

# Genome-wide evolutionary signatures of climate adaptation in spotted sea bass inhabiting different latitudinal regions

Baohua Chen<sup>1</sup>, Zhixiong Zhou<sup>2</sup>, Yue Shi<sup>1</sup>, Jie Gong<sup>1</sup>, Chengyu Li<sup>1</sup>, Tao Zhou<sup>1</sup>, Yun Li<sup>3</sup>, Dianchang Zhang<sup>4</sup>, and Peng Xu<sup>1</sup>

<sup>1</sup>Xiamen University

<sup>2</sup>Xiamen University College of Ocean and Earth Science

<sup>3</sup>Ocean University of China

<sup>4</sup>Chinese Academy of Fishery Sciences

March 12, 2021

## Abstract

Temperature is an important climatic factor that shapes the distribution of eurythermal species. Thermal adaptation of species is important to both evolutionary biology and climate-change biology because it frequently leads to latitudinal gradients of various phenotypes among populations. Spotted sea bass (*Lateolabrax maculatus*) has a broad latitudinal distribution range along the marginal seas of the Northwest Pacific, providing an excellent teleost model for climate adaptation studies. We generated over 8.57 million SNP loci using whole genome re-sequencing from 100 samples collected at 14 geographic loci. We built the phylogeographic structure and demographic history of *L. maculatus* and determined sea surface temperature as the key environmental factor and major driving force for genetic divergence and local adaptation. We also identified distinct selective signatures and functional genes underlying adaptive mechanisms and ecological tradeoffs in the southernmost and northernmost populations inhabiting distinct climatic and latitudinal zones. The results offer an opportunity to better understand the genetic basis of the phenotypic variation in eurythermal fishes inhabiting different climatic regions.

## Genome-wide evolutionary signatures of climate adaptation in spotted sea bass inhabiting different latitudinal regions

Baohua Chen<sup>1,2</sup>, Zhixiong Zhou<sup>1,2</sup>, Yue Shi<sup>1,2</sup>, Jie Gong<sup>1,2</sup>, Chengyu Li<sup>1,2</sup>, Tao Zhou<sup>1,3</sup>, Yun Li<sup>4</sup>, Dianchang Zhang<sup>5\*</sup>, Peng Xu<sup>2,3\*</sup>

<sup>1</sup>State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China

<sup>2</sup>Shenzhen Research institute of Xiamen University, Shenzhen 518000, China

<sup>3</sup>Fujian Key Laboratory of Genetics and Breeding of Marine Organisms, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China

<sup>4</sup>The Key Laboratory of Mariculture, Ministry of Education, Ocean University of China, Qingdao, China

<sup>5</sup>South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510300, China

\*Corresponding author: Dr. Peng Xu, e-mail: xupeng77@xmu.edu.cn; Dr. Dianchang Zhang, email: zhangdch@163.com

## Abstract

Temperature is an important climatic factor that shapes the distribution of eurythermal species. Thermal adaptation of species is important to both evolutionary biology and climate-change biology because it frequently leads to latitudinal gradients of various phenotypes among populations. Spotted sea bass (*Lateolabrax maculatus*) has a broad latitudinal distribution range along the marginal seas of the Northwest Pacific, providing an excellent teleost model for climate adaptation studies. We generated over 8.57 million SNP loci using whole genome re-sequencing from 100 samples collected at 14 geographic loci. We built the phylogeographic structure and demographic history of *L. maculatus* and determined sea surface temperature as the key environmental factor and major driving force for genetic divergence and local adaptation. We also identified distinct selective signatures and functional genes underlying adaptive mechanisms and ecological tradeoffs in the southernmost and northernmost populations inhabiting distinct climatic and latitudinal zones. The results offer an opportunity to better understand the genetic basis of the phenotypic variation in eurythermal fishes inhabiting different climatic regions.

Keywords: Spotted sea bass, phylogeographic structure, climate adaptation

## Introduction

Temperature is one of the most critical climatic factors for ectothermic animals, affecting their growth, development, reproduction and distribution. Teleosts are a main category of ectothermic animals with ranges of temperature tolerance that vary widely from species to species. Teleosts can be divided into two subcategories, eurythermal fishes and stenothermal fishes. Habitats of most stenothermal fishes are confined in regions with stable water temperatures, such as tropical or polar zones. In contrast, eurythermal fishes can tolerate greater temperature fluctuations than stenothermal fishes and usually disperse over broad latitudinal zones. Many studies have suggested that eurythermal fishes living in different latitudes undergo spatially heterogeneous natural selection (Bradbury et al., 2010; Hasselman, Ricard, & Bentzen, 2013; Lannig et al., 2003), which leads to population structure variation and inheritable phenotype differentiation linked with optimized fitness in response to selective pressures imposed by their local environment (Kawecki & Ebert, 2004; Leimu & Fischer, 2008). In addition to the population structure being shaped by environmental fluctuations, current structured populations probably play a role in the response to large-scale environmental change, such as global climate change (Zakharov & Hellmann, 2008). A sufficiently high rate of gene flow among populations could give rise to common responses to environmental changes. In contrast, fully differentiated populations with distinct genetic components may respond to climate change in different ways because of their accumulated local adaptation and fitness variation (Davis, Jenkinson, Lawton, Shorrocks, & Wood, 1998).

Many ecological studies have revealed that ectothermic animals distributed at high latitudes tend to have a higher capacity for growth (i.e., maximum growth potential) than their low-latitude conspecifics ranging from invertebrates to fish (D. O. Conover, Brown, & Ehtisham, 1997; D. O. Conover & Present, 1990; Imsland et al., 2000; Levins, 1969; Lindgren & Laurila, 2005). Such counter-gradient growth variations in teleosts are presumably due to the limited length of the growing season for high-latitude dwellers. A high growth rate in the growing season may help fish evade size-selective winter mortality (fish with large body sizes have a higher probability of overwintering), a phenomenon that is present in many high-latitude teleost species (David O. Conover, 1990). The loosed lethal cold stress experienced by low-latitude populations results in reduced predation frequency, which decreases the risk of mortality and allocates more living resources to other biological processes, such as movement and reproduction, rather than growth, reflecting the tradeoffs in physiological and life historical traits. These kinds of tradeoffs have been investigated in some teleost species distributed over a relatively wide latitudinal range. Atlantic silversides (*Menidia menidia*) belong to the high-latitude population of Nova Scotia and were found to have a significantly higher growth rate than southern conspecifics, and this high-latitude fish had much lower maximum prolonged and burst swimming speeds (Billerbeck, Lankford, & Conover, 2001). A study conducted in medaka (*Oryzias latipes*) obtained very similar results: medaka living at high latitudes also makes a tradeoff between a high growth rate and vulnerability to predation in the same way as Atlantic silverside.

Spotted sea bass (*Lateolabrax maculatus*) was recently discriminated from Japanese sea bass (*L. japonicus*

) as a newly re-characterized species with obvious morphological and genetic differences (Yokogawa, 1998). Taking advantage of its eurytherm and euryhaline characteristics, *L. maculatus* is distributed along the Chinese coast with a wide range spanning from the tropical Beibu Gulf in the South China Sea to the mid-temperate Bohai Gulf (**Figure 1a**). Such a broad latitudinal range results in a highly variable sea surface temperature (SST) of the habitats of different populations. The SST can drop to near the freezing point in the northern Bohai Gulf in winter but remain as high as 15 degrees Celsius in the Beibu Gulf. The SST difference between the above two areas is very small in summer. Similar to other ectothermic animals, many of the characteristics of *L. maculatus* inhabiting high- and low-latitude areas differ, implying tradeoffs in life-history strategies. The northern population in the Bohai Gulf have a stronger growing potential, motor ability and higher cold resistance than their southern conspecifics in the South China Sea (Y. B. Hu et al., 2019). Phenotype variation and biogeographic divergence make this species a proper model for population genetics and adaptive evolution studies on fish species in continual marginal seas. In the last few decades, various molecular tools have been employed in population genetic and phylogeographic studies on *L. maculatus*, such as mitochondrial gene sequences (J. X. Liu, Gao, Yokogawa, & Zhang, 2006), isozymes (Gao, Dong, Liu, Zhang, & Koji, 2004), random amplified polymorphic DNAs (RAPDs) (Z. M. Hu, Gao, Han, & Song, 2007) and microsatellites (Jiang et al., 2009). A recent population genetic study employing double digest restriction-site associated DNA (ddRAD) revealed distinct population structures and genetic divergences of the northern population inhabiting the Bohai Gulf and the southern population inhabiting the Beibu Gulf (Zhao, et al, 2017), implying potential differences in the adaptive mechanisms and ecological strategies in response to significant differences in the local environment and climate.

Herein, we investigate the genomic signatures underlying climate adaptation in *L. maculatus* inhabiting a broad latitudinal zone along the coastline of China. We performed whole genome re-sequencing on *L. maculatus* samples collected from 14 populations and collected genome-wide single nucleotide polymorphisms (SNPs), which allowed us to build the phylogeographic structure and demographic history of *L. maculatus* and determine the key environmental factors serving as the major driving force of genetic divergence and local adaptation. Selective signatures and adaptive genetic variants were also identified and illustrated as potential mechanisms and strategies in response to climate divergence. The results will improve the understanding of the adaptive mechanism and genetic basis of phenotypic variation in changing environments in eurythermal fishes.

## Results and Discussion

### Reference genome improvements and genotyping

A high-quality reference genome is crucial for genome-wide detection of SNPs, construction of a credible population structure and high-resolution selective signature scanning. Our previous reference genome assembly was completed with high accuracy and contiguity at the scaffold level (GenBank accession: GCA\_004028665.1) (B. H. Chen et al., 2019). However, obsoleted software, LACHESIS, was employed in processing the raw Hi-C data and chromosome-level scaffolding, which may have introduced sub-optimal orderings and orientations (Zhang, Zhang, Zhao, Ming, & Tang, 2019). We therefore re-assembled the reference genome at the chromosome level to correct potential assembly mistakes by employing the recently developed software 3D-DNA (Dudchenko et al., 2017). This improved genome assembly contained 480 scaffolds with a total length of 589.99 mega bases, showing higher integrity than the previous assembly. The integration level was also slightly improved to 98.76% (**Figure S1, Table S1**). The new assembly showed a significantly increased concordance with the linkage map of *L. maculatus* (Y. Liu et al., 2020) and the pseudo-chromosomes assembled based on the Asian seabass reference genome (Vij et al., 2016), providing an accurate reference genome for downstream population genomics and genome evolution studies (**Figure S2**). The improved reference genome assembly has been updated in GenBank with the accession number GCA\_004028665.2.

### Population structure and phylogenetic analysis

In total, 8.57 million SNPs were obtained from 100 fish collected from 14 continuous sites along the coastline of China (**Figure 1a, Figure S3, Table S2-4**). To reduce the redundant information of highly linked

SNPs, we produced a pruned subset of SNPs containing 3.373 million loci that were in approximate linkage equilibrium with each other. Principal component analysis across all geographic sampling sites was performed based on these SNPs. The results showed that all populations were clustered into 3 groups with the first two principal components. The first principal component that accounted for 2.17% of the variance isolated FC, HK and TS samples from all other samples. The second principal component that accounted for 1.21% of the variance further divided the rest of the samples. There was no overlap among the three confidence ellipses, implying obvious genetic differentiation among these three clusters. The Bohai (BH) populations included samples from three sample sites (TJ, YT and WD) distributed in the Bohai Gulf and the northern Yellow Sea; the Beibu Gulf (BB) populations included samples from three sample sites (FC, TS and HK) restricted to the Beibu Gulf; and the intermediate populations (IM), including all other sample sites except SZ (LY, LS, SZ, WZ, DS, ST and ZJ, from north to south), ranged from the southern Yellow Sea to the northern South China Sea. Cluster IM was more decentralized than the other two clusters in the scatter plot because it contained more geographical populations rather than higher genetic diversity. The gene diversity ( $H$ ) of each IM population was lower than that of the BH populations but higher than that of the BB populations. Intriguingly, the IM populations had significantly higher heterozygosity than the other populations on both observed heterozygosity  $H_O$  and expected heterozygosity  $H_E$  (**Figure S2**), which might result from genomic admixture of the two distinct clusters of BH and BB. To relatively quantitatively assess the genomic admixture among clusters and populations, we estimated the ancestry proportions of each sample. IM populations were depicted as a series of admixtures of the significantly divergent BH and BB populations when K was set as 2, while independent ancestry was depicted when K was set as 3. Higher-level gene flow was observed between the IM and BH populations than between the IM and BB populations (**Figure 1d**).

We performed phylogenetic analysis using all polymorphic SNPs in all samples with (**Figure 1c**) and without (**Figure S4**) taking a related species (Asian sea bass, AS) as the outgroup. The rooted phylogenetic tree showed that the IM and BB populations were spread from the BH populations and in an incomplete parapatric speciation process. In addition, historical population effective sizes were estimated to discover key events and time nodes shaping the population structure of spotted sea bass during this process. We found that BB populations might have experienced rapid expansion 410,000 years ago followed by a bottleneck approximately 290,000 years ago. The period bound between these two historical events was surprisingly similar to the boundary between Beiweitan II marine transgression and regression events that occurred in the South China Sea in the Middle Pleistocene (Zhaoshu Liu, 2002) (**Figure 1e**). The BB populations were chronically isolated from other populations in the South China Sea before the Beiweitan II marine transgression event due to the closure of the Qiongzhou Strait between Hainan Island and the Leizhou Peninsula (Zhaoshu Liu, 2002). This Beiweitan II marine transgression reconnected the Beibu Gulf and South China Sea via the emerged Qiongzhou Strait, inflated the BB populations and intensified the gene flow between the BB populations and IM populations. The Beiweitan II marine regression event severely vanished the whole Beibu Gulf and subsequently introduced a dramatic decrease in the effective size of the BB populations.

Both the phylogenetic topology and population structures of *L. maculatus* populations along the coastline of China indicated that the Shandong Peninsula could be the primary physical barrier for gene flow between the BH and IM populations, and the Leizhou Peninsula and Hainan Island could be major barriers for the BB and IM populations. *L. maculatus* lives in near-shore waters with limited migration ability (Sun et al., 2014). Physical barriers such as peninsulas and islands, together with coastal currents, limit gene flow and contribute to distinctive genetic divergence in isolated BH and BB populations. The pairwise fixation indexes between the ZJ and BB populations were much higher than those between the LS and BH populations, indicating that the Leizhou Peninsula and Hainan Island block gene flow more than the Shandong Peninsula. As isolated marginal populations, the genetic diversity ( $H$ ) of BB populations was conceivably much lower than that of other populations (**Table S5**), which was consistent with previous reports based on much fewer genetic markers (Gong et al., 2020; Zhao et al., 2018). *L. maculatus* is a euryhaline fish and is capable of living in a wide range of salinities. Therefore, we did not find evidence of gene flow disturbance caused by massive

discharge of freshwater from great rivers, such as the Yangtze River and the Pearl River, as has been found for other marine species.

Notably, the SZ population that we sampled from the South China Sea was clustered or grouped with BH populations rather than IM populations in both phylogenetic analysis and population structure (**Figure 1**), providing evidence that escaped farmed sea bass has replaced the native population in the estuary of the Pearl River. Spotted sea bass is one of the most important aquaculture species, with an annual yield greater than 150,000 tons, ranking as the second most abundant aquaculture marine fish in China. The estuary area of the Pearl River contributes approximately two-thirds of the spotted sea bass production in China. Germplasms from northern populations have a favourably high growth rate and have been translocated to the Pearl River estuary and dominated sea bass aquaculture in recent decades. All five SZ samples were collected from Dapeng Bay in the Pearl River Delta. Their genetic components were identical to those of the samples from northern populations in BH, suggesting that they were either escaped farm sea bass or their descendants. Escape prevention for farmed sea bass is urgently needed in the Pearl River estuary area.

### Isolation by distance and environment analysis

To explore the main force driving the genetic differentiation of *L. maculatus* populations, we performed isolation-by-distance (IBD) and isolation-by-environment (IBE) analyses based on pairwise fixation indexes (*Fst*) and extragenetic variables. In IBD analysis, we modelled the linear relationship between the genetic distance (pairwise *Fst*) and pairwise geographic distance of all populations. For most IBD analyses, the great-circle distance ( $D_{gcc}$ , also called the crow-fly distance) is typically used as the geographic distance. We additionally considered the coastline length ( $D_{cst}$ ) between sample sites in the analysis since we presume it may be more meaningful to the biological characteristics of fishes. Models were obtained with P-values much smaller than typical significance thresholds when *Fst* exhibited pairwise correlation with  $D_{cst}$  (P-value =  $2.91E-13$ ) or  $D_{gcc}$  (P-value =  $9.74E-12$ ). However, a model was obtained with a higher coefficient of determination ( $R^2$ ) and a better relative quality indicated by the Akaike information criterion (AIC) when  $D_{cst}$  was used rather than  $D_{gcc}$  ( $0.506$  vs.  $0.459$  for  $R^2$  and  $-612.91$  vs.  $-605.80$  for AIC, **Table 1**). When both were included in one model,  $D_{cst}$  was significantly correlated with *Fst* (P-value= $0.003$ ), while  $D_{gcc}$  was not (P-value= $0.141$ ). Our results verified the hypothesis that coastline length ( $D_{cst}$ ) between sampling sites is a proper method for IBD analysis for movable species living in coastal waters (**Figure 2a**, **Figure S5**).

Genetic differentiation will be enhanced when genetic drift acts within populations more quickly than it is ameliorated by gene flow between populations. Therefore, an increase in geographic distance will lead to increased genetic differentiation by reducing the effective dispersal rate between populations, which is a frequent occurrence in widespread natural populations (Meirmans, 2012). As we mentioned above, the limited migration ability of spotted sea bass should introduce a strong linear correlation between distance along the coastline ( $D_{cst}$ ) and genetic distance. Although  $D_{cst}$  showed the most significant positive correlation with *Fst* values, the coefficient of determination of this correlation was somehow lower than expected in our result. The most reasonable explanation is that their genetic differentiation was enhanced by environmental heterogeneity, which was summarized as “isolation by environment” (Wang, 2012). For teleosts, which are exclusively poikilothermic species, temperature is the single most crucial environmental factor impacting assorted physiological activities such as growth, reproduction and movement (Johansen & Jones, 2011; Pankhurst & Munday, 2011; Rountrey, Coulson, Meeuwig, & Meekan, 2014).

We therefore observed pairwise correlations of *Fst* with three climatic factors: annual maximum sea surface temperature ( $SST_{max}$ ), which represents extremely hot summer climates, annual minimum SST ( $SST_{min}$ ), which represents extremely cold winter climates, and annual median SST ( $SST_{med}$ ), which represents long-term mild climate differences between sample sites. Our results showed that all three climatic factors were significantly correlated with genetic differences.  $SST_{max}$  had the highest significance (P-value =  $9.76E-15$ ) and goodness of fit for the model ( $R^2 = 0.548$  and AIC =  $-619.80$ , **Figure 2b**, **Figure S6**), both of which exceeded those obtained for  $D_{cst}$ , the most significant geographic factor. Nevertheless, we cannot simply deduce that an extremely hot summer climate is the most important factor in the Beibu Gulf and is the main driving force of genetic differentiation. There was unavoidable internal correlation between geographic

factors and climatic factors when populations were distributed approximately along a latitudinal gradient (**Figure 2c**). This result means that both geological and climatic differences always contributed to the effectiveness of the model, regardless of whether either kind of exogenous factor was used. Furthermore, drift and selection, two main driving forces of genetic differentiation, are always interlinked in nature. Their separation has been a goal of many genetic studies. We tried to use multiple linear regression to model the genetic differentiation of *L. maculatus* simultaneously driven by geological and climatic factors. As a result, the best model that was constructed contained three significant factors,  $D_{cst}$ ,  $SST_{max}$  and  $SST_{min}$ , whose P-values were 9.57E-11, 1.49E-10, and 6.81E-05, respectively. The goodness of fit of this model ( $R^2 = 0.757$  and  $AIC = -664.23$ ) also exceeded that of any single-factor model (**Figure 2d, Table 1**). Although we still cannot compare the influences of these three factors in this way because the predictive power or reliability of the calculations regarding individual predictors were reduced when multicollinearity was exorbitant among predictors, this result also indicated that the extremely cold water in northern areas in winter is a potential driving force of genetic differences together with geographic distance and high temperatures in summer. Most laboratory temperature tolerance estimates for fishes involve high temperatures, whereas the death of most fish in nature has been caused by exposure to low temperatures (Beitinger, Bennett, & McCauley, 2000). The death rate of sea bass, unsurprisingly, was higher in winter than in summer owing to their decreased swimming ability, decreased ingestion and growth rates, increased accumulated mortality during the prolonged early period and weakened retention mechanism within the estuarine turbidity maxima zone (Shoji & Tanaka, 2007).

### Candidate genomic regions under extreme positive selection

Highly differentiated SNPs between the BH and BB populations were identified with high  $F_{ST}$  values. The results showed that overall, only a few loci had nearly approached fixation between these two populations. On account of the lack of knowledge of the distribution of  $F_{ST}$  values, we simply selected the top 2% of SNPs with the highest  $F_{ST}$  values (171,868 SNPs, with a low threshold of  $F_{ST}$  [?] 0.201) as candidate SNPs under positive selection. Then, the weighted average  $F_{ST}$  values of the SNPs were calculated as described by Weir and Cockerham in every 50 kb sliding window with a sliding step size of 10 kb, and the top 5% of windows with the highest  $F_{ST}$  values (with a threshold of 0.2367) were selected as candidate genomic regions differentiated in the allele frequency spectrum (**Figure 3a and 3b**). However, both natural selection and random drift can cause differences in the allele frequency spectrum. Therefore, we complemented this analysis by calculating the ratio of nucleotide diversity ( $\omega$ ) and contrast in the extent of haplotype homozygosity ( $R_{sb}$ ) to eliminate the effect of random drift on the detection of selective signals. Low nucleotide diversity indicates the result of the selective sweep process of positively selected mutations. The absolute  $R_{sb}$  index will be inflated when the presentation of an allele extends the length of homozygosity in either population of the comparison. In this way, the  $R_{sb}$  index highlights recently or nearly fixed selective loci. To integrate the  $R_{sb}$  index with  $F_{ST}$  and  $\omega$ , we simply averaged it in every 50 kb sliding window (with a step size of 10 kb). Because of the lack of reliable distributions of the ratio of  $\pi$  and averaged  $R_{sb}$ , we selected windows with the top 2.5% on both sides of the ratio of  $\pi$  and  $R_{sb}$  as candidate genomic regions (**Figure 3, Figure S7 and S8**). Only if a window was included by all three genome scanning methods was it considered to be undergoing or have undergone severe positive selection in the BH or BB population. In summary, 1,901 50 kb windows were identified as under natural selection, occupying a 13.65 Mb genome region (**Figure 3b, Table S6**) distributed on all chromosomes except chromosome 15.

In our results, 287 and 227 protein-coding genes were detected as positively selected genes (PSGs) by overlapping with genomic regions with positive selection signatures (**Table S7 and S8**) in the BH and BB populations, respectively (6 genes were in both populations). In addition, all non-synonymous SNPs with significant differences in allele frequency were identified on all PSGs (**Table S9**). To further explore which biological processes were affected by positive selection, we conducted overrepresentation analysis (ORA) of Gene Ontology (GO) categories or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways based on PSGs in the two populations (**Table S10 and S11**). Insulin-like growth factor binding (GO:0005520), cytoskeleton (GO:0005856) and 11 other GO terms related to these two GO terms closely or remotely were overrepresented in the BH PSGs, involving 26 independent genes. Another 10 BB PSGs were enriched in

5 cell-adhesion-related GO terms, such as cell-cell adhesion via plasma membrane adhesion molecules (GO: 0098742) and homophilic cell adhesion via plasma membrane adhesion molecules (GO: 00007156).

### Selective signatures on GH/IGF-I axis-associated genes

We identified positively selected genes in BH populations that were significantly enriched in the GO term “regulation of insulin-like growth factor receptor signaling pathway” (GO:0043567, p-value = 1.476E-04, **Figure 4a and 4b**). Four insulin-linked growth factor-binding protein (Igfbp) genes (*igfbp1*, *igfbp2b*, *igfbp3* and *igfbp5*) were positively selected in the BH populations. IGFBPs determine the actions of IGFs, a series of potent mitogens in vertebrates mainly secreted by the liver as a result of stimulation by growth hormone (GH), by competitively binding IGFs and preventing their binding to IGF receptors (IGFR) or potentiate the effects of IGFs. Consequently, the interactions between IGFBPs, IGFs and IGFRs are restated as the GH-IGF-IGFBP axis, a powerful, complex system that regulates the cell growth of various organisms and tissues. Various studies have indicated that *igfbp* genes play crucial roles in skeletal, muscle and nervous growth in teleosts (several cites). In addition, we identified selective signatures of the suppressor of the cytokine signalling-2 (Socs-2) gene, an important negative regulator of the GH/IGF-I axis, which can competitively bind to Igf-1r to limit the action of Igf-I. Multiple *in vivo* experiments have provided substantial evidence that growth is enhanced when *SOCs-2* expression is inhibited or reduced in mice (Greenhalgh et al., 2002; Horvat & Medrano, 2001). Recent studies also indicated that *socs-2* plays a vital role in regulating *igf-1* expression in carp and tilapia, implying its role in growth regulation in teleosts.

BH populations can tolerate extremely cold sea water in the Bohai Gulf, where SST often reaches the freezing point in winter. BH populations are well recognized for their fast growth rate, e.g., common garden experiments indicated that *L. maculatus* from northern populations had a significantly higher growth rate (2.4 g/d) in the first year than their southern conspecifics (1.7 g/d) (D.-g. Chen, Gao, Zeng, Ren, & Ruan, 2001). The germplasm of BH populations has been dominantly exploited in the sea bass aquaculture industry in recent decades. Our findings of the selective signatures in GH/IGF-I axis-associated genes unveil the potential genetic basis of the high growth rate of BH populations.

Adaptation to cold environments repeatedly gives rise to thermal reaction norms for the growth rate. In the majority of instances, populations living at low temperatures tend to grow more rapidly than those living in warm environments (Angilletta, Niewiarowski, & Navas, 2002; D. O. Conover & Schultz, 1995). It has been widely reported that an improved growth rate increases the probability of survival of fishes from size-selective winter mortality in cold zones (Hurst & Conover, 1998), which is comprehensive but readily comprehensible. The major threats to creatures living in northern zones are cold conditions and starvation in deep winter. A large body size allows these creatures to accumulate more energy to protect against starvation and helps them maintain their body temperature by reducing the surface-to-volume ratio. On the one hand, sublethal cold stress allows these creatures to allocate more living resources to other biological processes, such as movement, reproduction and sensation, rather than growth, reflecting the tradeoffs in physiological and life historical traits (Angilletta & Dunham, 2003). The evolutionary signatures of growth regulation genes in BH populations underpin the adaptive strategies in spotted sea bass to cope with cold environments in the Bohai Gulf.

### Selective signatures on fibre muscle components

Although a high growth rate helps sea bass escape size-selective winter mortality in northern areas, it also comes at the cost of some disadvantages in terms of environmental fitness. For example, a high growth rate is highly dependent on a high frequency of foraging and decreased non-essential energy expenditure (Suzuki, Miyake, & Yamahira, 2010), which may drastically increase the risk of predation. Swimming ability therefore severely affects the success rates of predation and evasion and plays an important role in the cold acclimation of spotted sea bass. We identified substantial selective signatures of genes encoding fibre muscle components in BH populations, including *myh2*, *titin* and *cofilin-2* (**Figure 5a, Figure S9, Table S7**). All of these genes encode basic structural components or regulatory proteins of sarcomeres, the smallest contractile unit of striated muscle (Henderson, Gomez, Novak, Mi-Mi, & Gregorio, 2017), implying their potential roles in

cold climatic adaptation in high-latitude regions.

Intriguingly, we identified two SNP loci (c.373C>A and c.700G>A) in the myosin heavy chain gene *myhz2* with shifted allele frequencies in three BH populations compared to all other populations (**Figure 5b**). The myosin heavy chain consists of three domains: the head region, the tail region and the neck region. The head region contains a motor domain that binds filamentous actin and drives muscle contraction via ATP hydrolysis. We found that both SNPs are non-synonymous and cause amino acid substitutions (p.K125Q and p.I234V) in the head region of the myosin heavy chain (**Table S9**). Protein structure prediction showed that the substitutions did not change the tertiary structure of the myosin heavy chain (**Figure 5c**). However, their thermal stabilities were significantly changed, as indicated by distinct Gibbs free energy curves (**Figure 5d**). The two amino acid substitutes drastically increased the changes in the Gibbs free energy ( $\Delta G$ ) of the myosin heavy chain and improved its thermal stability in the BH populations. The increased thermostability of myosin may increase the stored potential energy, which will be released after its conformational change, thus resulting in strengthened muscle contractions. The shifting of thermal stability of myofibrillar proteins has been reported as an adaptive mechanism to cope with environmental temperature changes in teleosts and other vertebrates (Howell, Matthews, & Donnelly, 1991; Rodgers, Karr, Biedermann, Ueno, & Harrington, 1987). In addition, we identified substantial selective signatures of genes regulating muscle conduction, including *kcnd1*, *kcnd2* and *itpr3*, which also play a vital role in the athletic ability of animals. Overall, the evolutionary signatures of sarcomere-related genes in BH populations of spotted sea bass unveiled the adaptive mechanism of intrinsic muscle contractile properties for thermal acclimation in teleosts.

### Selective signatures on visual sense-related genes

Compared to the harsh environment in the high-latitude Bohai Gulf, the physical environment in the tropical Beibu Gulf is mild in terms of temperature, food biomass and photoperiod. Where physical conditions are appropriate, the interrelationships between species may become a paramount adaptive problem. For example, predator-prey interactions, one of the most important interspecific interactions driving the evolution of animal behaviour (Culumber, 2020), are widely considered to be stronger in tropical areas than in high-latitude zones (Schemske, Mittelbach, Cornell, Sobel, & Roy, 2009). Predator-introduced stress elicits various behavioural responses in fish, such as rapid locomotion (Camarillo, Rodriguez, & Tobler, 2020), prolonged concealment (Thurrow, Peterson, Chandler, Moffitt, & Bjornn, 2020) and reduced courtship behaviour (Uzendoski, Maksymovitch, & Verrell, 1993), to improve the probability of eluding and escaping from predators. In addition, the latitudinal gradient of anti-predator behaviours has been confirmed by studies in many species (Andrade & Blumstein, 2020).

We identified substantial selective signatures in BB populations inhabiting tropical regions harbouring 203 functional genes (**Table S8**). We conducted functional enrichment analysis of these positively selected genes in the BB populations and surprisingly found that many enriched categories were related to visual sense development and regulation, such as “fibroblast growth factor receptor signaling pathway” (GO:0008543), “embryonic neurocranium morphogenesis” (GO:0048702), “ceramide biosynthetic process” (GO Biological Processes) and “embryonic eye morphogenesis” (GO:0048048) (**Figure 6a and 6b, Table S11**). We further investigated visual sense-related genes under significant evolutionary selection in the BB populations, such as *fgf8*, *vax1*, *pard3ab*, *foxg1* and *ipo13* (**Figure 6c, Figure S10, Table S8**). *Fgf8* has been suggested to play a vital role in retinal differentiation and lens and optic nerve morphogenesis in zebrafish (cites). Interestingly, we also identified selective signatures in the genes correlated to *fgf8*, such as *ptpro* (Liao et al., 2013), *htra1a* (Kim et al., 2012), and *trim71* (Huttlin et al., 2017) (**Figure S10, Table S8**). *Vax1* is a novel homeobox-containing gene that directs the development of the basal forebrain and visual system (Hallonet, Hollemann, Pieler, & Gruss, 1999; Mui, Kim, Lemke, & Bertuzzi, 2005). *Pard3ab* has been reported to be required for cilia growth, separation of the eye fields and retinal lamination in zebrafish (Krock & Perkins, 2014; Wei et al., 2004). We recognized a non-synonymous SNP locus (c.37A>G) in *pard3ab* with significantly shifted allele frequencies in the BB populations compared to the northern populations (**Figure 6d, Table S7**), implying that functional variation is potentially linked to visual adaptation in spotted sea bass in



tropical regions. Visual sense influences the environmental sensitivity of fish by determining their ability to detect predators and prey. The selective signatures of visual sense-related genes suggest that they could serve as an important adaptive mechanism in response to highly intense predator-prey interactions in tropical habitats.

## Materials and Methods

### Refinement of chromosome-level scaffolding of the *L. maculatus* genome

The scaffold-level genome assembly was downloaded from the NCBI Genome Assembly database (Accession ID: ASM402866v1). A total of 163.3 Gb raw reads were downloaded from the NCBI Short Reads Archive database (accession ID: SRR8240566). Raw reads were filtered out by SolexaQA++ (version v.3.1.7.1) if the proportion of uncertain bases exceeded 10% of the total length, the proportion of low-quality bases exceeded 50% of the total length or the reads contained sequencing adaptors. First, we used Juicer (version 1.5.5) (Durand, Shamim, et al., 2016) to transform raw sequencing data into a list of Hi-C contacts (pairs of genomic positions that were adjacent to each other in 3D space during the experiment). Then, we employed a 3D-DNA pipeline (Dudchenko et al., 2017) to generate a candidate chromosome-level assembly. Finally, the candidate assembly was reviewed and finalized manually based on an associated contact map using Juicebox (version 1.5.3) (Durand, Robinson, et al., 2016).

### Comparison between refined and original chromosome-level scaffolding

To validate the correctness of the refined and original chromosome-level scaffolding, we compared both to a high-density linkage map of *L. maculatus* (Y. Liu et al., 2020) and a chromosome assembly of a related species, Asian seabass (*Lates calcarifer*). The Spearman's rank correlation coefficient between the original serial number of markers in the linkage map and the rearranged marker serial number of both assemblies for each chromosome was determined. Such coefficients can indicate which assembly has the best collinearity with the linkage map and hence the highest correctness. Similarly, we also employed the Spearman's rank correlation coefficient to indicate high collinearity of alignments, which were generated by Mummer4 (version 4.0.0beta2) (Marcais et al., 2018), between chromosomes of *L. maculatus* and *L. calcarifer*. Correlations were performed using an in-house Python script with basic statistical functions imported from the SciPy stats package.

### Sampling and sequencing

We performed whole genome re-sequencing on a total of 100 *L. maculatus* samples collected from 14 continuous sites along the coastline of China (**Figure 1a**). White muscles were collected from all fish after eugenol anaesthesia (200 mg/L for 15 min) and lysed in SDS digestion buffer with proteinase K for DNA extraction using the phenol-chloroform method. Sequencing libraries with 250 bp insert sizes were constructed according to the manufacturer's instructions (Illumina, San Diego, CA, USA), and paired-end sequencing was performed using the Illumina HiSeq 4000 platform with a read length of  $2 \times 150$  bp. All reads containing adaptor sequences were discarded before the removal of uncertain or low-quality bases using SolexaQA++. The retained and trimmed reads were used for variant calling. High-throughput sequencing generated 3.97 billion pairs of raw reads with a total length of 1.19 trillion bases and an average read depth of 19.64X per sample. Quality control retained 98.42% (1.17 T bases) of all bases for downstream steps. A total of 7.836 G of clean reads were mapped to the reference genome of *L. maculatus* with mapping ratios ranging from 97.2% to 99.1% and for genotyping. Finally, we collected 8.57 million non-redundant SNPs.

### Variant calling and filtering

Variant calling was conducted under the guidance of GATK Best Practices recommendations (Van der Auwera et al., 2013). There were three stages in our workflow: data pre-processing, variant discovery, and callset filtering. Pre-processing of data was performed independently and in parallel for each sample. First, we aligned clean reads to the newly refined reference genome of *L. maculatus* using the BWA-MEM algorithm (Li & Durbin, 2009) (version 0.7.17-r1188) with default parameters. Then, we used Picard implementation (version 2.18.0) MarkDuplicates to perform duplicate marking followed by SortSam to sort the reads. In

the variant discovery stage, we first called the variants per sample to an intermediate file called a GVCF using HaplotypeCaller in GATK. Before joint genotyping, we consolidated the GVCF files using GenomicsDB-Import across all samples to improve scalability and speed. Then, joint genotyping was conducted using GenotypeGVCFs, and a squared-off matrix of genotypes was produced that provided information about all sites in all samples, which was important for downstream callset filtering. Since there were no genome-wide known sites or genotyped data in the *L. maculatus* genome, we performed hard filtering on this callset instead of the recommended step, Variant Quality Score Recalibration (VQSR). SNPs with less ideal quality were marked using VariantFiltration with a compound filtering expression “QD < 2.0 || QUAL < 30.0 || SOR > 3.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0” and then removed by SelectVariants. All abovementioned tools and programs except BWA were included in a powerful toolkit, GATK (McKenna et al., 2010) (version 4.0.2.1). Finally, we performed an extra step of filtering on the resulting SNPs using VCFtools (version 0.1.15) (Danecek et al., 2011) to remove SNPs that were multiallelic, with minor allele counts less than 2, genotype missing counts greater than 2 or minor allele frequency (MAF) less than 0.05. The final set of SNPs was annotated with SnpEff (version 4.3t) (Cingolani et al., 2012).

### Population genetics analysis

To reduce the redundant information of highly linked SNPs, we first purged the final set by applying plink (version 1.90b6.16) (Purcell et al., 2007) with the parameter “--indep-pairwise 10 2 0.8”, which pruned all pairs of variants within 10 kbp windows with squared correlation greater than 0.8. We employed VCFtools (version 0.1.17) (Danecek et al., 2011) to filter the pruned SNP set with MAF [?] 0.05 per population to calculate the number of polymorphism loci and then calculated the *Tajima’s D* for each population. Observed heterozygosity (Ho), gene diversity (Hs) and Fis statistics were calculated using R package hierfstat (version 0.5-7). CoreArray Genomic Data Structure (GDS) data files were generated from variant call format (VCF) files using the R package “gdsfmt” (version 1.22.0) and then imported to the R package “SNPRelate” (version 1.16.0) (Zheng et al., 2012) for principal component analysis. The software frappe (version 1.0) (Tang, Peng, Wang, & Risch, 2005) (version 1.1) was used to estimate the genetic ancestry of each sample, which has high efficiency when large SNP genotype data are used. The maximum iteration of expectation maximization (EM) was set as 100,000, and the number of populations (K) was set from 2 to 4 for each calculation. The neighbour-joining tree of all samples were constructed using a command-line-based collection of utilities, VCF-kit (version 0.1.6) (Cook & Andersen, 2017). The size history of all populations was estimated from whole genome sequence data by SMC++ (version 1.15.4) (Terhorst, Kamm, & Song, 2017).

### Isolation by distance and environment analysis

In this study, the length of the coastline was estimated on the basis of pre-processed OpenStreetMap data obtained from Planet OSM (<https://osmdata.openstreetmap.de>). Downloaded shapefiles with the “Mercator” projection were transformed to the R data frame using the package “sf” (version 0.9.7) and then separated by 0.1 km grids. The coastline distance between populations ( $D_{cst}$ ) was calculated as the sum of the diagonal lengths of all grids. NOAA high-resolution SST data (Reynolds et al., 2007) provided by the NOAA/OAR/ESRL PSL (<https://psl.noaa.gov/>) were obtained for isolation by environment (IBE) analysis. We used an in-house Python script to parse the NetCDF4 file under the support of the “netCDF4” Python library (version 1.5.5.1). For a single site, SSTs were averaged in a 0.25-degree grid covering the site and each date from 1971 to 2010. Genetic distances were measured by mean pairwise *Fst* values of all loci calculated by VCFtools (Danecek et al., 2011) (version 0.1.15). Linear regression between genetic distances and exogenous variables was implemented with the “statsmodels” Python library (version 0.12.1).

### Selective signature detection

We scanned selective signatures in the whole *L. maculatus* genome to compare the two most divergent clusters, the Bohai Gulf populations (BG) and the Beibu Gulf populations (BH). A sliding-window approach was applied in this analysis with a window size of 50 kbp and a sliding step size of 10 kbp. To reduce false positive signals, we calculated three different positive selection statistics for each window and took

their intersection as candidate selected regions. The fixation index ( $F_{st}$ ) and raw nucleotide diversity ( $\pi$ ) were calculated with VCFtools (version 0.1.15) (Danecek et al., 2011). Raw values of  $\pi$  were subsequently calculated as the logarithm rate of BH and BG populations (denoted as  $\omega$ ). The extent of haplotype homozygosity ( $R_{sb}$ ) was calculated for each SNP using the R package “rehh” (version 3.0.1) (Gautier, Klassmann, & Vitalis, 2017) and then averaged in each window. Windows that ranked in the top five percent by absolute values of all three statistics were selected as positively selected regions (PSRs). Positively selected genes were identified by the presence of overlap between genes and PSRs. In addition, we calculated *Tajima's D* index using plink (version 1.90b6.16) (Purcell et al., 2007) within all 50 kb sliding windows and all generic regions. Finally, we calculated the composite selection score (*CSS*) (Avalos et al., 2017) to compare and combine different statistics,  $F_{st}$ ,  $\omega$  and  $R_{sb}$ . For each statistic, we ranked all windows and subsequently performed Z-transformation on their fractional ranks. Then, three Z scores were averaged for each window and compared against a normal distribution of Z scores to derive a P-value. Metascape (version 3.5) (Zhou et al., 2019) was used to analyse differential enrichment of Gene Ontology Biological Processes, KEGG pathways and Reactome Gene Sets for PSGs in the BG and BH populations.

### **Prediction of tertiary structure and thermostability of myosin heavy chain (*myhz2*) in *L. maculatus***

First, we applied the dominant allele of each population and each non-synonymous mutation in the *myhz2* coding sequence followed by translation to the protein sequence. Two resulting primary structures were then imported to I-TASSER software (version 5.1) (Yang et al., 2015) to predict their tertiary structures. Finally, we employed the SCooP web server (version 1.0) (Pucci, Kwasigroch, & Rooman, 2017) to estimate the thermodynamic quantities that characterize the folding process of each structure.

### **Conclusion**

In this study, we generated millions of genome-wide SNP loci to explicate the genetic structure and evolutionary history of spotted sea bass populations that are continuously distributed from temperate to tropical zones. With the addition of long-term remote sensing SST data, we determined that extreme climates in both summer and winter have great significance to genetic differentiation. We further performed microevolutionary analysis employing three different algorithms to detect evolutionary selective signatures and genes associated with climatic adaptation. Our investigation indicated that a substantial number of genes related to growth, muscle contraction, visual development and regulation have undergone natural selection, providing a better understanding of the adaptive mechanism and genetic basis of the phenotypic variation in eurythermal fishes in diverse climatic habitats.

### **Acknowledgements**

We acknowledge financial support from the Knowledge Innovation Program of Shenzhen City (No. JCYJ20170818142601870), Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515010731) and Fundamental Research Funds for the Central Universities, Xiamen University (Nos. 20720190108, 20720190099 and 20720190102).

### **Authors' Contributions**

PX conceived and supervised the study. BC and GJ analysed the data. DZ, BC, ZX and CL collected the samples. ZX, SY and BC conducted nucleic acid extraction and sequencing library preparation. BC and PX wrote and revised the manuscript. All authors have read and approved the manuscript.

### **Conflicts of Interest**

The authors declare that they have no conflict of interest. The authors alone are responsible for the content and writing of the article.

### **Data availability**

All genomic re-sequence data of *L. maculatus* can be found on Sequence Read Archive database of National Center for Biotechnology Information (B. H. Chen & Xu, 2021).

The improved reference genome assembly are also available on National Center for Biotechnology Information (B. H. Chen & Xu, 2019).

## References

- Andrade, M., & Blumstein, D. T. (2020). Anti-predator behavior along elevational and latitudinal gradients in dark-eyed juncos. *Current Zoology*, *66* (3), 239-245.
- Angilletta, M. J., & Dunham, A. E. (2003). The temperature-size rule in ectotherms: Simple evolutionary explanations may not be general. *American Naturalist*, *162* (3), 332-342.
- Angilletta, M. J., Niewiarowski, P. H., & Navas, C. A. (2002). The evolution of thermal physiology in ectotherms. *Journal of Thermal Biology*, *27* (4), 249-268.
- Avalos, A., Pan, H. L., Li, C., Acevedo-Gonzalez, J. P., Rendon, G., Fields, C. J., . . . Zhang, G. J. (2017). A soft selective sweep during rapid evolution of gentle behaviour in an Africanized honeybee. *Nature Communications*, *8* .
- Beitinger, T. L., Bennett, W. A., & McCauley, R. W. (2000). Temperature tolerances of North American freshwater fishes exposed to dynamic changes in temperature. *Environmental Biology of Fishes*, *58* (3), 237-275.
- Billerbeck, J. M., Lankford, T. E., & Conover, D. O. (2001). Evolution of intrinsic growth and energy acquisition rates. I. Trade-offs with swimming performance in *Menidia menidia*. *Evolution*, *55* (9), 1863-1872.
- Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., . . . Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B-Biological Sciences*, *277* (1701), 3725-3734. doi:10.1098/rspb.2010.0985
- Camarillo, H., Rodriguez, L. A., & Tobler, M. (2020). Functional consequences of phenotypic variation between locally adapted populations: Swimming performance and ventilation in extremophile fish. *Journal of Evolutionary Biology*, *33* (4), 512-523.
- Chen, B. H., Li, Y., Peng, W. Z., Zhou, Z. X., Shi, Y., Pu, F., . . . Xu, P. (2019). Chromosome-Level Assembly of the Chinese Seabass (*Lateolabrax maculatus*) Genome. *Frontiers in Genetics*, *10* . doi:ARTN 275  
10.3389/fgene.2019.00275
- [dataset] Chen, B. H., & Xu, P. (2019). *Chromosome-level assembly of spotted sea bass* . NCBI Assembly Database. Retrieved from: [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_004028665.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_004028665.1)
- [dataset] Chen, B. H., & Xu, P. (2021). *Whole genome re-sequencing of spotted sea bass* . Sequence Read Archives. Retrieved from: <https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP305917>
- Chen, D.-g., Gao, T.-x., Zeng, X.-q., Ren, Y.-p., & Ruan, S.-h. (2001). Study on the fishery biology of Laizhou population of *Lateolabrax sp.* *Acta Oceanologica Sinica*, *23* (4), 81-86.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly*, *6* (2), 80-92.
- Conover, D. O. (1990). The Relation between Capacity for Growth and Length of Growing Season: Evidence for and Implications of Countergradient Variation. *Transactions of the American Fisheries Society*, *119* (3), 416-430. doi:10.1577/1548-8659(1990)119<0416:Trbcfg>2.3.Co;2

- Conover, D. O., Brown, J. J., & Ehtisham, A. (1997). Countergradient variation in growth of young striped bass (*Morone saxatilis*) from different latitudes. *Canadian Journal of Fisheries and Aquatic Sciences*, *54* (10), 2401-2409. doi:DOI 10.1139/cjfas-54-10-2401
- Conover, D. O., & Present, T. M. (1990). Countergradient variation in growth rate: compensation for length of the growing season among Atlantic silversides from different latitudes. *Oecologia*, *83* (3), 316-324. doi:10.1007/BF00317554
- Conover, D. O., & Schultz, E. T. (1995). Phenotypic Similarity and the Evolutionary Significance of Countergradient Variation. *Trends in Ecology & Evolution*, *10* (6), 248-252.
- Cook, D. E., & Andersen, E. C. (2017). VCF-kit: assorted utilities for the variant call format. *Bioinformatics*, *33* (10), 1581-1582. doi:10.1093/bioinformatics/btx011
- Culumber, Z. W. (2020). Behavioural response to simulated avian predation varies with latitude and predation intensity of natural populations. *Evolutionary Ecology*, *34* (6), 1037-1046.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Genomes Project Analysis, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27* (15), 2156-2158. doi:10.1093/bioinformatics/btr330
- Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B., & Wood, S. (1998). Making mistakes when predicting shifts in species range in response to global warming. *Nature*, *391* (6669), 783-786. doi:Doi 10.1038/35842
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., . . . Aiden, E. L. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, *356* (6333), 92-95. doi:10.1126/science.aal3327
- Durand, N. C., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., & Aiden, E. L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems*, *3* (1), 99-101. doi:10.1016/j.cels.2015.07.012
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3* (1), 95-98. doi:10.1016/j.cels.2016.07.002
- Gao, T. X., Dong, L., Liu, J. X., Zhang, Y. P., & Koji, Y. (2004). Genetic variation among sea bass populations. *Acta Hydrobiologica Sinica*, *28* (1), 69-73.
- Gautier, M., Klassmann, A., & Vitalis, R. (2017). REHH 2.0: a reimplement of the R package REHH to detect positive selection from haplotype structure. *Molecular Ecology Resources*, *17* (1), 78-90.
- Gong, J., Chen, B. H., Li, B. J., Zhou, Z. X., Shi, Y., Ke, Q. Z., . . . Xu, P. (2020). Genetic analysis of whole mitochondrial genome of *Lateolabrax maculatus* (Perciformes: Moronidae) indicates the presence of two populations along the Chinese coast. *Zoologia*, *37* .
- Greenhalgh, C. J., Bertolino, P., Asa, S. L., Metcalf, D., Corbin, J. E., Adams, T. E., . . . Alexander, W. S. (2002). Growth enhancement in suppressor of cytokine signaling 2 (SOCS-2)-deficient mice is dependent on signal transducer and activator of transcription 5b (STAT5b). *Mol Endocrinol*, *16* (6), 1394-1406. doi:10.1210/mend.16.6.0845
- Hallonet, M., Hollemann, T., Pieler, T., & Gruss, P. (1999). Vax1, a novel homeobox-containing gene, directs development of the basal forebrain and visual system. *Genes & Development*, *13* (23), 3106-3114.
- Hasselman, D. J., Ricard, D., & Bentzen, P. (2013). Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. *Molecular Ecology*, *22* (6), 1558-1573. doi:10.1111/mec.12197

- Henderson, C. A., Gomez, C. G., Novak, S. M., Mi-Mi, L., & Gregorio, C. C. (2017). Overview of the Muscle Cytoskeleton. *Comprehensive Physiology*, 7 (3), 891-944.
- Horvat, S., & Medrano, J. F. (2001). Lack of Socs2 expression causes the high-growth phenotype in mice. *Genomics*, 72 (2), 209-212. doi:10.1006/geno.2000.6441
- Howell, B. K., Matthews, A. D., & Donnelly, A. P. (1991). Thermal-Stability of Fish Myofibrils - a Differential Scanning Calorimetric Study. *International Journal of Food Science and Technology*, 26 (3), 283-295.
- Hu, Y. B., Li, Y., Wen, H. S., Sun, Y. L., Xu, Y. T., Wang, X., . . . Fang, X. (2019). Characterization of thermal tolerance of juvenile *Lateolabrax maculatus* from different communities. *Periodical of ocean university of China*, 49 (S2), 1-7.
- Hu, Z. M., Gao, T. X., Han, Z. Q., & Song, L. (2007). Studies on Genetic Differentiation of the Spotted Sea Bass ( *Lateolabrax maculatus* ) and Japanese Sea Bass ( *Lateolabrax japonicus* ). *Periodical of ocean university of China*, 37 (3), 413-418.
- Hurst, T. P., & Conover, D. O. (1998). Winter mortality of young-of-the-year Hudson River striped bass (*Morone saxatilis*): size-dependent patterns and effects on recruitment. *Canadian Journal of Fisheries and Aquatic Sciences*, 55 (5), 1122-1130.
- Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., . . . Harper, J. W. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, 545 (7655), 505-+.
- Imsland, A. K., Foss, A., Naevdal, G., Cross, T., Bonga, S. W., Ham, E. V., & Stefansson, S. O. (2000). Countergradient variation in growth and food conversion efficiency of juvenile turbot. *Journal of Fish Biology*, 57 (5), 1213-1226. doi:10.1006/jfbi.2000.1384
- Jiang, X., Yang, G.-P., Wei, Q.-W., Zou, G.-W., Wang, D.-Q., Luo, X.-Z., . . . Chen, D.-Q. (2009). Analysis of the genetic structure of spotted sea bass (*Lateolabrax maculatus*) inhabiting the Chinese coast. *Periodical of ocean university of China*, 39 (2), 271-274.
- Johansen, J., & Jones, G. (2011). Increasing ocean temperature reduces the metabolic performance and swimming ability of coral reef damselfishes. *Global Change Biology*, 17 (9), 2971-2979.
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7 (12), 1225-1241. doi:10.1111/j.1461-0248.2004.00684.x
- Kim, G. Y., Kim, H. Y., Kim, H. T., Moon, J. M., Kim, C. H., Kang, S., & Rhim, H. (2012). HtrA1 Is a Novel Antagonist Controlling Fibroblast Growth Factor (FGF) Signaling via Cleavage of FGF8. *Molecular and Cellular Biology*, 32 (21), 4482-4492.
- Krock, B. L., & Perkins, B. D. (2014). The Par-PrkC Polarity Complex Is Required for Cilia Growth in Zebrafish Photoreceptors. *Plos One*, 9 (8).
- Lannig, G., Eckerle, L. G., Serendero, I., Sartoris, F. J., Fischer, T., Knust, R., . . . Portner, H. O. (2003). Temperature adaptation in eurythermal cod (*Gadus morhua*): a comparison of mitochondrial enzyme capacities in boreal and Arctic populations. *Marine Biology*, 142 (3), 589-599. doi:10.1007/s00227-002-0967-6
- Leimu, R., & Fischer, M. (2008). A Meta-Analysis of Local Adaptation in Plants. *Plos One*, 3 (12). doi:ARTN e4010
- 10.1371/journal.pone.0004010
- Levins, R. (1969). Thermal Acclimation and Heat Resistance in *Drosophila* Species. *The American Naturalist*, 103 (933), 483-499. doi:10.1086/282616

- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25 (14), 1754-1760. doi:10.1093/bioinformatics/btp324
- Liao, W. H., Cheng, C. H., Hung, K. S., Chiu, W. T., Chen, G. D., Hwang, P. P., . . . Huang, C. J. (2013). Protein tyrosine phosphatase receptor type O (Ptpro) regulates cerebellar formation during zebrafish development through modulating Fgf signaling. *Cellular and Molecular Life Sciences*, 70 (13), 2367-2381.
- Lindgren, B., & Laurila, A. (2005). Proximate causes of adaptive growth rates: growth efficiency variation among latitudinal populations of *Rana temporaria*. *Journal of Evolutionary Biology*, 18 (4), 820-828. doi:10.1111/j.1420-9101.2004.00875.x
- Liu, J. X., Gao, T. X., Yokogawa, K., & Zhang, Y. P. (2006). Differential population structuring and demographic history of two closely related fish species, Japanese sea bass (*Lateolabrax japonicus*) and spotted sea bass (*Lateolabrax maculatus*) in Northwestern Pacific. *Molecular Phylogenetics and Evolution*, 39 (3), 799-811. doi:10.1016/j.ympev.2006.01.009
- Liu, Y., Wang, H., Wen, H., Shi, Y., Zhang, M., Qi, X., . . . Li, Y. (2020). First High-Density Linkage Map and QTL Fine Mapping for Growth-Related Traits of Spotted Sea bass (*Lateolabrax maculatus*). *Mar Biotechnol (NY)*, 22 (4), 526-538. doi:10.1007/s10126-020-09973-4
- Marcais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *Plos Computational Biology*, 14 (1). doi:ARTN e1005944 10.1371/journal.pcbi.1005944
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20 (9), 1297-1303. doi:10.1101/gr.107524.110
- Meirmans, P. G. (2012). The trouble with isolation by distance. *Molecular Ecology*, 21 (12), 2839-2846.
- Mui, S. H., Kim, J. W., Lemke, G., & Bertuzzi, S. (2005). Vax genes ventralize the embryonic eye. *Genes & Development*, 19 (10), 1249-1259.
- Pankhurst, N. W., & Munday, P. L. (2011). Effects of climate change on fish reproduction and early life history stages. *Marine and Freshwater Research*, 62 (9), 1015-1026.
- Pucci, F., Kwasigroch, J. M., & Rooman, M. (2017). SCooP: an accurate and fast predictor of protein stability curves as a function of temperature. *Bioinformatics*, 33 (21), 3415-3422.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., . . . Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81 (3), 559-575.
- Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, 20 (22), 5473-5496.
- Rodgers, M. E., Karr, T., Biedermann, K., Ueno, H., & Harrington, W. F. (1987). Thermal-Stability of Myosin Rod from Various Species. *Biochemistry*, 26 (26), 8703-8708.
- Rountrey, A. N., Coulson, P. G., Meeuwig, J. J., & Meekan, M. (2014). Water temperature and fish growth: otoliths predict growth patterns of a marine fish in a changing climate. *Glob Chang Biol*, 20 (8), 2450-2458. doi:10.1111/gcb.12617
- Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M., & Roy, K. (2009). Is There a Latitudinal Gradient in the Importance of Biotic Interactions? *Annual Review of Ecology Evolution and Systematics*, 40 , 245-269.

- Shoji, J., & Tanaka, M. (2007). Growth and mortality of larval and juvenile Japanese seaperch *Lateolabrax japonicus* in relation to seasonal changes in temperature and prey abundance in the Chikugo estuary. *Estuarine, Coastal and Shelf Science*, 73 (3-4), 423-430.
- Sun, P., Li, X., Gong, X., Liu, Y., Zhang, X., & Wang, L. (2014). Carbon, nitrogen and phosphorus ecological stoichiometry of *Lateolabrax maculatus* and *Acanthogobius ommaturus* in the Estuary of Yangtze River, China. *Acta Ecologica Sinica*, 34 (4), 196-203.
- Suzuki, Y., Miyake, T., & Yamahira, K. (2010). An acquisition trade-off with fast growth in a fish, the medaka *Oryzias latipes*: why do low-latitude ectotherms grow more slowly? *Evolutionary Ecology*, 24 (4), 749-759.
- Tang, H., Peng, J., Wang, P., & Risch, N. J. (2005). Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol*, 28 (4), 289-301. doi:10.1002/gepi.20064
- Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49 (2), 303-309. doi:10.1038/ng.3748
- Thurrow, R. F., Peterson, J. T., Chandler, G. L., Moffitt, C. M., & Bjornn, T. C. (2020). Concealment of juvenile bull trout in response to temperature, light, and substrate: Implications for detection. *Plos One*, 15 (9), e0237716. doi:10.1371/journal.pone.0237716
- Uzendoski, K., Maksymovitch, E., & Verrell, P. (1993). Do the Risks of Predation and Intermale Competition Affect Courtship Behavior in the Salamander *Desmognathus Ochrophaeus*. *Behavioral Ecology and Sociobiology*, 32 (6), 421-427.
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., . . . DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*, 43 , 11 10 11-33. doi:10.1002/0471250953.bi1110s43
- Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., . . . Orban, L. (2016). Chromosomal-Level Assembly of the Asian Seabass Genome Using Long Sequence Reads and Multi-layered Scaffolding. *Plos Genetics*, 12 (4), e1005954. doi:10.1371/journal.pgen.1005954
- Wang, I. J. (2012). Environmental and topographic variables shape genetic structure and effective population sizes in the endangered Yosemite toad. *Diversity and Distributions*, 18 (10), 1033-1041.
- Wei, X. Y., Cheng, Y., Luo, Y. Y., Shi, X. H., Nelson, S., & Hyde, D. R. (2004). The zebrafish *Pard3* ortholog is required for separation of the eye fields and retinal lamination. *Developmental Biology*, 269 (1), 286-301.
- Yang, J. Y., Yan, R. X., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12 (1), 7-8.
- Yokogawa, K. (1998). Genetic divergence of fishes in genus *Lateolabrax* (Perciformes: Percichthyidae). *Aquaculture Science*, 46 (3), 315-320.
- Zakharov, E. V., & Hellmann, J. J. (2008). Genetic differentiation across a latitudinal gradient in two co-occurring butterfly species: revealing population differences in a context of climate change (vol 17, pg 189, 2008). *Molecular Ecology*, 17 (12), 3035-3035. doi:10.1111/j.1365-294X.2008.03825.x
- Zhang, X., Zhang, S., Zhao, Q., Ming, R., & Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants*, 5 (8), 833-845. doi:10.1038/s41477-019-0487-8
- Zhao, Y., Peng, W., Guo, H., Chen, B., Zhou, Z., Xu, J., . . . Xu, P. (2018). Population genomics reveals genetic divergence and adaptive differentiation of Chinese sea bass (*Lateolabrax maculatus*). *Marine Biotechnology*, 20 (1), 45-59.



Zhaoshu Liu, H. Z., Shiqing Fan, Senqiang Chen. (2002). *Geology of the South China Sea* (H. Xie Ed. 1 ed.). Beijing: Science Press.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28 (24), 3326-3328. doi:10.1093/bioinformatics/bts606

Zhou, Y. Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., . . . Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, 10 .

## Figure Legends

**Figure 1. Population genomics and phylogenetic analysis indicating the genetic structure and evolutionary relationship of *L. maculatus* populations.** **A**, Sampling sites along the Chinese coastline. Offshore waters are coloured by the median annual SST. **B**, PCA showing genetic distance among samples. The genetic diversity of each cluster is shown by confidence ellipses (BH, BH and IM). **C**, Phylogenetic analysis was demonstrated by a rooted neighbour-joining tree. **D**, Ancestral admixture analysis revealed the population genetic structure. **E**, Demographic analysis indicated the historical effective population size of *L. maculatus* populations.

**Figure 2. Genetics isolation by distance and environment.** **A**, Ordinary least squares (OLS) regression was used to model the isolation-by-distance pattern of *L. maculatus* populations. **B**, OLS regression showed that the genetic distance was also highly linearly correlated with differences in the annual maximum SST. **C**, Correlation heatmap of geographic and environmental factors. **D**, Linear regression attempts to model the relationship between genetic distance and three exogenous factors: coastline distance, difference in annual minimum SST, and difference in annual maximum SST.

**Figure 3. Genome-wide selective signatures in the *L. maculatus* genome.** **A**, Probability distributions of *Fst*,  $\omega$  and *Rsb* values over all sliding windows. Candidate selected regions were coloured by their *Fst* values. **B**, Circus plot showing the distributions of *CSS* (green), *Fst* (blue),  $\omega$  (khaki) and *Rsb* (red) values along all chromosomes. **C**, An upset plot showing intersections among the results of three selective signature detection algorithms. Statistical distribution of *CSS*, *Fst*,  $\omega$ , *Pof*, and *Tajima's D* in BB populations and *Tajima's D* in BH populations of each intersection.

**Figure 4. Selective signatures of GH-IGF axis-related genes in BH populations.** **A**, Enriched function categories based on BH PSGs. **B**, *Tajima's D* values in BH and BB populations of genes classified in enriched categories and other categories. **C**, Representative selective signatures of related genes on the GH-IGF axis. **E**, Allele frequency of two non-synonymous substitutions in the *ibp3* CDS (c.320C>G).

**Figure 5. Selective signatures in fibre muscle components.** **A**, Selective signature on *myhz2* gene. **B**, Allele frequency of two non-synonymous substitutions in the coding region of *myhz2* (c.373C>A, left; c700G>A, right). **C**, 3D protein structure of two types of myosin heavy chain protein. **D**, Gibbs free energy curve showing the thermal stability change introduced by these substitutions.

**Figure 6. Selective signatures of visual sense-related genes in BB populations.** **A**, Enriched function categories based on BB PSGs. **B**, *Tajima's D* in BH and BB populations of genes classified in enriched categories and other categories. **C**, Selective signatures of genes related to the visual system. **D**, Allele frequency of a non-synonymous substitution in the *pard3ab* CDS (c.37A>G) in all populations.

## Tables

**Table 1 .** Linear regression showing the variance *Fst* explained by differences of geological and environmental factors between populations.

Simple OLS <sup>a</sup>	R <sup>2</sup>	P-value of model <sup>c</sup>	AIC	BIC	P-values of each factor <sup>bc</sup>
D <sub>cst</sub>	<b>0.506</b>	2.91E-13	-612.91	-608.20	2.91E-13

Simple OLS <sup>a</sup>	R <sup>2</sup>	P-value of model <sup>c</sup>	AIC	BIC	P-values of each factor <sup>bc</sup>
D <sub>gcc</sub>	0.459	9.74E-12	-605.80	-601.09	9.74E-12
SST <sub>max</sub>	<b>0.548</b>	9.76E-15	-619.80	-615.09	9.76E-15
SST <sub>med</sub>	0.269	1.13E-06	-582.37	-577.66	1.13E-06
SST <sub>min</sub>	0.204	3.26E-05	-575.73	-571.02	3.26E-05
Multi-factor OLS	R <sup>2</sup>	P-value of model <sup>c</sup>	AIC	BIC	P-values of each factor <sup>c</sup>
D <sub>cst</sub> + D <sub>gcc</sub>	0.52	1.10E-12	-613.176	-606.106	0.003;0.141
D <sub>cst</sub> + SST <sub>max</sub>	0.575	1.14E-14	-622.69	-615.62	0.031;7.97E-04
D <sub>cst</sub> + SST <sub>med</sub>	0.672	7.05E-19	-642.84	-635.77	1.11E-14;3.34E-08
D <sub>cst</sub> + SST <sub>min</sub>	0.698	3.01E-20	-649.39	-642.32	1.84E-17;1.33E-09
D <sub>cst</sub> +SST <sub>min</sub> +SST <sub>max</sub>	0.757	1.13E-22	-664.23	-654.80	9.57E-11;1.49E-10;6.81E-05

<sup>a</sup> OLS: abbreviation of “ordinary least squares”.

<sup>b</sup> In simple ordinary least squares models, column “P-values of each factor” have only one value equals to P-value of model.

<sup>c</sup> Values less than 0.001 and greater than -0.001 were expressed in scientific notation.







