

## Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity

Ao Li<sup>1,2,8</sup>, He Dai<sup>3,8</sup>, Ximing Guo<sup>4,8</sup>, Ziyang Zhang<sup>1,5,6</sup>, Kexin Zhang<sup>1,5,6</sup>, Chaogang Wang<sup>1,5,6</sup>, Xinxing Wang<sup>1,5,6</sup>, Wei Wang<sup>1,5,7</sup>, Hongju Chen<sup>3</sup>, Xumin Li<sup>3</sup>, Hongkun Zheng<sup>3</sup>, Li Li<sup>1,5,6,7</sup>✉ & Guofan Zhang<sup>1,2,7</sup>✉

Understanding the roles of genetic divergence and phenotypic plasticity in adaptation is central to evolutionary biology and important for assessing adaptive potential of species under climate change. Analysis of a chromosome-level assembly and resequencing of individuals across wide latitude distribution in the estuarine oyster (*Crassostrea ariakensis*) revealed unexpectedly low genomic diversity and population structures shaped by historical glaciation, geological events and oceanographic forces. Strong selection signals were detected in genes responding to temperature and salinity stress, especially of the expanded *solute carrier* families, highlighting the importance of gene expansion in environmental adaptation. Genes exhibiting high plasticity showed strong selection in upstream regulatory regions that modulate transcription, indicating selection favoring plasticity. Our findings suggest that genomic variation and population structure in marine bivalves are heavily influenced by climate history and physical forces, and gene expansion and selection may enhance phenotypic plasticity that is critical for the adaptation to rapidly changing environments.

<sup>1</sup>CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. <sup>2</sup>Laboratory for Marine Biology and Biotechnology, Pilot National Laboratory for Marine Science and Technology, Qingdao, China. <sup>3</sup>Biomarker Technologies Corporation, Beijing, China. <sup>4</sup>Haskin Shellfish Research Laboratory, Department of Marine and Coastal Sciences, Rutgers University, Port Norris, NJ, USA. <sup>5</sup>Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao, China. <sup>6</sup>University of Chinese Academy of Sciences, Beijing, China. <sup>7</sup>National and Local Joint Engineering Key Laboratory of Ecological Mariculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China. <sup>8</sup>These authors contributed equally: Ao Li, He Dai, Ximing Guo. ✉email: [lili@qdio.ac.cn](mailto:lili@qdio.ac.cn); [gzhang@qdio.ac.cn](mailto:gzhang@qdio.ac.cn)

Climate change poses a serious threat to global biodiversity and the stability of ecosystems. Oceans bear the blunt of climate change as they absorb most of the heat energy from the sun and about one-third of the carbon dioxide created by human activities<sup>1</sup>. World oceans become warmer and more acidic. Sea levels are rising, and ocean circulation patterns are changing. With the rising of global temperature, dry regions are getting drier, while wet regions become wetter<sup>2</sup>, increasing salinity difference among estuaries<sup>3</sup>. These rapid changes in world's oceans are having a great impact on marine ecosystems, much of which remains poorly understood. The impact may vary among different organisms, and changes in keystone species may have a large effect on stability of an ecosystem. Understanding how organisms adapt to environmental shifts is fundamental to evolutionary biology and critical in assessing their adaptive potential under climate change.

The ability of an organism to survive and adapt to sudden environmental changes depends on available genetic variation and phenotypic plasticity. The standing genetic variation of a species is the outcome of its evolutionary history shaped by mutation, selection and genetic drifts<sup>4–6</sup>. Past and on-going climate and geological events may leave signatures of selection or bottleneck in the genome and influence the geographic distribution of genetic variation<sup>7–9</sup>. Many recent studies have revealed the important role of phenotypic plasticity in environmental adaptation<sup>10–15</sup>. Phenotypic plasticity may buffer against sudden environmental changes and provides time for adaptation to occur<sup>12–14,16</sup>. Phenotypic plasticity is particularly pronounced and important for sessile organisms inhabiting estuarine and intertidal zones that cannot use avoidance to cope with wide fluctuations in environmental conditions<sup>11,12,17</sup>, although the genetic basis of the enhanced plasticity is not clear<sup>18,19</sup>.

Oysters are keystone species in coastal and estuarine ecology providing critical ecological services as filter-feeders and habitat engineers. As sessile bivalves thriving in the coastal zone, oysters are well adapted to highly dynamic environmental conditions, as reflected in their high genetic diversity and plasticity<sup>12,20–23</sup>, making them interesting models for studying responses to climate change. In addition to threats from climate change, oyster populations worldwide have been decimated by over-fishing, habitat destruction and infectious diseases<sup>24,25</sup>. The estuarine (Jinjiang or Suminoe) oyster (*Crassostrea ariakensis*, Fig. 1a) is broadly distributed in estuaries of East Asia over large latitude and experiences wide ranges of temperature and salinity<sup>24–28</sup>. Unlike the Pacific oyster (*Crassostrea gigas*), a sympatric sister-species that is highly abundant in high-salinity waters of higher latitude, the estuarine oyster is only found in lower-salinity estuaries within its wide distribution range and at low abundance. The evolutionary pathways leading to the stark differences in abundance, distribution and salinity preference are not clear but important to our understanding of adaptive evolution.

While genetic changes can be inferred with genetic markers, whole-genome analysis is essential for exploring all genetic variation and identifying genes and selection events that are critical for adaption. Using multi-omic analyses, we previously showed that the Pacific oyster has a highly polymorphic genome with extensive expansion of environment-responsive genes, and plasticity is positively correlated with local adaptation<sup>12,23</sup>. In this study, we produced a chromosome-level assembly of the estuarine oyster genome, re-sequenced 264 wild oysters collected from 11 estuaries, and conducted transcriptomic studies of environmental response to understand its genetic variation, population structure, phenotypic plasticity and genomic signatures of selection or bottleneck that may be linked to environmental changes. Our results show that the estuarine oyster has significantly lower genetic diversity than its sister-species probably due to the impact of past

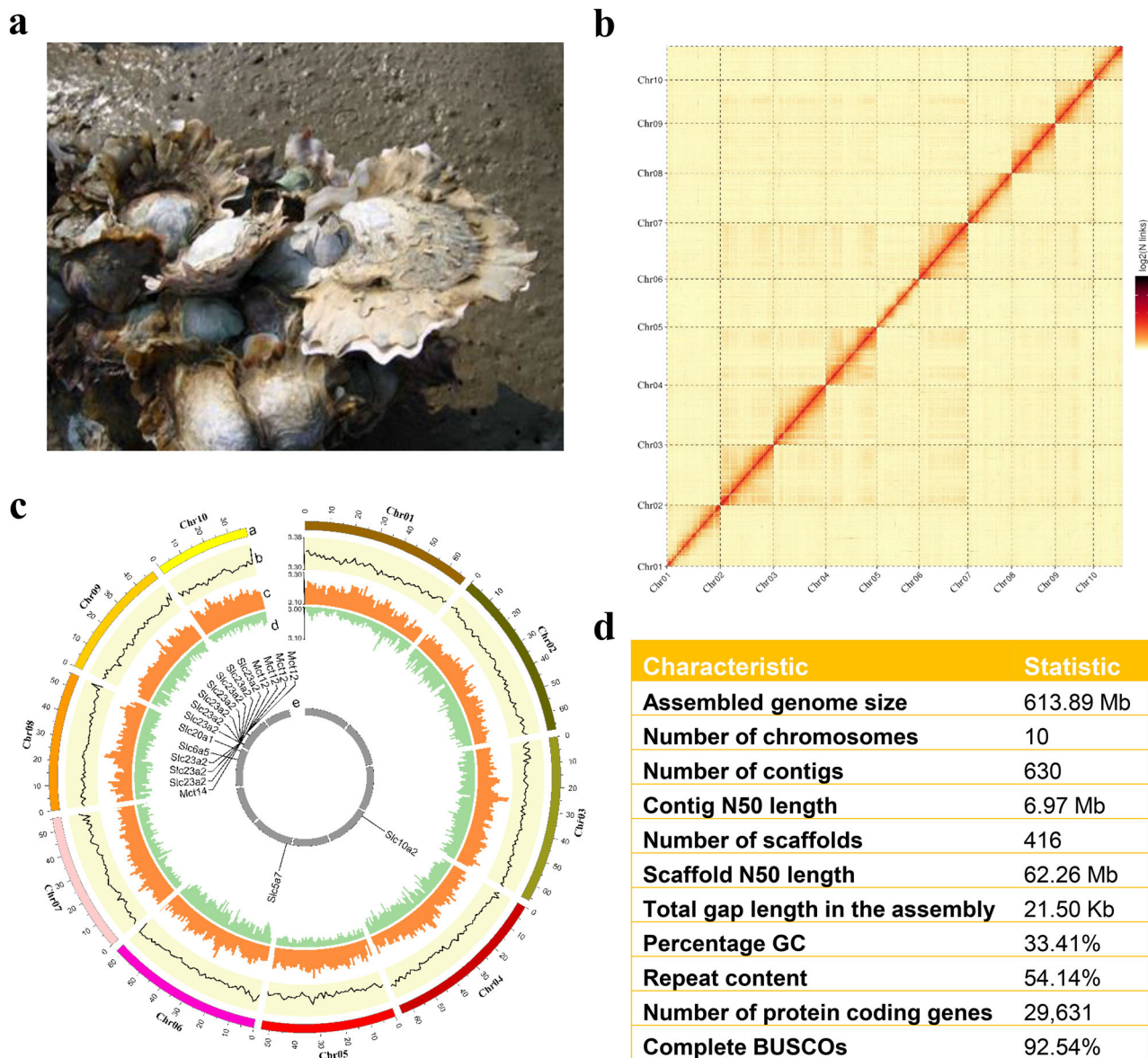
glaciation on its unique estuarine lifestyle. Its population structure is heavily influenced by geological events and ocean currents. Expansion and selection in regulatory regions of environment-responsive genes may enhance phenotypic plasticity that is critical for the adaptation to rapidly changing environments.

## Results and discussion

**Genome assembly and annotation.** We produced a chromosome-level assembly of the estuarine oyster genome of an estimated size of 614.05 Mb (19-mer analysis) (Supplementary Fig. 1, Supplementary Table 1), using a combination of nanopore long-reads generated from nine flow cells on the PromethION platform (184 Gb, 299.24×), Illumina paired-end short-reads (64 Gb) and Hi-C (106 Gb) sequences and a hierarchical assembly approach. The final assembly consisted of 630 contigs with a N50 of 6.97 Mb, spanning 613.89 Mb, 99.6% of which or 416 scaffolds with a scaffold N50 of 62.26 Mb were assembled into 10 chromosomes corresponding to the haploid number (Fig. 1b and c, Supplementary Table 2). To our knowledge, this is the most contiguous assembly (by contig N50) produced with nanopore long-read sequencing for a bivalve mollusc<sup>29–38</sup> (Supplementary Table 3). The coverage of the assembled genome was assessed by mapping RNA-seq reads and Illumina genomic reads, and over 97.9% of genomic short-reads and 97.2% RNA-seq reads were mapped to the assembly respectively (Supplementary Table 4). Also, the genome assembly captured 92.24% of the Benchmarking Universal Single Copy Orthologs (BUSCO) datasets (Fig. 1d, Supplementary Data 1), indicating the assembly has a high level of completeness. The accuracy of genome sequence was 98.32% as compared with Sanger sequencing (Supplementary Table 5). Analysis with REAPR showed that both fragment coverage distribution error and low fragment coverage over a gap were zero, indicating that the genome assembly was accurate. These results indicate that the estuarine oyster genome assembly is of high quality and can be used for down-stream analyses.

The estuarine oyster genome encoded 29,631 protein-coding genes as indicated by homology, *de novo* prediction and mRNA transcripts, 96.13% of which were functionally annotated (Supplementary Table 6). Various non-coding RNA sequences were also identified and annotated in the genome, including 1,077 transfer RNAs, 20 microRNAs and 131 ribosomal RNAs. A total of 332.40 Mb (54.14%) of repetitive elements were identified (Supplementary Table 7), which is higher than the 43% observed in the Pacific oyster<sup>32</sup>. Gene density was inversely correlated to the repetitive element content across all chromosomes (Supplementary Fig. 2,  $\rho = -0.308$ ,  $p < 0.001$ ). Genome-wide polymorphism was 0.58%, which is less than half as that of the Pacific oyster<sup>23</sup>, probably due to population bottleneck caused by past glaciations more severely impacting its estuarine lifestyle (see below).

We analyzed macrosynteny between *C. ariakensis* and two other oyster species inhabiting low-salinity estuarine, *C. virginica* and *C. hongkongensis*. High colinearity was found between *C. ariakensis* and *C. hongkongensis* across 205.48 Mb covering 20,571 genes of *C. ariakensis* genome (Supplementary Data 2), while a lower colinearity between *C. ariakensis* and *C. virginica* was detected across 194.17 Mb covering 18,692 genes of *C. ariakensis* genome (Supplementary Fig. 3, Supplementary Data 3). Our findings are consistent with the fact that the two Asian species *C. ariakensis* and *C. hongkongensis* have a closer phylogenetic relationship that diverged 22.3 Myr ago, while the Atlantic *C. virginica* diverged from the Asian species about 82.7 Myr ago<sup>39</sup>. Assembly errors in the *C. virginica* genome as indicated by the discrepancy with linkage map (X. Guo, personal communication) may also explain some of the differences.

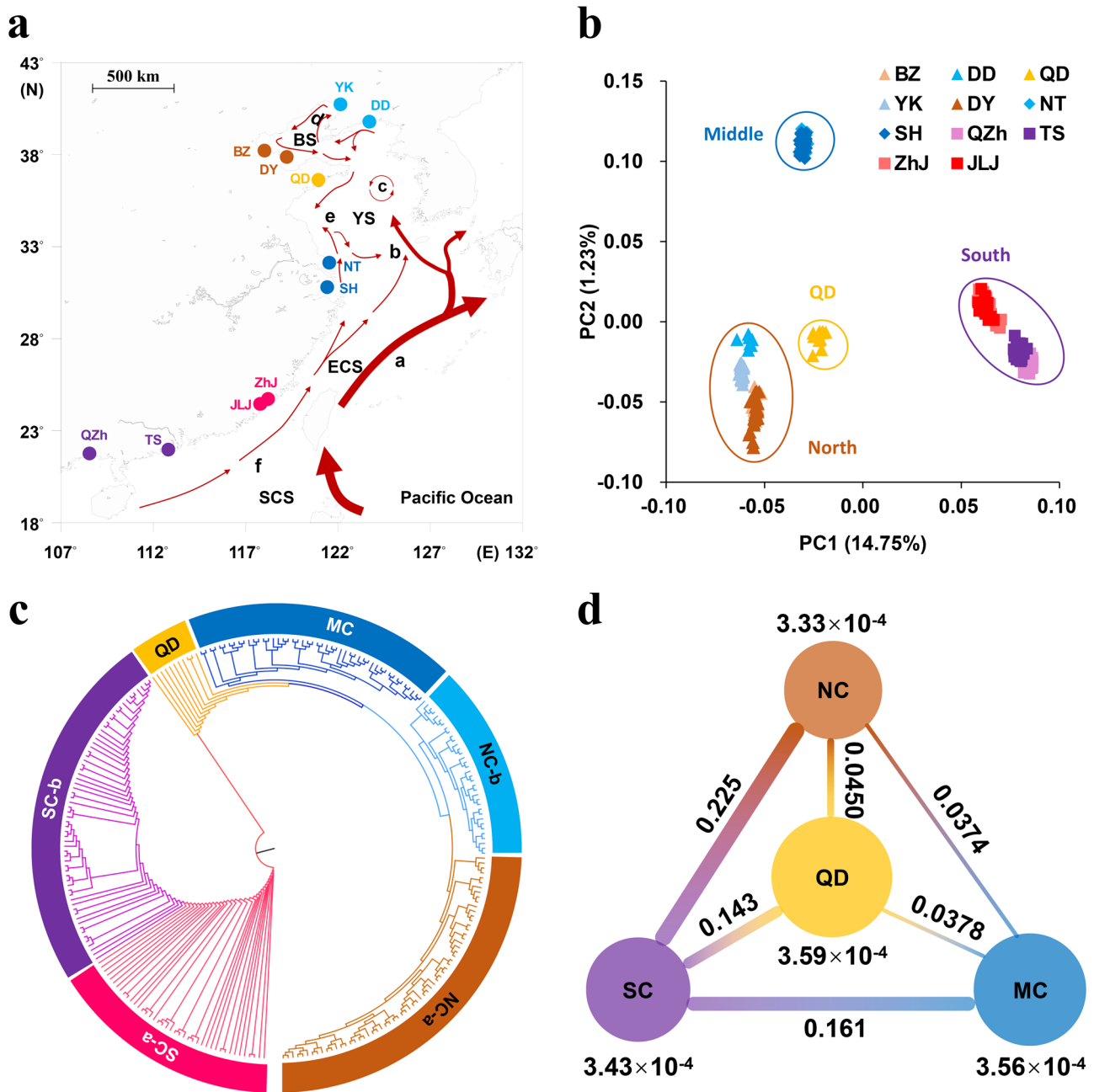


**Fig. 1** Chromosome-level assembly of the estuarine oyster *Crassostrea ariakensis* genome. **a** Estuarine oyster (photo by Lumin Qian). **b** Hi-C interaction heatmap showing 10 chromosomes of the estuarine oyster. **c** CIRCOS plot showing 10 chromosomes (a), the distribution of GC content (b), transposable elements (c), coding sequences (d), and duplicated gene cluster of the *solute carrier* families showing selection signals (e, also see Supplementary Fig. 13). **d** Summary statistics of the genome assembly.

**Genome resequencing, variation calling and population structure.** We generated 3.81 Tb clean whole-genome resequencing data from 264 wild oysters collected from 11 estuaries covering most of the distribution range<sup>26,28</sup> (Fig. 2a). The genome mapping rate averaged 95.3% ranging from 86.5% to 96.7%, and the mapped read depth averaged 19.89x. The data generated 145,271,754 SNPs (ranging from 487,881 to 640,962 per individual) and 103,080,822 indels (ranging from 342,486 to 443,381 per individual). Overall, there were 0.47 heterozygous SNPs per Kb per individual (Supplementary Data 4), which is ~35-fold lower than that in Pacific oyster populations<sup>12</sup>.

Structure analysis of genome-wide SNPs revealed significant differentiation among different geographic populations, consistent with previously analyses of fitness-related traits, neutral markers and transcriptomic data<sup>26,40–42</sup>, but with higher resolution. The optimal number of population clusters was identified as  $k = 3$  (Supplementary Fig. 4), representing three main regions along the coast of northern China (NC, 5 sites), middle China

(MC, 2 sites) and southern China (SC, 4 sites) (Supplementary Fig. 5). Principle component analysis (PCA), explaining 16% of genetic variance by two PCs, consistently revealed three distinct populations corresponding to NC, MC and SC. Further, fine-scale subpopulations were detected where oysters from Qingdao (QD) and southern subpopulation (SC-b) including Taishan (TS) and Qinzhou (QZh) were separated from NC and SC, respectively (Fig. 2b). Moreover, phylogenetic analysis using the Maximum Likelihood (ML) method supported PCA clustering, by first distinguishing southern populations from others and then identified a subpopulation (NC-a) within NC located in estuaries of southern Bohai Sea including Binzhou (BZ) and Dongying (DY) (Fig. 2c). A total of six subpopulations were identified for 11 estuaries, greatly improving the resolution of the fine-population structure. Pairwise  $F_{ST}$  using all polymorphic positions revealed exceptionally high divergence between SC and other populations, ranged from 0.143 to 0.225 (Fig. 2d), while divergence within SC, MC and NC populations was much lower (Supplementary

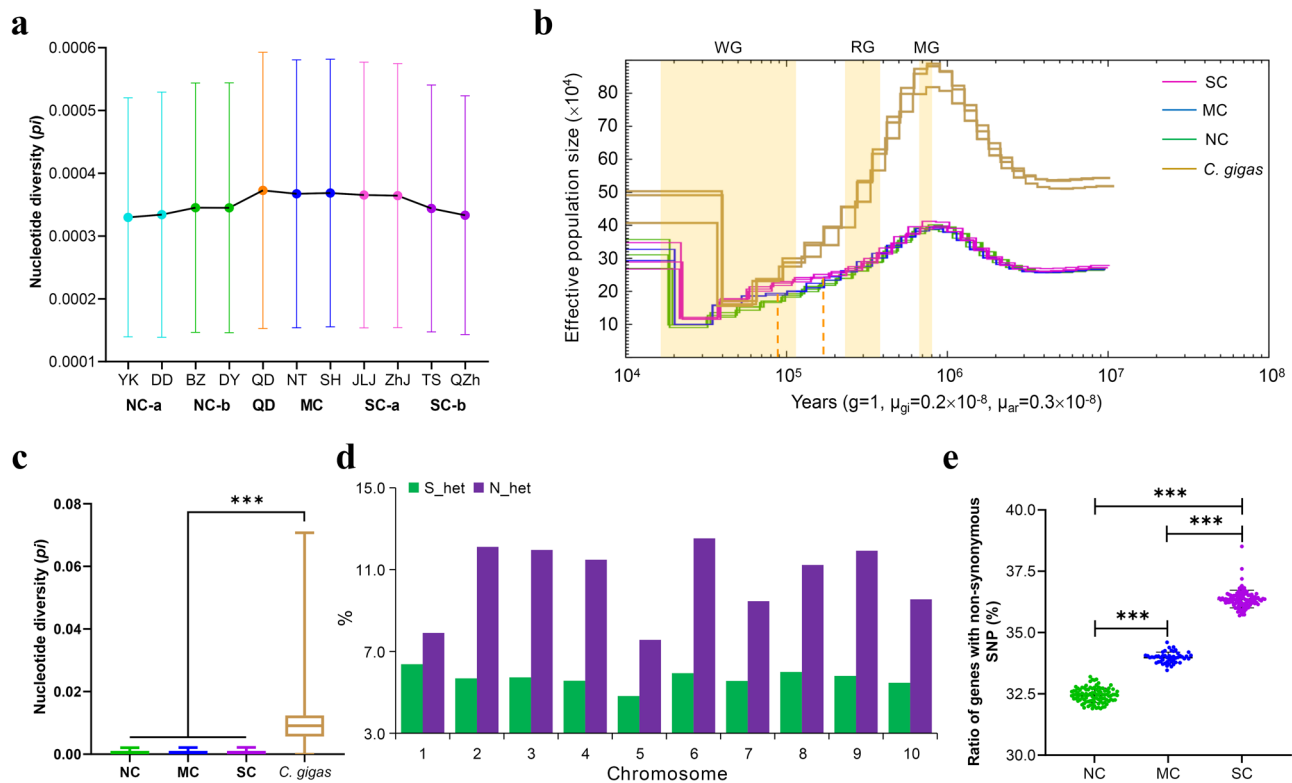


**Fig. 2 Geographic distribution and genetic differentiation of the estuarine oyster populations.** **a** Sampling locations of 264 resequenced wild oysters from 11 estuaries along the coast of China. The arrowed curves represent ocean currents in summer. **a**: Kuroshio Current, **b**: Yellow Sea Warm 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 264 resequenced oysters based on whole-genome data. **c** Phylogenetic tree of estuarine oysters inferred from genome-wide SNPs with the Maximum Likelihood (ML) method. NC-a: northern China, including BZ and DY; NC-b: northern China including DD and YK; MC: middle China including NT and SH; SC-a: southern China including JLJ and ZhJ; and SC-b: southern China including TS and QZh. **d** Nucleotide diversity ( $\pi$ , under or above the circles) and genetic divergence ( $F_{ST}$ , between populations) among four populations.

Table 8). Oysters from MC and NC were clustered together in phylogenetic tree with a lower genetic divergence ( $F_{ST} < 0.05$ ), which is comparable with the divergence among populations of the Pacific oyster in north China<sup>12</sup>. Linkage disequilibrium (LD, measured as  $r^2$ ) decreased to half of its maximum value around 2.54 – 3.00 kb in the three populations (Supplementary Fig. 6), which is substantially slower than the LD decay in the Pacific oyster ( $\sim 0.1$  kb)<sup>12</sup>. These results provide unprecedented insights into the fine-scale population structure of the estuarine oyster.

The finding of exceptionally low sequence diversity, slow LD decay and high population divergence in the estuarine oyster compared with the Pacific oyster suggests that the two species experienced different evolutionary forces.

**Historical glaciation, geological events and oceanographic forces shaping diversity and distribution.** Population structure of the estuarine oyster was largely concordant with the direction



**Fig. 3** Effects of gene flow, historical glaciation and natural selection on diversity and differentiation of the estuarine oyster. **a** Nucleotide diversity ( $\pi$ ) in 11 geographic populations along the coast of China. **b** Demographic histories of three populations of the estuarine oyster (ar) (SC, MC and NC), and Pacific oyster (*C. gigas* or gi) inferred by Pairwise Sequentially Markovian Coalescent. Periods of the Mindel glaciation (MG, 0.68–0.80 mya), Riss glaciation (RG, 0.24–0.37 mya) and Würm glaciation (WG, 10,000–120,000 years ago) were shaded by yellow. **c** Nucleotide diversity ( $\pi$ ) of three populations of the estuarine oyster, and Pacific oyster (*C. gigas*). The data are presented as box plots; the central rectangle spans the first to third quartiles of the distribution, and the ‘whiskers’ above and below the box show the maximum and minimum estimates. The line inside the rectangle shows the median, the circles represent outliers. **d** The ratio of SNPs that are highly heterozygous only in northern (N\_het) or only in southern (S\_het) populations across 10 chromosomes. **e** The ratio of genes with non-synonymous SNPs in three oyster populations. Data are presented as circle plot, with each circle representing an individual oyster and bars showing the mean value  $\pm$  SEM. Asterisks indicate significant difference (\*\*\*)  $p < 0.001$ . Error bars represent SEM values.

of ocean currents during the summer, where the northern or southern coastal currents are not crossed except over middle populations near the Yangtze River estuary (Fig. 2a and Supplementary Movie 1). Nucleotide diversity ( $\pi$ ) was highest in middle populations ( $3.56 \times 10^{-4}$ ) and QD oysters ( $3.59 \times 10^{-4}$ ), likely because converging southern and northern currents bring increased genetic diversity to the middle populations (Fig. 3a). These findings indicate that ocean currents play important role in shaping and maintaining population structure of the estuarine oyster<sup>12,43</sup>. The exceptionally high divergence between SC and NC populations can be explained by the fact that currents from the south and north do not cross over and cannot facilitate gene flow. Freshwater discharge from Yangtze River is considered as a major barrier limiting gene flow between north and south populations of marine organisms<sup>7,44–47</sup>. For the estuarine oyster, however, populations on the north (NT) and south (SH) side of Yangtze River were clustered together as MC, indicating Yangtze River is not a barrier to gene flow for the estuarine oyster that thrives in low salinity water of estuaries.

To reconstruct the demographic history, Pairwise Sequentially Markovian Coalescent (PSMC) method were used to assess fluctuations in effective population size ( $N_e$ ) in response to Quaternary climatic change using genome data of two or three estuarine oysters from each of the three populations, as well as three previously sequenced Pacific oysters for comparisons<sup>12</sup>. Both species were severely affected by glaciation events during the

past million year as their  $N_e$  peaked at  $\sim 0.90$  mya before the Mindel glaciation (MG, 0.68–0.80 mya) and then substantially decreased over three periods of glaciation: the MG, Riss (RG, 0.24–0.37 mya) and Würm (WG, 10,000–120,000 years ago) glaciation. The  $N_e$  of both species reached their bottom during the last glaciation WG, but Pacific oyster populations rebounded early and to higher levels than the estuarine oyster, suggesting the last ice age had a greater impact on the estuarine oyster. Also, during the interglacial period before the last glaciation at  $\sim 0.2$  mya, the SC and NC populations of the estuarine oyster started to diverge (Fig. 3b). We also estimated divergence time between SC and NC based on cytochrome oxidase I (COI) sequence data using divergence time between *C. gigas* and *C. angulata* as reference<sup>39,46</sup>. The COI-based estimate placed the divergence time at 0.14–0.63 mya, which is close to the 0.2 mya estimated by PSMC with whole-genome data (Fig. 3b). The divergence time coincided with the formation land bridge between Taiwan and mainland China from 0.2 mya to 25,000 years ago by tectonic movement<sup>48</sup>, creating a physical barrier to gene flow between southern and northern populations that is now maintained by ocean currents.

The nucleotide diversity of all three populations of the estuarine oyster was 25-fold lower than that of the Pacific oyster ( $\pi_{C. gigas} = 9.27 \times 10^{-3}$ , Fig. 3c). Similarly, stickleback fishes adapted to freshwater areas exhibited lower  $\pi$  and  $N_e$  values than counterparts dwelling in brackish areas<sup>49</sup>. These findings suggest

that climate change may affect species with different lifestyles differently and the estuarine oyster may be more sensitive to climate change than the Pacific oyster that inhabits more open waters of high salinity. Further, *Ne* curves of middle and northern populations split ~90,000 years ago, which corresponded to the great sea level fall at the sub-glaciation of WG when Bohai Sea dried up<sup>50</sup>. The lower *Ne* and nucleotide diversity of northern population (Fig. 3a, b) support that a stronger bottleneck occurred in estuarine oysters of Bohai Sea. Relatively lower nucleotide diversity was also found in Bohai populations of the Pacific oyster<sup>12</sup>, possibly due to the same sea level fall. The role of bottlenecks and geographic isolation resulting from historical glaciations and tectonic events in shaping phylogeography has been demonstrated in other molluscs, such as the divergence between Atlantic coast and Gulf of Mexico populations of the eastern oyster *Crassostrea virginica*<sup>9</sup>. Our results provide putative divergence times among estuarine oyster populations linked to historical events and suggest past climate and tectonic movement played an important role in shaping genetic diversity and population differentiation in the estuarine oyster.

To explore the effects of selection and genetic drift, we compared highly heterozygous SNPs (heterozygosity > 0.5) in northern and southern populations. The number of SNPs that are highly heterozygous in NC only (14,373) were 1.89-fold higher than those highly heterozygous in SC only (7,595) across 10 chromosomes (Fig. 3d). Both directional selection and genetic bottleneck can decrease the number of heterozygous SNPs. The finding that southern oysters had significantly higher ratio of genes with non-synonymous SNPs (35.67% ± 0.35%) than the middle (33.33% ± 0.22%) and northern (31.82% ± 0.29%) oysters ( $p < 0.01$ , Fig. 3e), suggests that the southern oysters experienced stronger selection. Stronger selection, potentially from strong environmental disturbance, decreased the numbers of heterozygous SNPs in southern oysters. Thus, in addition to isolation and bottlenecks from historical glaciation, geological events and oceanographic forces, selection or local adaptation may also play an important role in shaping variation and phylogeography of the estuarine oyster.

**Signatures of selection.** Oysters inhabiting northern and southern estuaries experience significantly different environmental conditions. Southern habitats are characterized by high temperature and low salinity. Satellite remote sensing data from 2000 to 2017 indicated that the monthly average sea surface temperature (SST) of southern habitats was 10.35 °C higher than that of northern habitats (Supplementary Fig. 7). Salinity at the northern BZ site was 10.98 ‰ higher than the southern TS site<sup>3</sup>. Climate change may enhance the difference and cause salinity to increase at north habitats but decrease at south habitats<sup>51,52</sup>. Thus, temperature and salinity are the two most important environmental factors driving adaptive divergence between northern and southern populations. With discontinuous distribution and limited gene flow, the southern population became locally adapted and evolved higher tolerance to high temperature and low salinity<sup>3</sup>. We expect that some genomic regions were subjected to selection and contributed to adaptation to higher temperature and lower salinity conditions in the southern population.

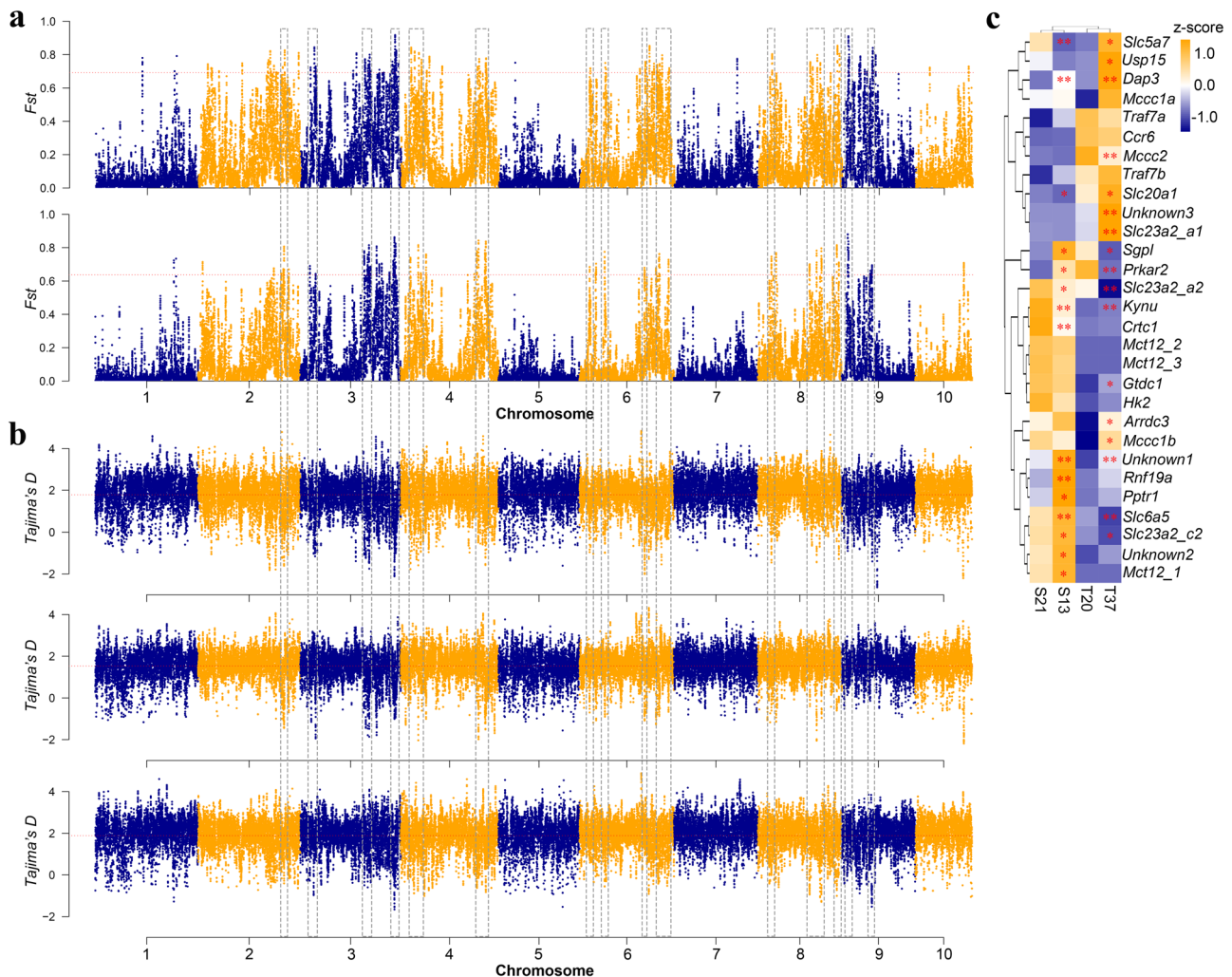
To identify signatures of selection sweeps, we calculated the fixation index ( $F_{ST}$ ) and selection statistics (Tajima's *D*) between two population pairs (SC vs. NC and SC vs. MC) and identified  $F_{ST}$  outliers (top 1%,  $F_{ST\_north}$  vs. south >= 0.693,  $F_{ST\_middle}$  vs. south >= 0.637). Only genomic regions surrounding selective peaks that overlapped between the two population pairs and located at the valleys of Tajima's *D* values in one of the three

populations were considered as under selection. 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 Figs. 8–13, Supplementary Data 5). Most of these candidate genes are involved in response to environmental disturbances in salinity and temperature<sup>53–60</sup>.

To assess transcriptomic response of the estuarine oyster to temperature and salinity disturbances, we conducted RNA-seq analysis of oysters exposed to elevated temperature (6 h under 37 °C) and high salinity (7 days under 60 ‰). Only genes with 10 or more aligned reads in >90% samples were used for subsequent analyses. About 44.17% (10,279 of 23,270) and 11.7% (2,757 of 23,650) of genes were differentially expressed under high salinity and temperature stresses, respectively, with 1,088 genes responsive to both stressors (Supplementary Fig. 14). For the 51 candidate genes from regions under selection, a total of 29 genes were expressed, 75.9% (22 of 29) and 58.6% (17 of 29) of which were differentially expressed under high salinity and high temperature, respectively ( $p < 0.05$ , Fig. 4c), indicating most genes under selection are involved in response to environmental disturbances in these two factors. Thirteen genes responded to both thermal and salinity stresses, while only three genes were not sensitive to the two stressors. Nine and four genes were highly responsive to high salinity and high temperature, respectively. The finding that most of genes from regions under selection are involved in response to temperature and salinity disturbances suggests that these two environmental factors are important drivers of adaptive evolution in the estuarine oyster.

**Expansion of genes contributing to temperature and salinity adaptation.** Among the 24 regions under selection, we found two clusters of tandemly duplicated genes belonging to *solute carrier* families, 10 copies of *Slc23a2* and four copies of *monocarboxylate transporter 12* (*Mct12*, also known as *Slc16a12*) (Supplementary Fig. 13c), located in two regions of chromosome 9 with high divergence ( $F_{ST} = 0.81$  and 0.76) among the three populations (Supplementary Fig. 13d). The genomic region spanning *Slc23a2* gene families exhibited extremely low Tajima's *D* values in northern oysters, while these spanning *Mct12* gene families had extremely low Tajima's *D* values in southern oysters (Supplementary Fig. 13e). These findings indicate that *Slc23a2* and *Mct12* genes were under directional selection in northern and southern environments respectively, highlighting the critical role of *Slc* family genes in adaptation of marine species, as also reported in porpoises and coral<sup>60–63</sup>, to salinity and temperature variations.

The ten copies of *Slc23a2* family belong to three orthogroups, where two of them are annotated as purine permease (a: OG0011985 and c: OG0000489) and the other is annotated as uric acid transporter (OG0000633). Four copies of *Mct12* gene family belong to one orthogroup annotated as purine efflux pump (OG0000571). All these orthogroups were extensively expanded in the estuarine oyster, as well as in the other two oyster species (*Crassostrea virginica* and *Crassostrea hongkongensis*) that occur in warm and low-salinity estuarine environments, but not so in *C. gigas* that inhabit cool, open and high-salinity waters (Supplementary Fig. 15). In addition to their presence in tandem arrays on chromosome 9, the expansion of the *solute carrier* families was further supported by phylogenetic analysis where duplicated genes were mostly clustered in orthogroup- and lineage/species-specific manners (Supplementary Fig. 16). Gene copies that did not respond to heat and high-salinity stresses were not expressed or expressed at very lower expression (FPKM < 0.5) (Supplementary Table 9). Conversely, three copies of *Slc23a2* (*Slc23a2\_a1*, *Slc23a2\_a2* and *Slc23a2\_c2*) in two expanded orthogroups



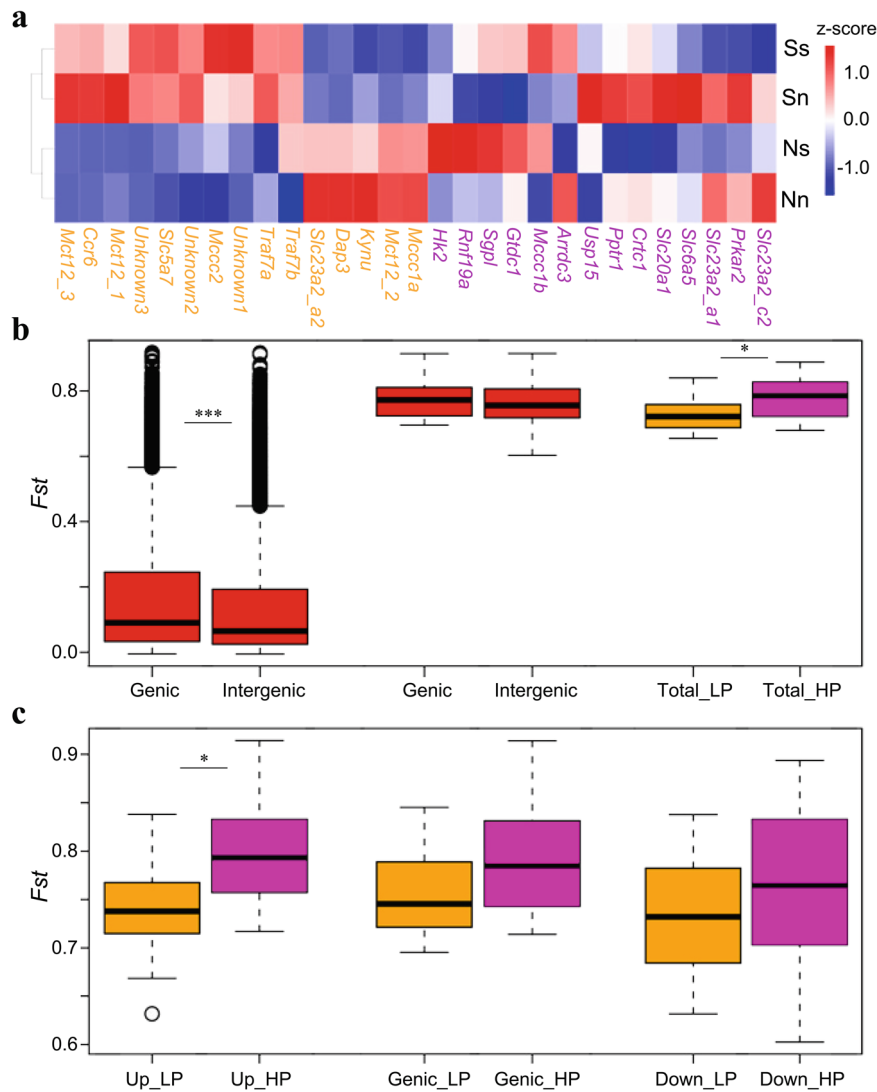
**Fig. 4** Genome-wide distribution of selective sweep signals and transcriptional responses of associated genes to thermal and salinity challenges in the estuarine oyster. **a** Global  $F_{ST}$  values (top 1%, red lines) between two population pairs: northern vs. southern (up) and middle vs. southern (bottom). **b** Global Tajima's  $D$  values in northern (up), middle (middle) and southern (bottom) populations. **c** Expression of genes under selection exposed to thermal (6 h at 37 °C) and high-salinity (7 days at 60‰) challenges. Asterisks indicate significant difference (\* $p < 0.05$ , \*\* $p < 0.01$ ). The blue-to-yellow color bar represents the relative expression level of the given gene from low to high.

responded to both temperature and salinity challenges, while three copies of *Mct12* (*Mct12\_1*, *Mct12\_2* and *Mct12\_3*) responded to salinity challenge only (Fig. 4c) and preferentially clustered together (*Mct12\_1/2/3* and *Slc23a2\_a1/2*, Supplementary Fig. 16). The expansion of these environment-responsive genes may enhance transcriptional complexity or plasticity and play an important role in adaption to diverse temperature and salinity conditions in the estuarine oyster and two other species inhabiting similar environments. This finding provides further evidence that gene duplication is critical for stress adaptation, as previously demonstrated for the expansion of *heat shock protein* and *inhibitor of apoptosis* gene families in the Pacific oyster and other marine bivalves<sup>23,35,64,65</sup>.

**Stronger selection in upstream regulatory regions of environment-responsive genes with high plasticity.** To understand the interaction between divergence and plasticity, we examined divergence in different genomic regions (genic, upstream and downstream) with  $F_{ST}$  and transcriptional plasticity of the 29 genes from regions under selection in response to environmental translocations.  $F_1$  progeny produced from northern and southern populations were acclimatized at both northern

and southern habitats for three months before transcriptional analysis. Fourteen of the 29 genes exhibited higher plasticity (HP) where they showed significantly differentiated expression when translocated to nonnative environments in both oyster populations ( $p < 0.05$ ), while other genes had lower plasticity (LP) and their expression didn't strongly respond to environmental translocation/change (Fig. 5a). The transcriptional plasticity of the HP genes was 19.15% higher than that of LP genes ( $p = 0.0255$ , Supplementary Fig. 17).

For the whole genome, 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 stronger selection in genic regions driving differentiation between northern and southern populations. For the 29 genes from regions under selection, both genic ( $F_{ST} = 0.7745$ ) and intergenic ( $F_{ST} = 0.7585$ ) regions showed strong selection signals or divergence compared with all genes in the genome, and the difference between genic and intergenic regions was not significant ( $p > 0.05$ ). For both genic and intergenic regions, genes with high plasticity showed significantly ( $p = 0.0230$ ) higher divergence ( $F_{ST} = 0.7816$ ) between northern and southern oysters than LP genes ( $F_{ST} = 0.7274$ ) (Fig. 5b).



**Fig. 5 Transcriptional and genomic divergence of genes with selection signals in the estuarine oyster.** **a** Expression level of genes under selection in northern (N) and southern (S) populations (capital letters) acclimated to northern (n) and southern (s) environments (lowercase letters), showing low plasticity (LP, orange) and high plasticity (HP, purple) genes. The blue-to-red color bar represents the relative expression level of the given gene from low to high. **b** Genetic divergence ( $F_{ST}$ ) for genic and intergenic regions at genome level (left) and for 29 genes under selection (middle), and for LP and HP genes in 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 LP and HP genes. Asterisks indicate significant difference (\* $p < 0.05$ , \*\*\* $p < 0.001$ ). The data are presented as box plots; the central rectangle spans the first to third quartiles of the distribution, and the ‘whiskers’ above and below the box show the maximum and minimum estimates. The line inside the rectangle shows the median, the circles represent outliers.

Specifically, the upstream intergenic regions of HP genes exhibited significantly ( $p = 0.01512$ ) higher divergence ( $F_{ST} = 0.7958$ ) between northern and southern populations than that of LP genes ( $F_{ST} = 0.7402$ ), while there was no significant difference between the two types of genes for both genic and downstream intergenic regions ( $p > 0.05$ , Fig. 5c). Further, LP genes showed higher divergence in genic regions than up and downstream regions. While selection preferentially targeted genic regions at genome level and for LP genes, HP genes showing high plasticity exhibited stronger selection in upstream intergenic regions, where critical regulatory elements such as promoter and enhancers reside, potentially regulate gene expression and enhance transcriptional plasticity<sup>18</sup>. These results suggest that selection and divergence in upstream regulatory regions may enhance phenotypic plasticity that contributes to the survival and adaption of organisms facing environmental disturbances<sup>66</sup>. The mechanism of selection favoring plasticity is not clear at this time. Directional

selection tends to reduce genetic variation and facilitate local adaptation. Selection acting on regulatory elements may create divergence and transcriptional complexity that enhance phenotypic plasticity<sup>18,19</sup> at species level. Directional selection acting on duplicated genes can potentially direct duplicated member genes to express under different conditions and therefore enhance overall phenotypic plasticity. It is also possible that balancing selection retains diversity in regulatory elements and increase plasticity. It has been suggested that balancing selection is pervasive in oysters and important for their evolution<sup>67</sup>.

In conclusion, analysis of the estuarine oyster genome revealed exceptionally low genetic diversity and fine-scale population structures, which were shaped and maintained by climate history, geological events and oceanographic forces. Integration of genomic and transcriptomic analyses revealed genes in regions under selection are mostly involved in environmental response including the expanded *solute carrier* gene families that are



important for environmental adaptation. Expansion and selection in upstream regulatory regions of environment-responsive genes may hence be critical for the adaptation to rapidly changing environments in the estuarine oyster.

## Methods

**Genome sequencing.** A wild adult estuarine oyster *Crassostrea ariakensis* was obtained from Binzhou, Bohai Sea, northern China. Four tissue samples (gill, mantle, adductor muscle and labial palp) were collected and flash-frozen in liquid nitrogen. Genomic DNA was extracted from the adductor muscle using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) and used to construct libraries for Oxford Nanopore Technologies' (ONT) long-read sequencing. gDNA (2 µg) was repaired using NEB Next FFPE DNA Repair Mix kit (M6630, USA) and subsequently processed using the ONT Template prep kit (SQK-LSK109, UK) according to the manufacturer's instructions. The large segment library was pre-mixed with loading beads and then pipetted into a previously used and washed R9 flow cell. The library was sequenced on the ONT PromethION platform with nine R9 cells and ONT sequencing reagent kit (EXP-FLP001.PRO.6, UK) according to the manufacturer's instructions. In addition, libraries with insert size of ~350 bp were prepared and sequenced using Illumina HiSeq 4000 platform.

**Genome size and heterozygosity.** We conducted k-mer copy number (KCN) analyses using  $k = 17, 19$  and  $21$  to estimate genome size.  $K$  was the least odd number that conform to the following formula:  $4^k/G > 200$ , where  $G$  indicates the genome size (bp). Previous study estimated the genome size of *C. gigas* and *C. ariakensis* by flow cytometry, where their respective C-values, 0.89 pg and 0.99 pg<sup>68</sup>, have a ratio of 1.112. Assuming the genome size of *C. gigas* is 586.8 Mb<sup>34</sup>, the genome size of *C. ariakensis* would be 652.7 Mb. Thus, we selected 19-mer for genome size estimation.

Trimmed Illumina short reads were used to generate the KCN distribution. The 19-mer KCN distribution showed two distinct peaks (Supplementary Fig. 1). The first peak (KCN = 44) represents heterozygous single copy k-mers while the second peak (KCN = 89) represents homozygous single copy k-mers in the genome. Genome size was estimated by the formula  $G = K\_num/peak\ depth$ .

**Contig-level assembly using long-reads.** 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<sup>69</sup> v1.5, and then were assembled using Canu, WTDBG2<sup>70</sup> and SMARTdenovo tools. Quickmerge<sup>71</sup> v0.2.2 was used to join the three assemblies, and the assembly was corrected for 3 cycles with long reads using Racon<sup>72</sup> and for 3 cycles with Illumina reads using Pilon<sup>73</sup> v1.22 with default parameters (Supplementary Tables 10 and 11). Then, the Purge Haplotigs<sup>74</sup> software was used to remove redundant heterozygous contigs with parameter 'purge\_haplotigs purge -a 55'.

**Chromosome-level assembly with Hi-C.** The same adductor muscle was used for crosslink with formaldehyde at room temperature and DNA was digested with HindIII. The 5' overhangs were adapted with biotinylated nucleotides, and free blunt ends were ligated. Then, the crosslinks were reversed and the DNA was purified and further sheared to 300–700 bp fragments. Streptavidin beads were used to isolate biotin-tagged fragments for PCR enrichment. The Hi-C library was constructed using Illumina's paired-end kits according to the manufacturer's instructions, which was subsequently sequenced on the Illumina HiSeq 4000 platform. Contigs and scaffolds which were then sorted and oriented into super scaffolds using LACHESIS<sup>75</sup> 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.** Hi-C contact heatmap of the number of Hi-C links between 100 kb windows on the pseudochromosomes was used to assess the accuracy of the Hi-C assembly. Benchmarking Universal Single-Copy Orthologs (BUSCO, obd10) with 954 conserved genes were used to assess the completeness and accuracy of gene coverage. The completeness and accuracy of the assembly was also evaluated by mapping rate of RNA-seq reads (see Protein coding gene annotation) and Illumina genomic reads with Burrows-Wheeler Aligner<sup>76</sup> v.0.7.8. Further, the accuracy of genome sequences was assessed by REAPR<sup>77</sup> v. 1.0.18 with default parameters to identify errors in genome assemblies.

**Repeat annotation.** Transposable elements (TEs) were identified and classified using homology and *de novo*-based approaches. RepeatScout and LTR\_FINDER were used to construct *de novo* repeat libraries. The *de novo*-based libraries were further classified with PASTEClassifier<sup>78</sup> to obtain a consensus library, and combined with the repeat library of Repbase data. RepeatMasker<sup>79</sup> v4.0.5 was used to identify TEs in the estuarine oyster genome with the combined library.

**Protein-coding gene annotation.** Protein-coding genes were identified with three methods: *de novo* prediction, homology-based prediction and mRNA-based

prediction. For *de novo* prediction, five ab initio gene prediction programs, Genscan<sup>80</sup> v1.0, Augustus<sup>81</sup> v2.4 (with transcriptomic data of the estuarine oyster used for training), GlimmerHMM<sup>82</sup> v3.0.4, GeneID<sup>83</sup> v1.4 and SNAP<sup>84</sup>, were used to predict genes in the repeat-masked genome (hard-masking). For homolog-based prediction, protein sequences from 10 well-annotated species that were annotated by the NCBI Eukaryotic Genome Annotation Pipeline, *Homo sapiens*, *Danio rerio*, *Aplysia californica*, *Strongylocentrotus purpuratus*, *C. gigas*, *C. virginica*, *Biomphalaria glabrata*, *Lingula anatina*, *Octopus bimaculoides* and *Mizuhopecten yessoensis*, were downloaded from NCBI and aligned to the repeat-masked estuarine oyster genome using tblastn<sup>85</sup> with E-value  $\leq 1E-05$ . We used GeMoMa<sup>86</sup> v1.3.1 to predict gene models based on alignment sequences. For mRNA-based prediction, four tissues (gill, mantle, adductor muscle and labial palp) were collected from the same oyster used for genome sequencing. Total RNA was extracted separately using TRIzol reagent (OMEGA, USA) according to the manufacturer's instructions. RNA quality and quantity were assessed using the Agilent 2100 instrument and Qubit 2.0 fluorometer (Thermo Fisher, USA), respectively. RNA from four tissues were combined in equimolar quantities into a single pool. Full-length coding DNA (cDNA) was obtained using SMARTer PCR cDNA Synthesis Kit (Clontech, USA). The library was constructed using SMARTbell<sup>®</sup> Template Prep Kit 2.0 (PacBio, USA). cDNA was sequenced by PacBio Sequel II platform with SMRT cell 8 M (Biomarker, China). RNA-seq data were filtered to remove adapters and then trimmed to remove low-quality bases. Clean reads were aligned to reference genome using TopHat2<sup>87</sup> and then assembled using Trinity<sup>88</sup>. Full transcriptome-based genome annotation was predicted using PASA<sup>89</sup> v2.2.2. Finally, EvidenceModeler<sup>90</sup> (EVM) v1.1.1 was used to generate a weighted and non-redundant gene set by integrating all gene models predicted by the three methods (Supplementary Table 12).

Homologous sequences in the genome were identified with genBlastA<sup>91</sup> v1.0.4 using the integrated gene set, and GeneWise<sup>92</sup> was used to identify pseudogenes. Transfer RNAs (tRNAs) were defined using tRNAscan-SE<sup>93</sup> v1.3.1 software with eukaryote default parameters. MicroRNA and rRNA were identified with Infernal BLASTN<sup>94</sup> against the Rfam<sup>95</sup> database v12.0.

Functional annotation of protein-coding genes was conducted by aligning them to the NCBI non-redundant protein<sup>96</sup> (NR), SwissProt<sup>97</sup>, KOG<sup>98</sup> and TrEMBL<sup>97</sup> databases using BLAST<sup>94</sup> v2.2.31 with a maximal e-value of  $1e-05$ . Domains were identified by searching against Pfam<sup>99</sup> database using HMMER<sup>100</sup> v3.0. Genes were mapped to Gene Ontology (GO) terms and KEGG pathways to identify their best functional classification.

**Syntenic analysis.** We analyzed macrosynteny between *C. ariakensis* and the other two oyster species living in low-salinity estuaries, *C. virginica* and *C. hongkongensis*. First, we aligned gene sequences between two oyster species using Diamond<sup>101</sup> v0.9.29.130, to identify homologous gene pairs (e-value  $< 1e^{-5}$ , C score  $> 0.5$ ). MCSanX<sup>102</sup> was used to analyze chromosome collinearity for syntenic blocks between *C. ariakensis* and the other two oyster species (alignment significance =  $1e^{-10}$ ), using Diamond result file and gff file. The maximum gap size was set to 25 genes and a minimum syntenic block required 5 genes.

**Whole-genome resequencing and mapping.** We collected 264 wild estuarine oysters from 11 estuaries (Fig. 3a), covering most of its distribution range<sup>26,28</sup>. Genomic DNA was isolated from gill tissue with standard phenol-chloroform extraction and used to construct paired-end libraries with an insert size of ~350 bp according to the manufacturer's instructions (Illumina Inc., San Diego, CA, USA) for sequencing on the Illumina HiSeq X Ten Sequencer. We obtained ~14.42 Gb of clean data for each sample, giving an average depth of 19.9X coverage (15–28X) (Supplementary Data 4 and 6). The 150-bp paired-end reads were mapped to the estuarine oyster reference genome (PRJNA715058) using BWA<sup>76</sup> with default parameters (bwa mem -M -t 10 -T 20). Mapping data were then converted into the BAM format and sorted by SAMtools v1.3.1<sup>103</sup> to remove duplicate reads. Read pairs with the highest mapping quality were retained if multiple pairs have identical external coordinates.

**Variant calling and annotation.** The Genome Analysis Toolkit (GATK) v3.7<sup>104</sup> module HaplotypeCaller was used to obtain high-quality variation calling of each sample. SNPs were further filtered with parameters 'QD < 2.0 || FS > 60.0 || MQ < 40.0'. Similarly, indels were called and filtered using parameters 'QD < 2.0 || FS > 60.0'. Filtered SNPs were annotated by SnpEff<sup>105</sup> and then classified into regions of exon, intron, splicing site, and upstream and downstream intergenic regions, and as heterozygous or homozygous variants. To characterize the types of variants in northern and southern oysters, Plink<sup>106</sup> was used to filter raw SNPs using parameters of MAF > 0.05 and Int > 0.8. The retained SNPs were singly classified as highly heterozygous in a population if more than half individuals were heterozygous in that population but not in the other 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 mutations were different between northern and southern populations, using the function *wilcoxsign-test* in R package "coin".

**Population analysis.** Population structure was inferred using ADMIXTURE v.1.23<sup>107</sup> with default settings. The number of assumed genetic clusters  $K$  ranged from 2 to 5, and the optimum  $K$  was assessed with cross-validation errors. The individual-based Maximum Likelihood (ML) phylogenetic tree was constructed using MEGA<sup>108</sup> under Jukes–Cantor model with 1000 bootstraps, and visualized using FigTree. PCA was performed with whole-genome SNPs of all 264 individuals using Eigensoft<sup>109</sup>. To evaluate LD decay, parameter  $r^2$  between any two loci was calculated within each chromosome using Plink v.1.07<sup>106</sup> 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 genome-wide LD was averaged across all chromosomes. The LD decay was plotted against the length of distance. Pop-genome R package<sup>110</sup> was used to calculate Tajima's  $D$ , global  $F_{ST}$  and nucleotide diversity ( $p$ ) using a 100-kb sliding window with the step size of 10-kb.

**Demographic history.** We implemented PSMC<sup>111</sup> to estimate dynamics of effective population size ( $N_e$ ) and possible divergence time over the past several million years. A total of eight estuarine oysters from northern ( $n = 3$ ), middle ( $n = 2$ ) and southern ( $n = 3$ ) populations and three Pacific oysters<sup>12</sup> with high sequencing depth ( $20 \times - 28\times$ ) were used. To minimize the probability of false positives, sequencing depth of SNPs was filtered with parameters: MinDepth = average depth/3, MaxDepth = average depth $\times 2$ . The PSMC parameters were set as: `-N25 -t15 -r5 -p '4 + 25 * 2 + 4 + 6'` to estimate historical  $N_e$ . The estimated generation time ( $g$ ) was set as 1 for both species, while mutation rates ( $\mu$ ) were calculated, following the formula  $T_{\text{divergence}} = Ks/2\mu$ , as  $0.3 \times 10^{-8}$  and  $0.2 \times 10^{-8}$  for estuarine and Pacific oysters, respectively.

**Selection signature detection.** To identify selection sweeps potentially contributing to adaptation to southern environments, we calculated fixation ( $F_{ST}$ ) and selection statistics (Tajima's  $D$ ) between two pairs of populations, north vs. south and middle vs. south, in a 100-kb sliding window with a step size of 10-kb. Genomic regions showing strong selection signals were defined as: (1) regions with top 1%  $F_{ST}$  values that overlapped in both pair comparisons; and (2) regions located at valleys of Tajima's  $D$  values distribution along the chromosome in one of the three populations.

**Expansion of candidate genes under selection.** Orthologs between genes from four oyster species with high-quality genomes, *C. ariakensis*, *C. hongkongensis*<sup>33</sup>, *C. virginica* (NCBI assembly Bioproject PRJNA376014, GCA\_002022765.4 *C. virginica*-3.0) and *C. gigas*<sup>34</sup>, were determined with Orthofinder<sup>112</sup> v2.3.7, with default parameters using protein sequences ( $e$ -value =  $1e^{-5}$ ). For expanded *Slc* gene families under selection, we constructed phylogenetic tree for orthologs belong to *Slc23a2* and *Mct12* gene families in *Crassostrea* oysters, including *C. gigas*, *C. ariakensis*, *C. hongkongensis* and *C. virginica*. The phylogenetic relationship was inferred with protein sequences using the Maximum Likelihood (ML) method and the Whelan And Goldman (WAG) model in MEGA7 software<sup>108</sup>, with 1,000 bootstraps.

**Response to temperature and salinity disturbance.** To investigate responses of estuarine oysters to challenges of elevated temperature and high salinity, we collected wild oysters and exposed them to different temperatures (20 °C and 37 °C for 6 h) and salinities (20 ‰ and 60 ‰ for 7 days), respectively. The control oysters were exposed to seawater of 20 ‰ salinity at 20 °C. Gills from five oysters were individually sampled and flash-frozen in liquid nitrogen for subsequent transcriptomic analysis.

**Reciprocal transplantation experiments.** To explore the plasticity of candidate genes from regions under selection, we measured transcriptional changes in oysters subjected to native and non-native environments after reciprocal transplantation<sup>3</sup>. Briefly, wild oysters derived from northern (Binzhou: BZ, Bohai Sea) and southern (Taishan: TS, East China Sea) environments were collected and used to produce  $F_1$  progeny. For each population, 40 males and 40 females were selected as parents. Eggs were pooled and then divided into 40 beakers, each fertilized with sperm from one of the 40 males to create all possible crosses. Zygotes fathered by every eight males were combined and cultured as one group, and the five groups were reared in separate tanks in the hatchery and nursery to juvenile stage. Two-month-old  $F_1$  juveniles from each of the two populations were outplanted at two source habitats to assess their responses to reciprocal transplantation or environmental change. After three months of acclimation at northern and southern environments, gills of five oysters were sampled from each of population at both habitats in situ and flash-frozen in liquid nitrogen for subsequent transcriptomic analysis.

**Transcriptomic analysis.** Total RNA was isolated from gills sampled from laboratory challenge experiment (high temperature and high salinity) and field reciprocal transplantation experiment, using the RNeasy Pure Tissue Kit (Qiagen) following the manufacturer's protocol. The RNA integrity and concentration were examined with 1.2% agarose gel electrophoresis and Nanodrop 2000 spectrophotometer, respectively. DNA contamination was removed with DNase I treatment. RNA integrity was further assessed using an Agilent Bioanalyzer 2100 and the RNA Nano 6000 Assay Kit. One  $\mu$ g RNA per sample was used to

construct sequencing libraries using NEBNext Ultra™ RNA Library Prep Kit for 150-bp paired-end sequencing on an Illumina HiSeq 4000 platform. Clean data were obtained by removing reads containing adapters, reads containing ploy-N and low-quality reads. TopHat2 was used to map clean reads to the estuarine oyster genome. StringTie v2.0 was used for read assembly. Only reads with 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 used DESeq2 to determine differentially expressed genes (DEGs) between different populations or treatments. Genes with an adjusted  $p$ -value  $< 0.01$  using the Benjamin and Hochberg's approach were accepted as DEGs. For genes from regions under selection, those significantly differentially expressed between northern and southern habitats in both oyster populations were defined as high plasticity genes (HP), while others were defined as low plasticity genes (LP). A hierarchical cluster analysis was performed to show differential expression of these genes, using the *heatmap* package in R software.

**Statistics and reproducibility.** The difference of gene expression level between northern and southern oyster populations at two environmental habitats and those under thermal and high-salinity stressed conditions were determined by DESeq2 in R software. A total of 264 wild oysters collected from 11 estuaries were used for whole-genome resequencing and 40 individuals were used to compared gene expression difference under different habitats, thermal and high-salinity conditions. For all figures, the criterion for statistical significance was set as  $*P < 0.05$ ,  $**P < 0.01$  and  $***P < 0.001$ .

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The genome, whole-genome re-sequencing and transcriptome datasets were deposited in the Sequence Read Archive (SRA) database under the accession number PRJNA715058, or GenBank under the accession number JAGFMG000000000.1. In addition, we uploaded annotation files of estuarine oyster genome, including gene locations (.gff3), protein models (.pep), CDS (.cds) and exon (.exon) sequences to FigShare online ([https://figshare.com/articles/dataset/Genome\\_of\\_the\\_estuarine\\_oyster\\_provides\\_insights\\_into\\_climate\\_impact\\_and\\_adaptive\\_plasticity/16557390](https://figshare.com/articles/dataset/Genome_of_the_estuarine_oyster_provides_insights_into_climate_impact_and_adaptive_plasticity/16557390)). Source data underlying Figs. 3 and 5b, c were provided in Supplementary Data 7.

Received: 4 July 2021; Accepted: 28 October 2021;

Published online: 12 November 2021

## References

- Hoegh-Guldberg, O. & Bruno, J. F. The impact of climate change on the world's marine ecosystems. *Science* **328**, 1523–1528 (2010).
- Chou, C. et al. Increase in the range between wet and dry season precipitation. *Nat. Geosci.* **6**, 263–267 (2013).
- Li, A. et al. Molecular and Fitness Data Reveal Local Adaptation of Southern and Northern Estuarine Oysters (*Crassostrea ariakensis*). *Front. Marine Sci.* **7**, <https://doi.org/10.3389/fmars.2020.589099> (2020).
- Crispo, E. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution* **61**, 2469–2479 (2007).
- Pfennig, D. W. et al. Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol. Evol.* **25**, 459–467 (2010).
- Schneider, R. F. & Meyer, A. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. *Mol. Ecol.* **26**, 330–350 (2017).
- 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. *Mol. Ecol.* **26**, 4591–4602 (2017).
- Xiao, Y. et al. The impact of Yangtze River discharge on the genetic structure of a population of the rock bream, *Oplegnathus fasciatus*. *Mar. Biol. Res.* **12**, 426–434 (2016).
- Buroker, N. E., Hershberger, W. K. & Chew, K. K. Population genetics of the family Ostreidae. II. Interspecific studies of the genera *Crassostrea* and *Saccostrea*. *Mar. Biol.* **54**, 171–184 (1979).
- Armarego-Marriott, T. Climatic selection and gene expression plasticity. *Nat. Clim. Change* **11**, 4–4 (2021).
- Kenkel, C. D. & Matz, M. V. Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nat. Ecol. Evol.* **1**, 0014 (2016).
- Li, L. et al. Divergence and plasticity shape adaptive potential of the Pacific oyster. *Nat. Ecol. Evol.* **2**, 1751–1760 (2018).
- Vargas, C. A. et al. Species-specific responses to ocean acidification should account for local adaptation and adaptive plasticity. *Nat. Ecol. Evol.* **1**, 84 (2017).

14. Yampolsky, L. Y., Schaer, T. M. & Ebert, D. Adaptive phenotypic plasticity and local adaptation for temperature tolerance in freshwater zooplankton. *Proc. Biol. Sci. R. Soc.* **281**, 20132744 (2014).
15. Sandoval-Castillo, J. et al. Adaptation of plasticity to projected maximum temperatures and across climatically defined bioregions. *Proc. Natl Acad. Sci. USA* **117**, 17112–17121 (2020).
16. Whitman, D. W. & Agrawal, A. A. *What is phenotypic plasticity and why is it important?* 1–63 (Science Publishers, 2009).
17. Eierman, L. E. & Hare, M. P. Reef-specific patterns of gene expression plasticity in Eastern Oysters (*Crassostrea virginica*). *J. Heredity* **107**, 90–100 (2016).
18. Grishkevich, V. & Yanai, I. The genomic determinants of genotype x environment interactions in gene expression. *Trends Genet.* **29**, 479–487 (2013).
19. Kelly, M. Adaptation to climate change through genetic accommodation and assimilation of plastic phenotypes. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 20180176 (2019).
20. 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 Biol. Evol.* **10**, 2518–2534 (2018).
21. Li, A., Li, L., Song, K., Wang, W. & Zhang, G. Temperature, energy metabolism, and adaptive divergence in two oyster subspecies. *Ecol. Evol.* **7**, 6151–6162 (2017).
22. Li, A., Li, L., Wang, W. & Zhang, G. Evolutionary trade-offs between baseline and plastic gene expression in two congeneric oyster species. *Biol. Lett.* **15**, 20190202 (2019).
23. Zhang, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
24. Beck, M. W. et al. Oyster reefs at risk and recommendations for conservation, restoration, and management. *BioScience* **61**, 107–116 (2011).
25. Guo, X. & Ford, S. E. Infectious diseases of marine molluscs and host responses as revealed by genomic tools. *Philos. Trans. R. Soc. London B Biol. Sci.* **371**, <https://doi.org/10.1098/rstb.2015.0206> (2016).
26. 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 (2004).
27. Wang, H. et al. Distribution of *Crassostrea ariakensis* in China. *J. Shellfish Res.* **25**, 789–790 (2006).
28. Zhou, M. F. & Allen, S. K. A review of published work on *Crassostrea ariakensis*. *J. Shellfish Res.* **22**, 1–20 (2003).
29. Belcaid, M. et al. Symbiotic organs shaped by distinct modes of genome evolution in cephalopods. *Proceedings of the National Academy of Sciences* **116**, 3030–3035 (2019).
30. Bai, C. M. et al. Chromosomal-level assembly of the blood clam, *Scapharca (Anadara) broughtonii*, using long sequence reads and Hi-C. *GigaScience* **8**, <https://doi.org/10.1093/gigascience/giz067> (2019).
31. Bao, Y. et al. Genomic insights into the origin and evolution of molluscan red-bloodedness in the blood clam *Tegillarca granosa*. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msab030/6126415> (2021).
32. Penalzoa, C. et al. A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *GigaScience* **10**, <https://doi.org/10.1093/gigascience/giab020> (2021).
33. Peng, J. et al. Chromosome-level analysis of the *Crassostrea hongkongensis* genome reveals extensive duplication of immune-related genes in bivalves. *Mol. Ecol. Resour.* **20**, 980–994 (2020).
34. Qi, H., Li, L. & Zhang, G. Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. *Mol. Ecol. Resources*, <https://doi.org/10.1111/1755-0998.13368> (2021).
35. Song, H. et al. The hard clam genome reveals massive expansion and diversification of inhibitors of apoptosis in *Bivalvia*. *BMC Biol.* **19**, 15 (2021).
36. Sun, J. et al. The Scaly-foot Snail genome and implications for the origins of biomineralised armour. *Nat. Commun.* **11**, 1657 (2020).
37. Varney, R. M., Speiser, D. I., McDougall, C., Degnan, B. M. & Kocot, K. M. The iron-responsive genome of the chiton *Acanthopleura granulata*. *Genome Biol. Evol.* **13**, <https://doi.org/10.1093/gbe/evaa263> (2021).
38. Zhang, Y. et al. The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-021-01448-6> (2021).
39. 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 Evol. Biol.* **10**, 394 (2010).
40. Kim, W.-J. et al. Mitochondrial DNA sequence analysis from multiple gene fragments reveals genetic heterogeneity of *Crassostrea ariakensis* in East Asia. *Genes Genom.* **36**, 611–624 (2014).
41. Liu, X. et al. Transcriptome and gene coexpression network analyses of two wild populations provides insight into the high-salinity adaptation mechanisms of *Crassostrea ariakensis*. *Mar. Biotechnol.* **21**, 596–612 (2019).
42. Xiao, J., Cordes, J. F., Wang, H., Guo, X. & Reece, K. S. Population genetics of *Crassostrea ariakensis* in Asia inferred from microsatellite markers. *Mar. Biol.* **157**, 1767–1781 (2010).
43. Li, C. et al. Genome sequences reveal global dispersal routes and suggest convergent genetic adaptations in seahorse evolution. *Nat. Commun.* **12**, 1094 (2021).
44. Dong, Y. & Somero, G. N. Temperature adaptation of cytosolic malate dehydrogenases of limpets (genus *Lottia*): differences in stability and function due to minor changes in sequence correlate with biogeographic and vertical distributions. *J. Exp. Biol.* **212**, 169–177 (2009).
45. 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 (2012).
46. Wang, H., Qian, L., Liu, X., Zhang, G. & Guo, X. Classification of a common cupped oyster from southern China. *J. Shellfish Res.* **29**, 857–866 (2010).
47. Wang, H., Zhang, G., Liu, X. & Guo, X. Classification of common oysters from North China. *J. Shellfish Res.* **27**, 495–503 (2008).
48. Kimura, M. Paleogeography of the Ryukyu Islands. *Tropics* **10**, 5–24 (2000).
49. Raeymaekers, J. A. M. et al. Adaptive and non-adaptive divergence in a common landscape. *Nat. Commun.* **8**, 267 (2017).
50. Qin, Y., Zhao, Y. & Zhao, S. *Geology of the Bohai Sea*. (Science Press, 1985).
51. Mekonnen, M. M. & Hoekstra, A. Y. Four billion people facing severe water scarcity. *Sci. Adv.* **2**, e1500323 (2016).
52. Skliris, N. et al. Salinity changes in the World Ocean since 1950 in relation to changing surface freshwater fluxes. *Clim. Dyn.* **43**, 709–736 (2014).
53. 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 rat kidney. *Am. J. Physiol.* **275**, F191–F197 (1998).
54. Huang, Y., Niwa, J., Sobue, G. & Breitwieser, G. E. Calcium-sensing receptor ubiquitination and degradation mediated by the E3 ubiquitin ligase dorf1n. *J. Biol. Chem.* **281**, 11610–11617 (2006).
55. 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 (2009).
56. Ribeiro, C. A., Balestro, F., Grando, V. & Wajner, M. Isovaleric acid reduces Na<sup>+</sup>, K<sup>+</sup>-ATPase activity in synaptic membranes from cerebral cortex of young rats. *Cell. Mol. Neurobiol.* **27**, 529–540 (2007).
57. Stern, D. B. & Lee, C. E. Evolutionary origins of genomic adaptations in an invasive copepod. *Nat. Ecol. Evol.* **4**, 1084–1094 (2020).
58. 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. *Sci. Rep.* **7**, 43575 (2017).
59. Weinman, E. J., Dubinsky, W. P. & Shenolikar, S. Reconstitution of cAMP-dependent protein kinase regulated renal Na<sup>+</sup>-H<sup>+</sup> exchanger. *J. Membr. Biol.* **101**, 11–18 (1988).
60. Zhou, X. et al. Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat. Commun.* **9**, 1276 (2018).
61. Barrio, A. M. et al. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. *eLife* **5**, e12081 (2016).
62. Hoglund, P. J., Nordstrom, K. J., Schioth, H. B. & Fredriksson, R. The solute carrier families have a remarkably long evolutionary history with the majority of the human families present before divergence of Bilaterian species. *Mol. Biol. Evol.* **28**, 1531–1541 (2011).
63. Kenkel, C. D., Meyer, E. & Matz, M. V. Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites astreoides*) from different thermal environments. *Mol. Ecol.* **22**, 4322–4334 (2013).
64. Guo, X., He, Y., Zhang, L., Lelong, C. & Jouaux, A. Immune and stress responses in oysters with insights on adaptation. *Fish. shellfish Immunol.* **46**, 107–119 (2015).
65. Zhang, G. et al. Molecular basis for adaptation of oysters to stressful marine intertidal environments. *Annu. Rev. Anim. Biosci.* **4**, 357–381 (2016).
66. Li, A. et al. Noncoding variation and transcriptional plasticity promote thermal adaptation in oysters by altering energy metabolism. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msab241> (2021).
67. Guo, X., Li, C., Wang, H. & Xu, Z. Diversity and evolution of living oysters. *J. Shellfish Res.* **37**, 755–771 (2018).
68. Xu, F. *Reproductive isolations between oysters of genus Crassostrea on the Xiaomiaohong oyster reef*. Doctoral thesis, Institute of Oceanology, Chinese Academy of Sciences, (2009). (In Chinese)
69. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
70. Jayakumar, V. & Sakakibara, Y. Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Brief. Bioinformatics* **20**, 866–876 (2019).
71. 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 Res.* **44**, e147 (2016).
72. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017).
73. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).

74. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *Bmc Bioinforma.* **19**, 460 (2018).
75. Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
76. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
77. Hunt, M. et al. REAPR: a universal tool for genome assembly evaluation. *Genome Biol.* **14**, R47 (2013).
78. Hoede, C. et al. PASTEC: an automatic transposable element classification tool. *PLoS ONE* **9**, e91929 (2014).
79. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 10 (2009).
80. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
81. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl 2), ii215–ii225 (2003).
82. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
83. Alioto, T., Blanco, E., Parra, G. & Guigo, R. Using geneid to identify genes. *Curr. Protoc. Bioinforma.* **64**, e56 (2018).
84. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
85. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
86. 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 (2018).
87. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
88. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
89. 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 (2006).
90. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
91. She, R., Chu, J. S., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009).
92. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988 (2004).
93. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids Res.* **25**, 955–964 (1997).
94. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
95. Griffiths-Jones, S. et al. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–D124 (2005).
96. Marchler-Bauer, A. et al. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).
97. Boeckmann, B. et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370 (2003).
98. Tatusov, R. L. et al. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–28 (2001).
99. Finn, R. D. et al. Pfam: clans, web tools and services. *Nucleic Acids Res.* **34**, D247–D251 (2006).
100. Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A. & Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **41**, e121 (2013).
101. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
102. Wang, Y. et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
103. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
104. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
105. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
106. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
107. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
108. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
109. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
110. 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 (2014).
111. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
112. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

## Acknowledgements

We thank X. Wang, Z. Jia, Z. She, Y. Zhang, Z. Yu, W. Quan, Z. Huo, X. Yan, Z. Zeng and Y. Ning for sample collection, and B. Yin and J. Qi for information on marine currents. L.L. is supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA23050402) and the National Key R&D Program of China (No. 2018YFD0900304). A.L. is supported by the National Natural Science Foundation of China (No. 32101353) 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). A.L. is also supported by the Distinguished Young Scientists Research Fund of Key Laboratory of Experimental Marine Biology, Chinese Academy of Sciences (No. KLEMB-DYS04). L.L. is also supported by the Technology and the Modern Agro-industry Technology Research System (No. CARS-49).

## Author contributions

L.L., G.Z. and X.G. conceived the study and participated in final data analysis, interpretation and manuscript writing. A.L. carried out 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., C.W. and X.W. collected and sampled oyster specimens. A.L. and W.W. produced the F<sub>1</sub> progeny. A.L., L.L., X.G. and G.Z. revised the manuscript. All authors approved the manuscript for publication.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02823-6>.

**Correspondence** and requests for materials should be addressed to Li Li or Guofan Zhang.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021