig. 3a, b) suggested the stronger recent bottleneck on oysters dwelling in the Bohai Sea. Decreased nucleotide diversity was also found in Bohai populations of marine oyster species[11]. The role of bottlenecks and geographic isolation resulted from historical glaciations and tectonic events in shaping species' distributions were also demonstrated in other molluscs, such as the divergence between Atlantic coast and Gulf of Mexico populations of eastern oyster[3]. Our results not only provide putative divergence times among northern, middle and southern estuarine oyster populations, but also pointed out historical glaciation and tectonic event was the critical evolutionary drivers to shape their differentiation.

We further characterized the genomic variations related to natural selection among three oyster populations. The number of SNPs showing heterozygous in northern oysters but homozygous in southern oysters (n = 14,373) were 1.89-fold higher than those of SNPs showing homozygous in northern oysters but heterozygous in southern oysters (n = 7,595) across 10 chromosomes (Fig. 3d). Correspondingly, homozygous variations in southern oysters (50.71% ± 1.92%) were significantly higher than that in middle (44.35% ± 0.39%) and northern oysters (44.53% ± 0.86%) ($p < 0.01$, Supplementary Fig. 7). Both divergent selection and genetic bottlenecks can result in purified homogeneous variations. Furthermore, southern oysters (35.67% ± 0.35%) evolved significantly higher ratio of genes with non-synonymous alleles than these in middle (33.33% ± 0.22%) and northern (31.82% ± 0.29%) oysters ($p < 0.01$, Fig. 3e), suggesting southern oysters were subjected to stronger natural selection. Our results revealed that stronger natural selection, potentially resulted from strong environmental gradients, preferred to purify the homozygous mutations in southern oysters.

In summary, genetic bottlenecks from historical glaciations and geographic events play important role in shaping geographic distribution patterns of estuarine oysters in China, and gene flow from ocean currents and natural selection from environmental gradients synergistically contribute to maintain their divergent patterns.

Genomic signatures related to environmental adaptation. Oysters inhabiting in northern and southern habitats subjected to significant environmental gradients. Southern habitats are characterized with higher temperature in comparison with northern and middle habitats. Monthly average sea surface temperature (SST) of six sampling sites from each of six subpopulations were collected from satellite remote sensing data during 2000 to 2017. Average SST of southern habitats showed a 10.35 □ higher than that of northern habitats (Supplementary Fig. 8). In addition, we found a 10.98 ‰ higher of salinity at northern (BZ) site than southern (TS) site[2]. And increasing ocean salinity contrast would intensify the salinity increases at north regions but decreases at south regions[47,48]. Thus, variations in temperature and salinity between northern and southern habitats are two of the critical environmental factors that driven adaptive divergence of northern and southern oyster populations. Combining geographically disconnected distribution and limited gene flow, local adaptation was revealed that southern population evolved higher thermal tolerance and greater sensitivity to high salinity[2]. Observing distinct population structure among three oyster populations, we propose that there were genomic regions under selection contribute to adaptation to higher temperature and lower salinity conditions in the south.

To gain insights into the adaptive genetic basis in adapting to southern environments, we scanned for selection signatures in two population pairs including a) north vs south and b) middle vs south. We calculated the fixation index (F_{ST}) and selection statistics (Tajima's D) between two population pairs, and used an outlier approach (top 1% F_{ST} values, $F_{ST\_north\ vs.\ south}$ = 0.693, $F_{ST\_middle\ vs.\ south}$ = 0.637) to identify genomic regions undergoing selective sweeps in these three oyster populations. Only genomic regions surrounding the selective peaks that overlapped between the two population pairs, and further located at the ravines of Tajima's D values along each chromosome in one of the three populations were considered as selective signals. A total of 24 selective regions spanning 51 candidate genes (44 annotated) were identified along chromosomes 2, 3, 4, 6, 8 and 9 among three oyster populations (Fig.

4a, b and Supplementary Fig. 9-14, Supplementary Table 9). Most of these candidate genes in other species have also been reported to respond to different environmental gradients of salinity and temperature[12,19,49-54].

We conducted RNA-seq analysis to examine the responses of genome-wide and selective genes using wild oysters, which were exposed to sublethal conditions of elevated temperature (6 hours under 37 □) and high-salt (7 days under 60 ‰). Low expressed genes that the aligned read counts were less than 10 in more than 90% samples were removed for subsequent analysis. About 44.17% (10,279 of 23,270) and 11.7% (2,757 of 23,650) of highly expressed genes were responsible for high salinity and temperature stresses respectively, and 1,088 genes were both responsive to the two stresses (Supplementary Fig. 15). For candidate genes with strong selective signals, a total of 29 genes were expressed under high salinity and temperature stress conditions, and 75.9% (22 of 29) and 58.6% (17 of 29) of all expressed genes were significantly increased or repressed their expression levels ($p < 0.05$, Fig. 4c), indicating genes under selection preferred to respond to these two environmental changes. Thirteen genes were responsive to both thermal and low salinity stresses, while three genes were not sensitive to these two stresses. Nine and four genes specifically responded to high salinity or temperature exposure, respectively. Our findings reveal that most of genes with strong selective signals were responsible for adaptation of estuarine oysters to environmental gradients especially for temperature and salinity.

Expansion of selective genes contribute to temperature and salinity adaptation. Among these selective regions, we found two tandem duplications for *Solute carrier family* including 10 copies of *Slc23a2* and four copies of *Monocarboxylate transporter 12* (*Mct12*, also known as *Slc16a12*) (Supplementary Fig. 14c), which located in the two peaks of chromosome 9 and showed high differentiations among three oyster populations, where average F_{ST} values were 0.81 and 0.76 respectively (Supplementary Fig. 14d). Genomic regions spanning *Slc23a2* gene families exhibited

317 extremely lower Tajima's *D* values in northern oysters, while these spanning *Mct12*
318 gene families had extremely lower Tajima's *D* values in southern oysters
319 (Supplementary Fig. 14e). These findings indicate that *Slc23a2* and *Mct12* genes were
320 under directional selection at northern and southern environments respectively, and
321 highlight the critical role of genes belonging to *Slc* families in adaptation of marine
322 species, such as porpoises and coral, to salinity and temperature gradients[17,19,55,56].
323 Moreover, ten copies of *Slc23a2* gene family belong to three orthogroups, where two
324 of them are annotated as purine permease (a: OG0011985 and c: OG0000489) and
325 another is annotated as uric acid transporter (OG0000633). Four copies of *Mct12* gene
326 family belong to one orthogroup annotated as purine efflux pump (OG0000571). All
327 of these orthogroups were extensively expanded in *C. ariakensis*, as well as in other
328 two estuarine oysters of *C. virginica* and *C. hongkongensis* in compare with marine
329 species of *C. gigas* (Supplementary Fig. 16). We found that three copies of *Slc23a2*
330 genes (*Slc23a2_a1*, *Slc23a2_a2* and *Slc23a2_c2*) in two expanded orthogroups
331 responded to both temperature and salinity challenges, while three copies of *Mct12*
332 (*Mct12_1*, *Mct12_2* and *Mct12_3*) genes were responsible for salinity challenges (Fig.
333 4c). The expanded selective genes of these two gene families responsive to the
334 temperature and salinity challenges mediates adaptive divergence among different
335 oyster populations. This finding highlighted the important role of gene duplication,
336 such as *Heat shock protein* (*Hsp*) gene family in Pacific oyster dwelling in the
337 intertidal zones[38,57], in adapting to species-specifically challenging habitats and
338 further mediate adaptive divergence among different subspecies or populations[58-61].
339
340 **Selective preference in upstream intergenic regions of environmentally**
341 **responsive genes.** We explored the divergence patterns of genomic variations,
342 including genic and intergenic (upstream and downstream) regions, and expression
343 responses of these 29 genes under selection to environmental changes by conducting
344 reciprocal transplant experiments. F_1 progenies bred from each of wild northern and
345 southern populations were acclimatized at both northern and southern habitats for

three months. F_{ST} were used to quantify the extent of divergence of different genomic regions and transcriptional changes upon transplantation between two habitats were used to qualify plastic changes of each selective genes.

Expression level of 29 candidate selective genes showed distinct population- and environment-specific patterns that ten and five genes were highly expressed at southern and northern oyster populations respectively ($p < 0.05$), while seven and four genes were highly expressed at northern and southern habitats respectively ($p < 0.05$), as well as three genes exhibited specifically higher or lower expression at native habitats ($p < 0.05$, Fig. 5a). Environment-specific selective genes correspondingly exhibited significant higher expression plasticity (19.15%) than that of population-specific selective genes when oysters were reciprocally transplanted between northern and southern habitats ($p = 0.0255$, Supplementary Fig. 17).

We further characterize the divergence of different genomic regions at genome-wide level, as well as total, population- and environment-specific candidate gene sets. At genome-wide profile, genic regions showed significantly higher F_{ST} values (mean $F_{ST\_genic} = 0.163$) than that of intergenic regions ($F_{ST\_intergenic} = 0.138$, $p < 0.001$, Wilcoxon signed-rank tests, Fig. 5b), indicating genic regions were substantially preferred to be under selection and evolved higher differentiation between northern and southern oysters. For all of the 29 candidate genes, both genic ($F_{ST} = 0.7745$) and intergenic ($F_{ST} = 0.7585$) regions were under strong selection, but no difference was detected between them ($p > 0.05$). However, at both genic and intergenic regions, environment-specific genes ($F_{ST} = 0.7816$) evolved significantly higher genomic divergence than population-specific genes ($F_{ST} = 0.7274$) between northern and southern oysters ($p = 0.0230$, Fig. 5b). Specifically, we found that the upstream intergenic regions of environment-specific genes exhibited significantly higher divergence than that of population-specific genes between northern and southern oysters ($F_{ST\_environment} = 0.7958$, $F_{ST\_population} = 0.7402$, $p = 0.01512$), while there was no difference of F_{ST} values at both genic and downstream intergenic regions between two gene sets ($p > 0.05$, Fig. 5c). Further, population-specific genes showed higher

375 divergence at genic region, while environment-specific genes evolved higher
376 divergence at upstream intergenic region (high divergence at genic region may be
377 resulted from the higher divergence at upstream intergenic region due to the 100-Kb
378 sliding window). Although the natural selection preferred to genic regions at
379 genome-wide level and for population-specific genes, environmentally responsive
380 genes were preference at upstream intergenic regions and under stronger selection,
381 which included critical regulatory elements such as promoter and enhancers that
382 potentially regulate their higher gene expression plasticity[24]. Our findings provide
383 insights into the genomic basis of plasticity that can evolve and has higher genetic
384 divergence by natural selection[23,24], facilitating the assessment of species'
385 vulnerabilities to climate-driven decline or extinction.

Methods

Genome sequencing. A wild adult estuarine oyster *Crassostrea ariakensis* was obtained from the northern China (Binzhou, Bohai Sea, China). Four tissue samples (gill, mantle, muscle and labial palp) were collected and flash-frozen in the liquid nitrogen. Genomic DNA extracted from the muscle was used to construct long insert genomic libraries. Oxford Nanopore Technologies' long-read sequencing platform generated ~184 Gb of data which corresponding to ~299× coverage. In addition, libraries with insert size of 350 bp were prepared and sequenced using Illumina HiSeq 4000 platform, which generated ~64 Gb of paired-end reads corresponding to ~105× coverage of the genome.

Genome size and heterozygosity. Trimmed Illumina short reads were used as input to calculate the distribution of k-mer copy number (KCN). We selected 21 to obtain the KCN distribution which showed two distinct peaks (Supplementary Fig. 1). The first peak (KCN = 45) represents the heterozygous single copy k-mer while the second peak (KCN = 90) represents the homozygous single copy k-mer in the genome. Genome size was estimated by the formula G = K_num/peak depth.

Contig-level assembly using long-read data. The Nanopore long reads, with a read N50 of 33,230 and a mean read length of 23,240 bp, were used for initial genome assembly. Error correction of clean data was conducted using Canu[62] v1.5, and then were assembled using Canu, WTDBG2[63] and SMARTdenovo tools. Quickmerge[64] v0.2.2 was used to join the three assemblies, and then was corrected for 3 cycles using long reads by Racon[65] and for 3 cycles using Illumina reads by Pilon[66] v1.22 with default parameters. The initial assembly of estuarine oyster genome was 613,892,480 bp in length with a contig N50 of 6,967,240 bp.

Chromosome-level assembly with Hi-C. The same genomic DNA extracted from the muscle was used to construct Illumina library, which was sequenced on the

Illumina HiSeq 4000 platform. A total of 106.34 Gb (173.22-fold coverage) of clean data was obtained, and 41.89% of all reads were truncated that containing enzyme cutting sites. HiC-Pro[67] was employed to evaluate the alignment efficiency and insert length distribution for valid interaction pairs (75.51%). Furthermore, the genome sequence contigs and scaffolds were interrupted in 50 kb length, which were then sorted and oriented into super scaffolds using LACHESIS[68] with the following parameters: CLUSTER_MIN_RE_SITES = 47, CLUSTER_MAX_LINK_DENSITY = 2, CLUSTER_NONINFORMATIVE_RATIO = 2, ORDER_MIN_N_RES_IN_TRUN = 40, ORDER_MIN_N_RES_IN_SHREDS = 41.

Genome evaluation. The Hi-C contact heatmap was used to assess the accuracy of the Hi-C assembly. The density of red color represents the number of Hi-C links between 100 kb windows on the pseudochromosomes of the final assembly. Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 with 978 conserved genes were used to assess the completeness and accuracy of estuarine oyster genome. The Illumina genomic reads were also aligned to the oyster genome to assess the completeness using sequence alignment tool BWA[69].

Repeat annotation. Transposable elements (TEs) were identified and classified using homology-based and *de novo*-based approaches. RepeatScout and LTR_FINDER was used to construct the *de novo* repeat libraries. The *de novo*-based library was further classified by PASTEClassifier[70] to obtain a consensus library, and combined with the repeat library of Repbase data. RepeatMasker[71] v4.0.5 was used to identify TEs in the estuarine oyster genome with the combined library.

Protein-coding genes annotation. We adopted three methods including *de novo*-based predictions, homology-based predictions and RNA-seq-based predictions to annotate the protein-coding genes of estuarine oyster *C. ariakensis*. For the RNA-seq-based prediction, RNA-seq data generated from four tissues (gill, mantle,

muscle and labial palp) using ONT long-reads sequencing platform was filtered to remove adaptors and then trimmed to remove low-quality bases. Clean reads were aligned to reference genome using TopHat2[72] and then assembled using Trinity[73]. Full transcriptome-based genome annotation was predicted using PASA[74] v2.2.2 software. For the *de novo* prediction, five *ab initio* gene prediction programs, including Genscan[75] v1.0, Augustus[76] v2.4, GlimmerHMM[77] v3.0.4, GeneID[78] v1.4 and SNAP[79], were used to predict genes in the repeat-masked genome. For homolog-based prediction, the protein sequences of 10 well-annotated species, including *Homo sapiens*, *Danio rerio*, *Aplysia californica*, *Strongylocentrotus purpuratus*, *C. gigas*, *C. virginica*, *Biomphalaria glabrata*, *Lingula anatina*, *Octopus bimaculoides* and *Mizuhopecten yessoensis*, were downloaded and aligned to the repeat-masked estuarine oyster genome using tblastn[80] with E-value ≤ 1E-05. We employed GeMoMa[81] v1.3.1 to predict gene models based on the alignment sequences. Finally, EVidenceModeler[82] (EVM) v1.1.1 was used to generate a weighted and non-redundant gene set by integrating all gene models predicted by the above three methods.

Homologous sequences in the genome were identified by genBlastA[83] v1.0.4 using the integrated gene set, and GeneWise[84] was used to identify pseudogenes. Transfer RNAs (tRNAs) were defined using tRNAscan-SE[85] v1.3.1 software with eukaryote default parameters. Micro RNA and Rrna were identified by Infernal BLASTN[86] against the Rfam[87] database v12.0.

Functional annotation of protein-coding genes was conducted by aligning them to the NCBI non-redundant protein[88] (NR), SwissProt[89], KOG[90] and TrEMBL[89] databases using BLAST[86] v2.2.31 with a maximal e-value of 1e-05. Domains were identified using HMMER[91] v3.0 to search against Pfam[92] databases. Gene set was mapped to Gene Ontology (GO) terms and KEGG pathway to identify the best match classification for each gene.

Whole-genome resequencing and mapping. We collected 264 wild oysters of *C.*

ariakensis from 11 estuarine areas (Fig. 3a), representing the major distribution range across north, middle and south Chinese coastlines[28,29]. Genomic DNA was isolated from gill of each oysters following the standard phenol-chloroform extraction procedure, and then was used to construct a library with an insert size of ~ 350 bp. Paired-end sequencing libraries were constructed according to the manufacturer's instructions (Illumina Inc., San Diego, CA, USA) and subsequently sequenced on the Illumina HiSeq X Ten Sequencer (Illumina Inc.). We obtained ~14.42 Gb of clean data for each sample, giving an average depth of 19.9× coverage (15-28×) (Supplementary Table 8). The 150-bp paired-end reads were mapped onto the *C. ariakensis* reference genome (PRJNA715058) with the Burrows-Wheeler Aligner v.0.7.8[69] using the default parameters (bwa mem –M -t 10 -T 20). Mapping data were then converted into the BAM format and sorted by SAMtools v.1.3.1[93], which was further used to remove duplicate reads. Read pair with the highest mapping quality was retained if multiple read pairs had identical external coordinates.

Genomic variation calling and annotation. The Genome Analysis Toolkit (GATK) v.3.7[94] module HaplotypeCaller was used to obtain high-quality variation calling of each sample. SNPs were further filtered with the parameter 'QD<2.0 || FS>60.0 || MQ<40.0'. Similarly, calling of INDELs was conducted and filtered using the command parameters as 'QD<2.0 || FS>60.0'. Filtered SNPs were annotated by the SnpEff[95] based on the *C. ariakensis* genome, and then were classified as variations in regions of exon, intron, splicing sites, and upstream and downstream intergenic regions, and in types of heterozygous and homozygous variations. To characterize the types of variations in northern and southern oysters, Plink[96] was used to filter the raw SNPs of each oyster populations using the parameters of MAF > 0.05 and Int > 0.8. The same SNPs were retrained and then were singly classified as homozygote or heterozygote in each oyster population (more than half individuals were the same type of variation in each oyster population). Variations in exons were further categorized as synonymous or non-synonymous SNPs. Two-sided two-sample Wilcoxon signed-rank

tests were conducted to test whether the ratios of genes with nonsynonymous variations were different between northern and southern geographic populations, using the function *wilcoxsign-test* in R package "coin".

Population genetic analysis. Plink[96] was used to filter the raw SNPs of all individuals using the parameters of MAF > 0.05 and Int > 0.8. Population structure was investigated using ADMIXTURE v.1.23[97] with default setting. The number of assumed genetic clusters *K* ranged from 2 to 5, and the optimum number of *K* was assessed by cross-validation (CV) errors. The individual-based neighbor-joining (NJ) phylogenetic tree was constructed using the MEGA[98] under the Kimura 2-Parameter model with 1000 bootstraps, and was then visualized using FigTree. We performed PCA for whole-genome SNPs of all 264 individuals using Eigensoft[99]. To evaluate linkage disequilibrium decay, the parameter r^2 between any two loci was calculated within each chromosome using Plink v.1.07[96] with the command (–ld-window-r2 0 –ld-window 99999 –ld-window-kb 500). The average r^2 values were calculated for each length of distance and the whole-genome LD was averaged across all chromosomes. The LD decay plot was depicted against the length of distance. Popgenome R package[100] was used to calculate Tajima's *D*, global F_{ST} and nucleotide diversity (π) using a 100-kb sliding window with the step size of 10-kb.

Demographic history of marine and estuarine oyster species. We implemented PMSC[101] to estimate dynamics of effective population size (*Ne*) and the possible divergence time over the past several million years ago (mya). A total of eight estuarine oysters (*C. ariakensis*) from northern (n = 3), middle (n = 2) and (n = 3) southern populations and three marine oysters (*C. gigas*)[11] with a high sequencing depth (*C. ariakensis*: 25~28×, *C. gigas*: ~20×) were used. To alleviate the probability of false positive, sequencing depth of SNPs was filtered with parameters: MinDepth = average depth/3, MaxDepth = average depth×2. The PSMC parameters were set as: -N25 -t15 -r5 -p '4 + 25*2 + 4 + 6' to estimate the historical *Ne*. The estimated

generation time (g) was set as 1 for both species, while mutation rates (μ) were calculated, following the formula $T_{\_divergence} = Ks/2\mu$, as 0.3×10^{-8} and 0.2×10^{-8} for *C. ariakensis* and *C. gigas*, respectively.

Detection of selective signals for adaptation to southern environments. To identify candidate selective signals potentially contributing to adaptation of oysters to southern environments, we calculated two pairs of population fixation statistics (F_{ST}) and selection statistics (Tajima's D), including a) north vs south and b) middle vs south, in a 100-kb sliding window with a step size of 10-kb. Genomic regions showing strong selective signals were defined as following: 1) regions showed top 1% F_{ST} values were overlapped in both comparison pairs; 2) regions located at the ravines of Tajima's D values along each chromosome in one of the three oyster populations.

Exposure to high temperature and salinity. To investigate environmental responses of estuarine oysters to challenges of elevated temperature and high salinity, we collected wild oysters and acutely exposed to different gradients of temperature of 20 □ and 37 □ for 6 hours and of salinity of 20 ‰ and 60 ‰ for 7 days, respectively. Gills from five oysters were individually sampled and immediately flash-frozen in liquid nitrogen for subsequent RNA-seq analysis.

Reciprocal transplantation experiments. Reciprocal transplantation experiments were described in our previously study[2]. Briefly, wild oysters derived from northern (Binzhou: BZ, Bohai Sea) and southern (Taishan: TS, East China Sea) environments were collected and used to reproduce F_1 generation within population. To potentially maintain an effective population size, a total of 80 mature male and female oysters were selected as parental individuals after excluding hermaphrodites by microscopic examination. Eggs were mixed and then divided into 40 beakers. Sperm from each of 40 male oysters were individually crossed with each beaker of mixed eggs, which warranted each sperm can fertilized with eggs from different female oysters. Zygotes

fathered by eight males were combined into one group. Five groups were reared to the D-shaped stage and then cultured in one nursery pond during larvae to spat stages. Two-month-old juvenile oysters from each of two populations were outplanted to two source habitats to test their responses to reciprocal transplantation. After three months of acclimation at northern and southern environments, we sampled gills of five oysters from each of population at both habitats in situ that gills were dissected out immediately on the boat and flash-frozen in liquid nitrogen for subsequent RNA-seq analysis.

RNA-seq analysis. Total RNA was isolated from gills sampled from acute stress experiment (high temperature and salinity) and reciprocal transplant experiment, using the RNAprep Pure Tissue Kit (Tiangen) following the manufacturer's protocol. The RNA integrity and concentration were examined by 1.2% gel electrophoresis and Nanodrop 2000 spectrophotometer, respectively. DNA contamination was removed with DNAse I treatment. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system. A total amount of 1 μg RNA per sample was used to construct sequencing libraries using NEBNext Ultra^{TM} RNA Library Prep Kit, and then were sequenced on an Illumina HiSeq 4000 platform to generate 150-bp paired-end raw reads. Clean data were obtained by removing reads containing adapter, reads containing ploy-N and low-quality reads. TopHat2[66] was used to map clean reads to the estuarine oyster *C. ariakensis* reference genome. StringTie v2.0 was used for reads assembly. Only reads with a perfect match or one mismatch were further analyzed and annotated. Gene expression levels were estimated by fragments per kilobase of transcript per million fragments mapped (FPKM). We employed DESeq2 to analysis differentially expressed genes (DEGs) between different populations at northern and southern environments. Genes with an adjusted *p*-value < 0.01 using the Benjamin and Hochberg's approach were assigned as DEGs. A hierarchical cluster analysis was performed to indicate expression level of candidate genes showing strong selective signals, using the *pheatmap* package in R

589 software.

590

591 ## Data availability

592 The genome, whole-genome re-sequencing and transcriptome datasets were deposited
593 in the Sequence Read Archive (SRA) database under the accession number
594 PRJNA715058.

Reference

1 Chou, C. *et al.* Increase in the range between wet and dry season precipitation. *Nature Geoscience* **6**, 263-267, doi:10.1038/ngeo1744 (2013).

2 Li, A. *et al.* Molecular and Fitness Data Reveal Local Adaptation of Southern and Northern Estuarine Oysters (Crassostrea ariakensis). *Frontiers in Marine Science* **7**, doi:10.3389/fmars.2020.589099 (2020).

3 Buroker, N. E., Hershberger, W. K. & Chew, K. K. Population Genetics of the Family Ostreidae. II. Interspecific Studies of the Genera *Crassostrea* and *Saccostrea*. *Marine Biology* **54**, 171-184 (1979).

4 Liu, J. X., Gao, T. X., Wu, S. F. & Zhang, Y. P. Pleistocene isolation in the Northwestern Pacific marginal seas and limited dispersal in a marine fish, Chelon haematocheilus (Temminck & Schlegel, 1845). *Molecular ecology* **16**, 275-288, doi:10.1111/j.1365-294X.2006.03140.x (2007).

5 Dong, Y. *et al.* The Impact of Yangtze River Discharge, Ocean Currents and Historical Events on the Biogeographic Pattern of *Cellana toreuma* along the China Coast. *PloS ONE* **7**, e36178, doi:10.1371/journal.pone.0036178.g001 (2012).

6 Ni, G., Li, Q., Kong, L. & Zheng, X. D. Phylogeography of bivalve *Cyclina sinensis*: testing the historical glaciations and Changjiang River outflow hypotheses in northwestern Pacific. *PLoS ONE* **7**, e49487, doi:10.1371/journal.pone.0049487.g001 (2012).

7 Ni, G., Kern, E., Dong, Y. W., Li, Q. & Park, J. K. More than meets the eye: The barrier effect of the Yangtze River outflow. *Molecular ecology* **26**, 4591-4602,

doi:10.1111/mec.14235 (2017).

8 Sanford, E. & Kelly, M. W. Local adaptation in marine invertebrates. *Annual review of marine science* **3**, 509-535, doi:10.1146/annurev-marine-120709-142756 (2011).

9 Somero, G. N. The physiology of global change: linking patterns to mechanisms. *Annual review of marine science* **4**, 39-61, doi:10.1146/annurev-marine-120710-100935 (2012).

10 Miller, A. D. *et al.* Local and regional scale habitat heterogeneity contribute to genetic adaptation in a commercially important marine mollusc (Haliotis rubra) from southeastern Australia. *Molecular ecology* **28**, 3053-3072, doi:10.1111/mec.15128 (2019).

11 Li, L. *et al.* Divergence and plasticity shape adaptive potential of the Pacific oyster. *Nat Ecol Evol* **2**, 1751-1760, doi:10.1038/s41559-018-0668-2 (2018).

12 Stern, D. B. & Lee, C. E. Evolutionary origins of genomic adaptations in an invasive copepod. *Nat Ecol Evol* **4**, 1084-1094, doi:10.1038/s41559-020-1201-y (2020).

13 Kenkel, C. D. & Matz, M. V. Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nature Ecology & Evolution* **1**, 0014, doi:10.1038/s41559-016-0014 (2016).

14 Ho, W. C. & Zhang, J. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. *Nature communications* **9**, 350, doi:10.1038/s41467-017-02724-5 (2018).

15 Gienapp, P., Teplitsky, C., Alho, J. S., Mills, J. A. & Merila, J. Climate change and evolution: disentangling environmental and genetic responses. *Molecular ecology* **17**,

167-178, doi:10.1111/j.1365-294X.2007.03413.x (2008).

16 Pfennig, D. W. *et al.* Phenotypic plasticity's impacts on diversification and speciation. *Trends in ecology & evolution* **25**, 459-467, doi:10.1016/j.tree.2010.05.006 (2010).

17 Barrio, A. M. *et al.* The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife* **5**, e12081, doi:10.7554/eLife.12081.001 (2016).

18 Zong, S. B., Li, Y. L. & Liu, J. X. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol Biol Evol*, doi:10.1093/molbev/msaa290 (2020).

19 Zhou, X. *et al.* Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nature communications* **9**, 1276, doi:10.1038/s41467-018-03722-x (2018).

20 Sandoval-Castillo, J. *et al.* Adaptation of plasticity to projected maximum temperatures and across climatically defined bioregions. *Proceedings of the National Academy of Sciences* **117**, 17112-17121, doi:10.1073/pnas.1921124117/-/DCSupplemental (2020).

21 Eierman, L. E. & Hare, M. P. Reef-Specific Patterns of Gene Expression Plasticity in Eastern Oysters (Crassostrea virginica). *The Journal of heredity* **107**, 90-100, doi:10.1093/jhered/esv057 (2016).

22 Bernal, M. A. *et al.* Species-specific molecular responses of wild coral reef fishes during a marine heatwave. *Science Advances* **6**, eaay3423 (2020).

23 Kelly, M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philosophical Transactions of the Royal Society B:*

Biological Sciences **374**, 20180176, doi:10.1098/rstb.2018.0176 (2019).

24 Grishkevich, V. & Yanai, I. The genomic determinants of genotype x environment interactions in gene expression. *Trends in genetics : TIG* **29**, 479-487, doi:10.1016/j.tig.2013.05.006 (2013).

25 Ahlgren, J., Yang, X., Hansson, L. A. & Bronmark, C. Camouflaged or tanned: plasticity in freshwater snail pigmentation. *Biology Letters* **9**, 20130464-20130464, doi:10.1098/rsbl.2013.0464 (2013).

26 Li, A., Li, L., Song, K., Wang, W. & Zhang, G. Temperature, energy metabolism, and adaptive divergence in two oyster subspecies. *Ecology and evolution* **7**, 6151-6162, doi:10.1002/ece3.3085 (2017).

27 Gagnaire, P.-A. *et al.* Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters *Crassostrea gigas* and *Crassostrea angulata*. *Genome biology and evolution* **10**, 2518-2534, doi:10.12770/dbf64e8d-45dd-437f-b734-00b77606430a10.1093/gbe/evy194 (2018).

28 Wang, H., Guo, X., Zhang, G. & Zhang, F. Classification of jinjiang oysters Crassostrea rivularis (Gould, 1861) from China, based on morphology and phylogenetic analysis. *Aquaculture* **242**, 137-155, doi:10.1016/j.aquaculture.2004.09.014 (2004).

29 Zhou, M. F. & Allen, S. K. A review of published work on *Crassostrea ariakensis*. *Journal of Shellfish Research* **22**, 1-20 (2003).

30 Wang, H. *et al.* Distribution of *Crassostrea ariakensis* in China. *Journal of Shellfish Research* **25**, 789-790 (2006).

683 31 Zhang, Q., Allen, S. K., Jr. & Reece, K. S. Genetic variation in wild and hatchery
684 stocks of Suminoe Oyster (Crassostrea ariakensis) assessed by PCR-RFLP and
685 microsatellite markers. *Marine biotechnology* **7**, 588-599,
686 doi:10.1007/s10126-004-5105-7 (2005).
687 32 Xiao, J., Cordes, J. F., Wang, H., Guo, X. & Reece, K. S. Population genetics of
688 Crassostrea ariakensis in Asia inferred from microsatellite markers. *Marine Biology*
689 **157**, 1767-1781, doi:10.1007/s00227-010-1449-x (2010).
690 33 Kim, W.-J. *et al.* Mitochondrial DNA sequence analysis from multiple gene fragments
691 reveals genetic heterogeneity of Crassostrea ariakensis in East Asia. *Genes Genom.*
692 **36**, 611-624, doi:10.1007/s13258-014-0198-5 (2014).
693 34 Liu, X. *et al.* Transcriptome and Gene Coexpression Network Analyses of Two Wild
694 Populations Provides Insight into the High-Salinity Adaptation Mechanisms of
695 Crassostrea ariakensis. *Marine biotechnology* **21**, 596-612,
696 doi:10.1007/s10126-019-09896-9 (2019).
697 35 Bai, C. M. *et al.* Chromosomal-level assembly of the blood clam, Scapharca (Anadara)
698 broughtonii, using long sequence reads and Hi-C. *GigaScience* **8**,
699 doi:10.1093/gigascience/giz067 (2019).
700 36 Peng, J. *et al.* Chromosome-level analysis of the Crassostrea hongkongensis genome
701 reveals extensive duplication of immune-related genes in bivalves. *Molecular ecology*
702 *resources* **20**, 980-994, doi:10.1111/1755-0998.13157 (2020).
703 37 Song, H. *et al.* The hard clam genome reveals massive expansion and diversification
704 of inhibitors of apoptosis in Bivalvia. *BMC Biol* **19**, 15,

doi:10.1186/s12915-020-00943-9 (2021).

38 Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54, doi:10.1038/nature11413 (2012).

39 Hu, Y. *et al.* Spatial patterns and conservation of genetic and phylogenetic diversity of wildlife in China. *science advances* **7**, eabd5725, doi:10.1126/sciadv.abd5725 (2021).

40 Li, C. *et al.* Genome sequences reveal global dispersal routes and suggest convergent genetic adaptations in seahorse evolution. *Nature communications* **12**, 1094, doi:10.1038/s41467-021-21379-x (2021).

41 Wang, H., Qian, L., Liu, X., Zhang, G. & Guo, X. Classification of a common cupped oyster from southern China. *Journal of Shellfish Research* **29**, 857-866, doi:10.2983/035.029.0420 (2010).

42 Wang, H., Zhang, G., Liu, X. & Guo, X. Classification of Common Oysters from North China. *Journal of Shellfish Research* **27**, 495-503, doi:10.2983/0730-8000(2008)27[495:COCOFN]2.0.CO (2008).

43 Raeymaekers, J. A. M. *et al.* Adaptive and non-adaptive divergence in a common landscape. *Nature communications* **8**, 267, doi:10.1038/s41467-017-00256-6 (2017).

44 Kimura, M. Paleogeography of the Ryukyu Islands. *Tropics* **10**, 5-24 (2000).

45 Ren, J., Liu, X., Jiang, F., Guo, X. & Liu, B. Unusual conservation of mitochondrial gene order in *Crassostrea* oysters: evidence for recent speciation in Asia. *BMC evolutionary biology* **10**, 394, doi:10.1186/1471-2148-10-394 (2010).

46 Qin, Y., Zhao, Y. & Zhao, S. *Geology of the Bohai Sea*. (Science Press, 1985).

47 Skliris, N. *et al.* Salinity changes in the World Ocean since 1950 in relation to changing

surface freshwater fluxes. *Climate Dynamics* **43**, 709-736, doi:10.1007/s00382-014-2131-7 (2014).

48 Mekonnen, M. M. & Hoekstra, A. Y. Four billion people facing severe water scarcity. *Science Advances* **2**, e1500323 (2016).

49 Ribeiro, C. A., Balestro, F., Grando, V. & Wajner, M. Isovaleric acid reduces Na+, K+-ATPase activity in synaptic membranes from cerebral cortex of young rats. *Cellular and molecular neurobiology* **27**, 529-540, doi:10.1007/s10571-007-9143-3 (2007).

50 Huang, Y., Niwa, J., Sobue, G. & Breitwieser, G. E. Calcium-sensing receptor ubiquitination and degradation mediated by the E3 ubiquitin ligase dorfin. *The Journal of biological chemistry* **281**, 11610-11617, doi:10.1074/jbc.M513552200 (2006).

51 Vienken, H. *et al.* Characterization of cholesterol homeostasis in sphingosine-1-phosphate lyase-deficient fibroblasts reveals a Niemann-Pick disease type C-like phenotype with enhanced lysosomal Ca(2+) storage. *Scientific reports* **7**, 43575, doi:10.1038/srep43575 (2017).

52 Pagano, M. *et al.* Insights into the residence in lipid rafts of adenylyl cyclase AC8 and its regulation by capacitative calcium entry. *Am J Physiol Cell Physiol* **296**, C607–C619, doi:10.1152/ajpcell.00488.2008.-Adenylyl (2009).

53 Weinman, E. J., Dubinsky, W. P. & Shenolikar, S. Reconstitution of cAMP-Dependent Protein Kinase Regulated Renal Na^+-H^+ Exchanger. *J. Membr. Biol.* **101**, 11-18 (1988).

54 Fonteles, M. C., Greenberg, R. N., Monteiro, H. S. A., Currie, M. G. & Forte, L. R. Natriuretic and kaliuretic activities of guanylin and uroguanylin in the isolated perfused

749 rat kidney. *The American Journal of Physiology* **F191-F197** (1998).

750 55 Kenkel, C. D., Meyer, E. & Matz, M. V. Gene expression under chronic heat stress in
751 populations of the mustard hill coral (*Porites astreoides*) from different thermal
752 environments. *Molecular ecology* **22**, 4322-4334, doi:10.1111/mec.12390 (2013).

753 56 Hoglund, P. J., Nordstrom, K. J., Schioth, H. B. & Fredriksson, R. The solute carrier
754 families have a remarkably long evolutionary history with the majority of the human
755 families present before divergence of Bilaterian species. *Mol Biol Evol* **28**, 1531-1541,
756 doi:10.1093/molbev/msq350 (2011).

757 57 Guo, X., He, Y., Zhang, L., Lelong, C. & Jouaux, A. Immune and stress responses in
758 oysters with insights on adaptation. *Fish & shellfish immunology* **46**, 107-119,
759 doi:10.1016/j.fsi.2015.05.018 (2015).

760 58 Li, A., Li, L., Wang, W., Song, K. & Zhang, G. Transcriptomics and Fitness Data
761 Reveal Adaptive Plasticity of Thermal Tolerance in Oysters Inhabiting Different Tidal
762 Zones. *Front Physiol* **9**, 825, doi:10.3389/fphys.2018.00825 (2018).

763 59 Zhang, G. *et al.* Molecular Basis for Adaptation of Oysters to Stressful Marine
764 Intertidal Environments. *Annual review of animal biosciences* **4**, 357-381,
765 doi:10.1146/annurev-animal-022114-110903 (2016).

766 60 Li, A., Li, L., Wang, W. & Zhang, G. Evolutionary trade-offs between baseline and
767 plastic gene expression in two congeneric oyster species. *Biology Letters* **15**,
768 20190202, doi:10.1098/rsbl.2019.0202 (2019).

769 61 Ghaffari, H., Wang, W., Li, A., Zhang, G. & Li, L. Thermotolerance Divergence
770 Revealed by the Physiological and Molecular Responses in Two Oyster Subspecies of

Crassostrea gigas in China. *Front Physiol* **10**, 1137, doi:10.3389/fphys.2019.01137 (2019).

62 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome research* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).

63 Jayakumar, V. & Sakakibara, Y. Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Briefings in bioinformatics* **20**, 866-876, doi:10.1093/bib/bbx147 (2019).

64 Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic acids research* **44**, e147, doi:10.1093/nar/gkw654 (2016).

65 Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research* **27**, 737-746, doi:10.1101/gr.214270.116 (2017).

66 Walker, B. J. *et al.* Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE* **9**, e112963, doi:10.1371/journal.pone.0112963.g001 (2014).

67 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).

68 Burton, J. N. *et al.* Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nature biotechnology* **31**, 1119-1125, doi:10.1038/nbt.2727 (2013).

69 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

70 Hoede, C. *et al.* PASTEC: An Automatic Transposable Element Classification Tool. *PLoS ONE* **9**, e91929, doi:10.1371/journal.pone.0091929.t001 (2014).

71 Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10, doi:10.1002/0471250953.bi0410s25 (2009).

72 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36 (2013).

73 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**, 644-652, doi:10.1038/nbt.1883 (2011).

74 Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M. & Buell, C. R. Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC genomics* **7**, 327, doi:10.1186/1471-2164-7-327 (2006).

75 Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**, 78-94 (1997).

76 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225, doi:10.1093/bioinformatics/btg1080 (2003).

77 Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878-2879, doi:10.1093/bioinformatics/bth315 (2004).

78 Alioto, T., Blanco, E., Parra, G. & Guigo, R. Using geneid to Identify Genes. *Curr Protoc Bioinformatics* **64**, e56, doi:10.1002/cpbi.56 (2018).

79 Korf, I. Gene finding in novel genomes. *Bmc Bioinformatics* **5**, 59 (2004).

80 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**, 3389-3402 (1997).

81 Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O. & Grau, J. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *Bmc Bioinformatics* **19**, 189, doi:10.1186/s12859-018-2203-5 (2018).

82 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome biology* **9**, R7, doi:10.1186/gb-2008-9-1-r7 (2008).

83 She, R., Chu, J. S., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome research* **19**, 143-149, doi:10.1101/gr.082081.108 (2009).

84 Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome research* **14**, doi:10.1101/ (2004).

85 Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research* **25**, 955-964 (1997).

86 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *Journal of molecular biology* **215**, 403-410 (1990).

87 Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic acids research* **33**, D121-124, doi:10.1093/nar/gki081 (2005).

837 88 Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the functional
838 annotation of proteins. *Nucleic acids research* **39**, D225-229,
839 doi:10.1093/nar/gkq1189 (2011).
840 89 Boeckmann, B. *et al.* The SWISS-PROT protein knowledgebase and its supplement
841 TrEMBL in 2003. *Nucleic acids research* **31**, 365-370, doi:10.1093/nar/gkg095 (2003).
842 90 Tatusov, R. L. *et al.* The COG database: new developments in phylogenetic
843 classification of proteins from complete genomes. *Nucleic acids research* **29**, 22-28
844 (2001).
845 91 Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology
846 search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic acids*
847 *research* **41**, e121, doi:10.1093/nar/gkt263 (2013).
848 92 Finn, R. D. *et al.* Pfam: clans, web tools and services. *Nucleic acids research* **34**,
849 D247-251, doi:10.1093/nar/gkj149 (2006).
850 93 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
851 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
852 94 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
853 analyzing next-generation DNA sequencing data. *Genome research* **20**, 1297-1303,
854 doi:10.1101/gr.107524.110 (2010).
855 95 Cingolani, P. *et al.* A program for annotating and predicting the effects of single
856 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster
857 strain w1118; iso-2; iso-3. *Fly* **6**, 80-92, doi:10.4161/fly.19695 (2012).
858 96 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based

linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).

97 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655-1664, doi:10.1101/gr.094052.109 (2009).

98 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).

99 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909, doi:10.1038/ng1847 (2006).

100 Pfeifer, B., Wittelsburger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* **31**, 1929-1936, doi:10.1093/molbev/msu136 (2014).

101 Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496, doi:10.1038/nature10231 (2011).

Acknowledgements

L.L. is supported by the National Key R&D Program of China (No. 2018YFD0900304) and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA23050402). A.L. is supported by the Distinguished Young Scientists Research Fund of Key Laboratory of Experimental Marine Biology, Chinese Academy of Sciences (No. KLEMB-DYS04) and by China Postdoctoral Science Foundation (No. 2019TQ0324). A.L. and L.L. are supported by Key Deployment Project of Centre for Ocean Mega-Research of Science, Chinese Academy of Sciences (No. COMS2019Q06). L.L. is also supported by the National Natural Science Foundation of China (No. 31572620) and the Technology and the Modern Agro-industry Technology Research System (No. CARS-49). We thank X. Wang, Z. Jia, Z. She, Y. Zhang, Z. Yu, W. Quan, Z. Zeng and Y. Ning for sampling collection, and B. Yin and J. Qi for information on marine currents.

Author contributions

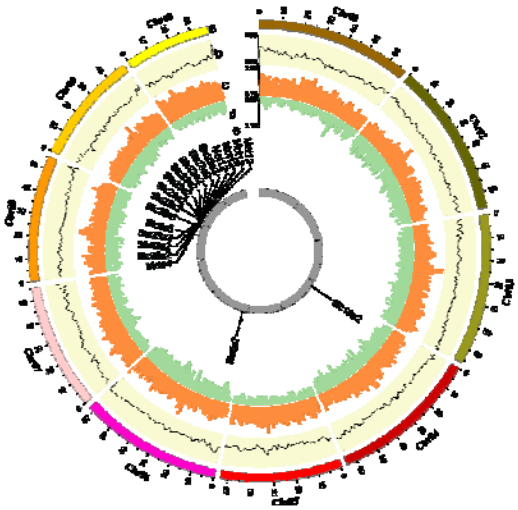
L.L., G.Z. and X.G. conceived the study and participated in final data analysis, interpretation and drafting the manuscript. A.L. carried out the data analysis and drafted the manuscript. H.D., A.L., H.C., X.L. and H.Z. contributed to the selective sweep analysis. A.L., Z.Z., K.Z. and C.W. collected and sampled oyster specimens. A.L. and W.W. produced the F_1 progeny. A.L., L.L., X.G. and G.Z. revised the manuscript. All authors approved the manuscript for publication. The authors declare no competing interests.

Competing interests

The authors declare no competing interests.

a

b



c

Characteristic	Statistic
Assembled genome size	613.89 Mb
Number of chromosomes	10
Number of contigs	630
Contig N50 length	6.97 Mb
Number of scaffolds	416
Scaffold N50 length	62.26 Mb
Total gap length in the assembly	21.50 Kb
Percentage GC	33.41%
Repeat content	54.14%
Number of protein coding genes	29,631
Complete BUSCOs	92.54%

Fig. 1 | High-quality assembly of the genome of estuarine oyster *Crassostrea ariakensis*. **a**, Estuarine oyster (photograph by Lumin Qian). **b**, CIRCOS plot showing the distribution of GC content, transposable elements (TE), coding sequences (CDS) and candidate genes (*solute carrier families*) surrounding selective sweep signals (see **Fig. 4**) in each chromosome of the *C. ariakensis* genome. **c**, Summary of the *C. ariakensis* genome assembly and gene annotation statistics.

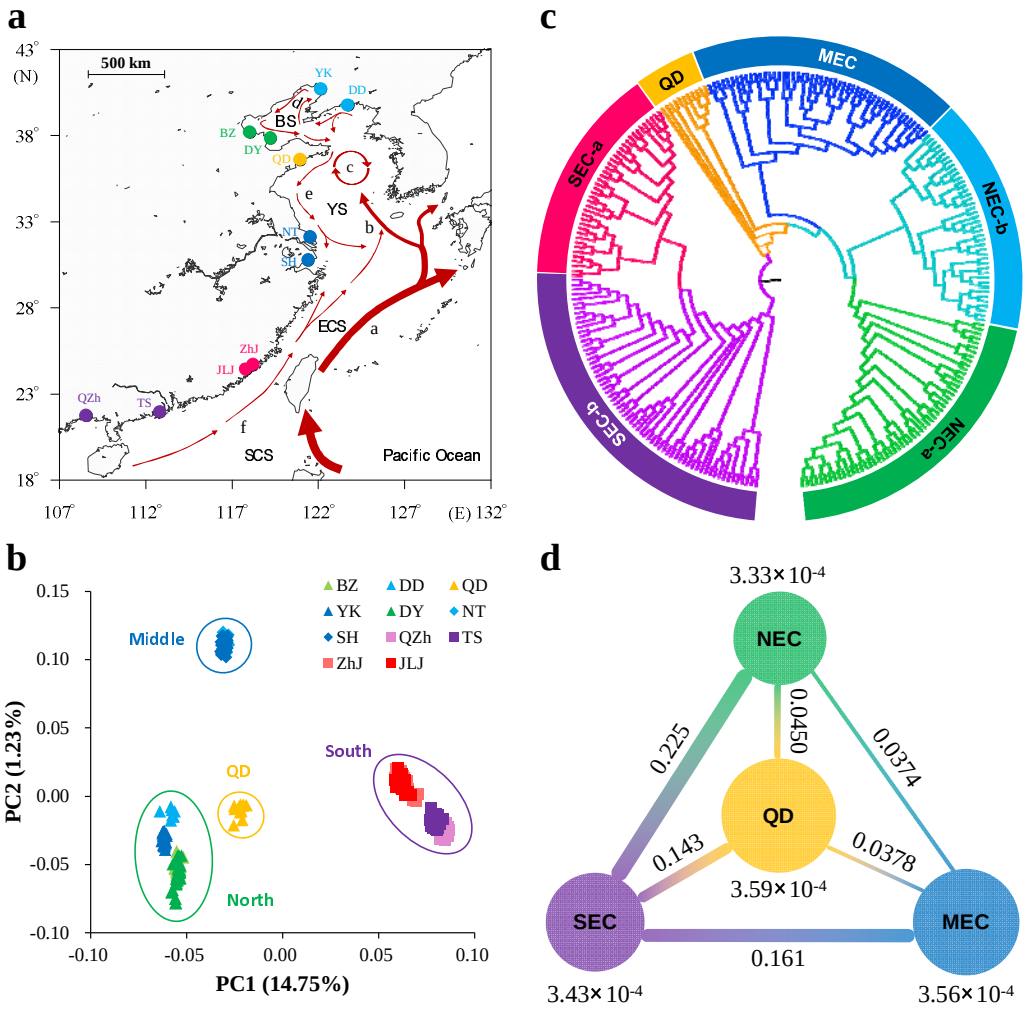


Fig. 2 | Geographic distribution and genetic structure of the *Crassostrea ariakensis* across the estuaries of China. a, Sampling locations of 264 wild estuarine oysters across 11 estuaried in China. The red and arrowed curves represent ocean currents in summer. a: Kuroshio current, b: Yellow Sea Warm Current, c: Yellow Sea Cold Water Mass, d: Bohai Sea Circulation, e: China Coastal Current, f: South China Sea Warm Current. SCS: South China Sea, ECS: East China Sea, YS: Yellow Sea, BS: Bohai Sea. DD: Dandong, YK: Yingkou, BZ: Binzhou, DY: Dongying, QD: Qingdao, NT: Nantong, SH: Shanghai, JLJ: Jiulongjiang, ZhJ: Zhangjiang, TS: Taishan, QZh: Qinzhou. **b**, Plots of principal components 1 and 2 of the 264 oyster individuals. **c**, Phylogenetic tree of estuarine oysters inferred from whole-genome SNPs by the neighbour-joining (NJ) method. NEC-a: north estuaries of China, including BZ and DY, NEC-b: DD and YK, MEC: middle estuaries of China, including NT and SH, SEC-a: south estuaries of China, including JLJ and ZhJ, and SEC-b: TS and QZh. **d**, Nucleotide diversity and genetic divergence across the four populations. The value under the circle is nucleotide diversity (π) for the oyster population, and values between population pairs indicate genetic divergence (F_{ST}).

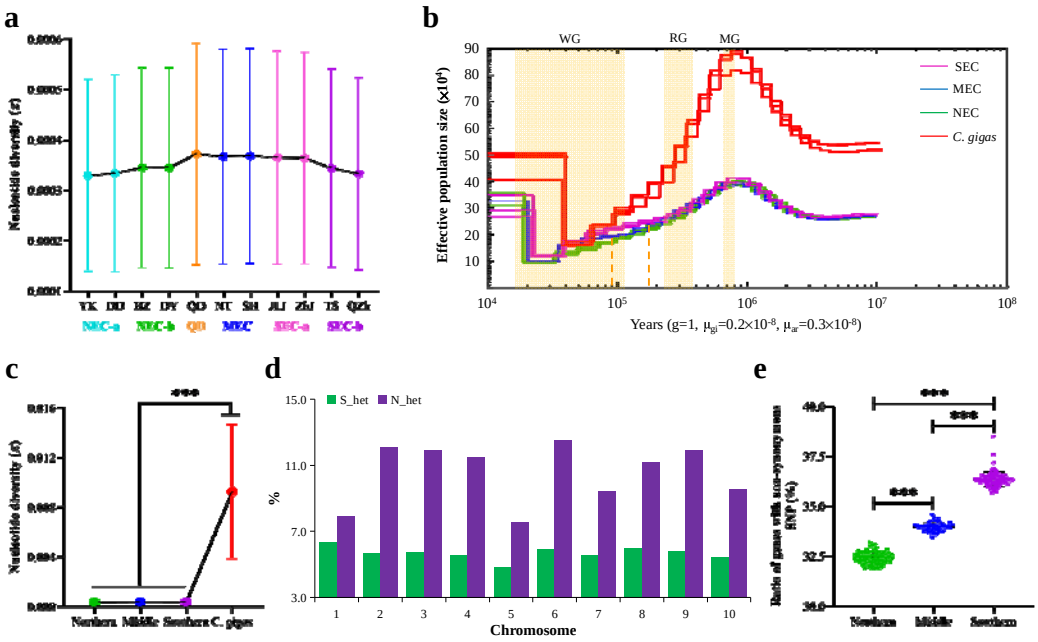


Fig. 3 | Effects of gene flow, historical glaciation and natural selection on population structure of *C. ariakensis*. **a**, Nucleotide diversity (π) of 11 oyster populations of *C. ariakensis*. **b**, Demographic histories of marine oyster species *C. gigas* (gi) and estuarine oyster species *C. ariakensis* (ar) including southern, middle and northern populations (SEC, MEC and NEC), inferred by the PSMC model. The period of the Mindel glaciation (MG, 0.68~0.80 mya), Riss glaciation (MG, 0.24~0.37 mya) and Würm glaciation (WG, 10,000~120,000 years ago) were shaded by pink. **c**, The ratios of SNPs showing heterozygous in northern oysters but homozygous in southern oysters (N_het) and those of SNPs showing homozygous in northern oysters but heterozygous in southern oysters (S_het) across 10 chromosomes. **d**, The ratio of genes with non-synonymous SNPs in three oyster populations. Asterisks indicate significant difference (*** $p < 0.001$).

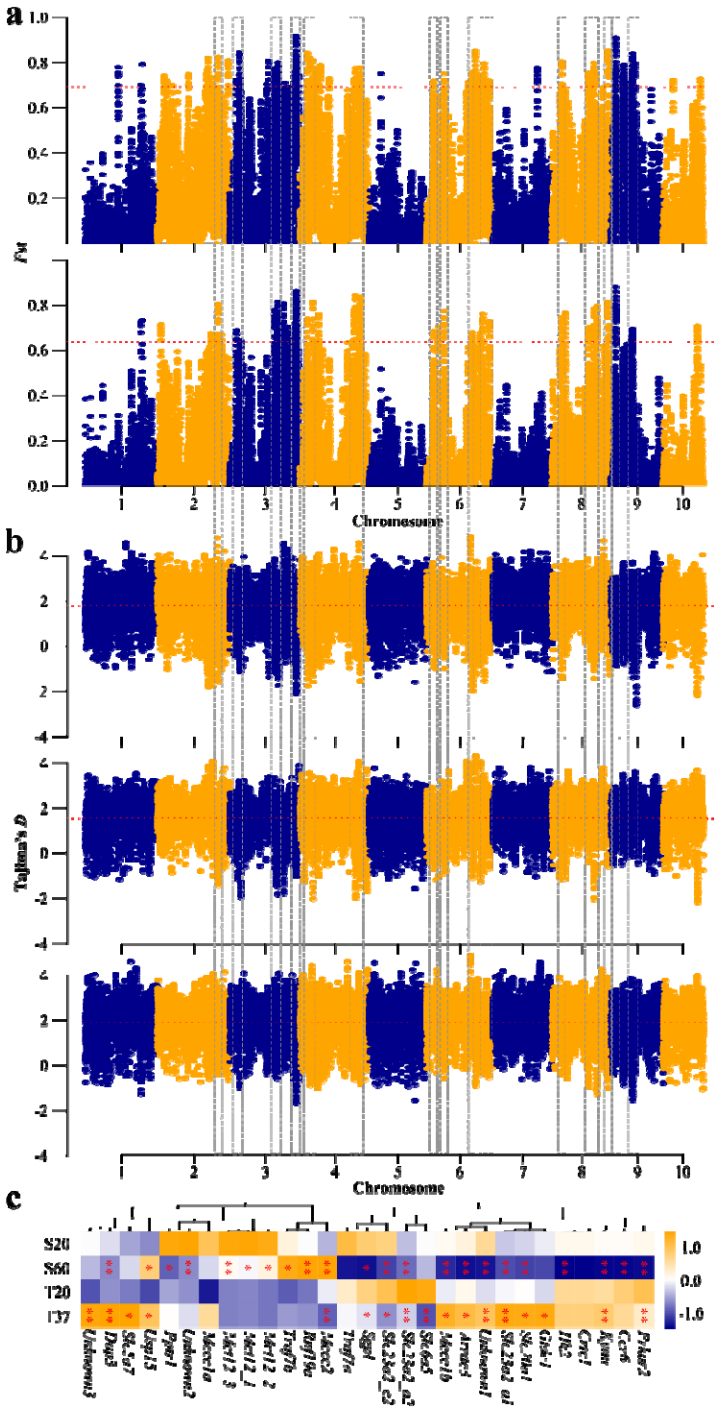


Fig. 4 | Genome-wide distribution of selective sweep signals and transcriptional responses of selective genes to stresses. a, Global F_{ST} values (top 1%, red lines) in two population pairs including northern versus southern (up) and middle versus southern (bottom). **b**, Global Tajima's *D* values in northern (up), middle (middle) and southern (bottom) oyster populations. **c**, Expression level of genes under selection in estuarine oysters when exposed to thermal (6 hours under 37 □) and high-salt (7 days under 60 ‰) conditions. Asterisks indicate significant difference (* $p < 0.05$, ** $p < 0.01$).

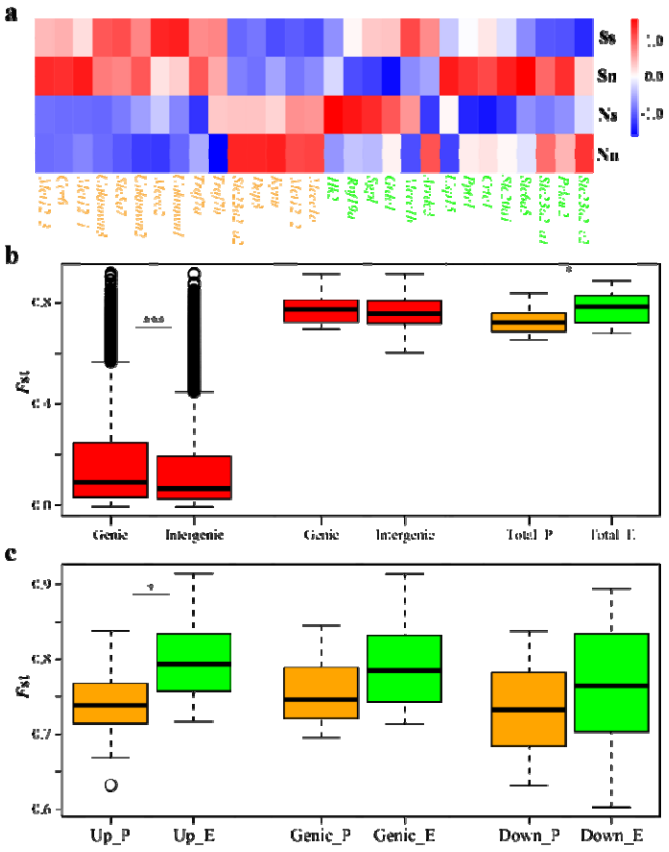


Fig. 5 | Transcriptional and genomic divergence of genes with selective signals. a, Expression level of selective genes in northern (N) and southern (S) oyster populations (capital letters) acclimated at northern (n) and southern (s) habitats (lowercase letters), which showed population- (orange) and environment- (green) specific expression patterns. **b**, Genetic divergence (F_{ST}) for genic and intergenic regions at genome-wide level (left) and 29 candidate genes with selective signals (middle), as well as for population- (P) and environment- (E) specific genes at both genic and intergenic regions (Total, right). **c**, Genetic divergence (F_{ST}) for intergenic [Up: upstream (left) and Down: downstream (right)] and genic (middle) regions of population- (P) and environment- (E) specific genes. Asterisks indicate significant difference (* $p < 0.05$, *** $p < 0.001$).