

# Revisiting the ‘Marine Speciation Paradox’ using genomic, morphometric, and ecological evidence in pipefish

Xin Wang<sup>1</sup>, Zhixin Zhang<sup>2</sup>, Stefano Mammola<sup>3</sup>, Adeljean Ho<sup>4</sup>, Yanhong Zhang<sup>5</sup>, Geng Qin<sup>5</sup>, and Qiang Lin<sup>6</sup>

<sup>1</sup>CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, South China Sea Institute of Oceanology, Institute of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences

<sup>2</sup>Tokyo University of Marine Science and Technology Graduate School of Marine Science and Technology

<sup>3</sup>Water Research Institute, National Research Council of Italy

<sup>4</sup>Florida Institute of Technology

<sup>5</sup>Affiliation not available

<sup>6</sup>South China Sea Institute of Oceanology

June 3, 2020

## Abstract

General theory predicts that ecological specialization should be rare in marine ecosystems, given that barriers to dispersal are less effective in the vastness of the sea compared to terrestrial settings. This paradigm, however, hardly fits with classical theories of local adaptation, leaving the question open of as to how marine diversity could originate at a restricted spatial scale. We tackled this so-called “Marine Speciation Paradox” by investigating how local specialization could arise in a widely distributed marine species, the seaweed pipefish *Syngnathus schlegeli*. We integrated morphological, genomic, and niche-based evidences to unravel geographical structuring in *S. schlegeli* populations. We revealed the existence of a north-to-south phenotypic gradient in eye size among *S. schlegeli* populations. This morphological differentiation was paralleled by genetic divergence, with South China Sea populations emerging as relatively independent. The north-to-south phylogeographical structuring was further corroborated by ecological analyses. We observed high niche differentiation among northern, central, and southern populations, resulting from both niche expansion and niche shift processes. Projected habitat suitability onto the Last Glacial Maximum revealed the existence of historical barriers to dispersal between the South and East China seas. We showed that the effect of this historical segregation, in concert with niche-driven ecological differentiation, lead to establishment of three distinct clades across the widely distributed marine pipefish. Ultimately, our study demonstrates that even the sea environment maintains the potential for adaptive radiation and ecological specialization, suggesting that ‘marine speciation’ may actually be far from being ‘paradoxical’.

## Introduction

Ecological specialization, the spatial and temporal differentiation of species in the use of habitat and resources (Devictor *et al.*2010), is a pivotal eco-evolutionary paradigm to explain phylogeographical patterns, community assembly rules, and the majority of adaptive radiations (Gillespie *et al.* 2020; Mammola *et al.* 2020; Todd Streelman & Danley 2003; Wilson *et al.*2020). Despite the fact that marine fishes often show complex adaptations in spatially heterogeneous environments (e.g. Wainwright *et al.*2004; Wainwright *et al.*2015), ecological specialization is thought to be rare in marine species mainly because the geographic

barriers in the vastness of seas are weak and planktonic larvae have the potential to maintain large population sizes with high gene flow (Palumbi 1994; Zhang *et al.* 2016). Therefore, marine organisms challenge the classical theories of local adaptation and speciation, what has been dubbed the “Marine Speciation Paradox” (Bierne *et al.* 2003).

The importance of natural selection in shaping adaptive trait differentiation among natural populations has long been recognized (Garant *et al.* 2005; Smith *et al.* 1997). Theory predicts that once populations are exposed to diverse ecological environments, rapid evolution of adaptive traits should occur (Hendry *et al.* 2000). Although empirical examples of deep divergence in marine ecosystems are limited, molecular phylogenetic reconstructions of some marine fishes suggest that ecological specialization has a great impact on their diversification (e.g. Rocha *et al.* 2005; Wainwright *et al.* 2004). This is especially true for geographically widespread species that may experience considerably different ecological conditions across their distribution (Fox & Morrow 1981; Loxdale *et al.* 2011; Shipley *et al.* 2009). Divergent selection is an important element of natural selection, and occurs when different environments favor different phenotypes, leading to increased differences between populations (Bolnick & Stutz 2017). Theoretically, divergent selection should promote the evolution of traits in local populations that provide an advantage under local environmental conditions (Kawecki & Ebert 2004), and selection pressure may lead to genetic divergence and eventual speciation if the homogenizing effects of gene flow are insufficient to prevent it. In this regard, uncovering the differentiation among marine populations is of considerable importance for advancing our knowledge of the molecular mechanisms of adaptive evolution against a background of high-level gene flow.

The seaweed pipefish *Syngnathus schlegeli* (Kaup, 1856) is a well-known species in which there is exclusive paternal care of eggs (Watanabe & Watanabe 2001). The offsprings develop within the male brood pouch and are released as free-living juveniles. *Syngnathus schlegeli* is distributed along coastlines of the northwestern Pacific Ocean and generally inhabits sheltered areas in shallow waters; geographic range of this species encompasses a wide variety of climatic conditions and nearshore microhabitats (Chen *et al.* 2017; Zhang *et al.* 2017). The vast region of the northwestern Pacific is characterized by distinct tectonic and geographical features, with a series of marginal seas separating the eastern Asian continent from the Pacific Ocean (Tamaki & Honza 1991). During the Pleistocene glacial cycles, the sea level fluctuations resulted in successive exposure and inundation of continental shelves, closure and opening of seaways, and separation and reunion of marginal seas (Wang 1999). When sea level falls, the seas were separated by land bridges extending between islands and the mainland, hindering the spread of most marine species (Ni *et al.* 2014; Wang 1999). The historical influence of glaciation, which has been proposed as a dominant factor shaping present-day phylogeographical patterns (Hewitt 2004), distinguishes this area from other marine realms to such an extent that the glacial effects on its biota would be distinctly different from those revealed in other regions (Xu *et al.* 2009).

In this study, we investigated the cranial morphology, genetic structure, and ecological niche of *S. schlegeli* along China seashores to reveal the phylogeographical patterns resulting from adaptive radiation and ecological specialization. Firstly, we used geometric morphometrics to partition the variations in the cranial morphology among populations. Secondly, we employed the restriction-site-associated DNA sequencing (RADseq) technique to evaluate the level of genetic differentiation among seaweed pipefish populations based on the genome-wide genotypic and sequence data. Finally, we compared the ecological niches of independent *S. schlegeli* populations using  $n$ -dimensional hypervolumes, and mapped habitat suitability of this species during present-day and the Last Glacial Maximum (LGM) using species distribution model (SDM). By revisiting the phylogeographical history of the widespread pipefish using an integrative approach, our over-arching goal is to investigate the dynamics through which local specialization could arise in the apparent absence of dispersal barriers.

## Materials and methods

### Sample collection and DNA extraction

A total of 280 adult individuals of seaweed pipefish were collected from 9 sites along the coastal waters of China ranging from the Yellow Sea (YS) to the South China Sea (SCS) (Table 1, Figure S1). We derived

morphological data from 163 seaweed pipefish specimens: 63 from the YS, 34 from the East China Sea (ECS), and 66 from the SCS. The body of the remaining 117 specimens was damaged, preventing to obtain reliable morphological measures. The total length (TL) of sampled specimens of *S. schlegeli* ranged 18.1–21.6 cm, and no significant differences in fish sizes existed among sampled localities (ANOVA, Duncan’s multiple range tests,  $p > 0.05$ ). Genomic DNA was extracted from muscle samples and then treated with RNase A to digest any RNAs present in the samples. The quality and integrity of the DNA samples were checked by agarose gel electrophoresis, and DNA concentrations were quantified using a Qubit 2.0 fluorometer.

We performed RAD-sequencing for 69 individuals (31 from the Yellow Sea, 22 from the East China Sea, and 16 from the South China Sea) (Table 1, Figure S1). The mitochondrial Cytochrome b (*cytb*) and nuclear Sorting nexin 33 (*snx33*) amplified from 270 pipefish individuals, and we obtained 248 and 265 valid *sh3* and *cytb* sequences after data correction (Table 1). All experimental procedures performed in this study were in accordance with relevant animal ethics guidelines and approved by the Ethical Committee of South China Sea Institute of Oceanology, Chinese Academy of Sciences.

### Sequence quality determination and filtering of RADtags

The DNA from individual fish was digested with *Eco* RI and paired-end sequencing libraries with an insert size of 200–400 bp were constructed. These RAD libraries were subsequently sequenced using the Illumina HiSeq PE150 platform.

The raw sequence data were filtered by removal of reads with low-quality, reads with adapter contamination, reads with [?] 10% unidentified nucleotides (N), and reads for which [?] 50% of the length had a phred quality [?] 5.

### SNP calling

All quality-filtered reads were mapped to the *Syngnathus scovelli* reference genome using BWA-MEM with the parameters `aln -e 10 -t 4 -l 32 -i 15 -q 10` (Li & Durbin 2009). The alignment files were imported to SAMtools for sorting and removal of duplicated reads (Li 2011). Sequencing coverage and depth for each sample were calculated. SNPs were detected using SAMtools and then filtered to meet the following criteria: (1) quality score [?] 20, (2) coverage depth [?] 5 and [?] 1000, (3) SNPs with minor allele frequency (MAF)  $> 0.01$ , and (4) a missing ratio of samples within each group  $< 50\%$ .

### Population genetics analyses

We constructed a neighbor-joining tree using the pairwise genetic distance matrix data of all individuals calculated using TreeBest v1.9.2 (Vilella *et al.* 2009), and the number of bootstrap replicates was set to 1000 to assess the statistical support for nodes in the tree.

Principal components analysis (PCA) based on SNP markers was performed using GCTA tools (Yang *et al.* 2011). We transformed the population genotypes into a matrix that included the numbers 0, 1, and 2, where 0 was used to represent a genotype that is homozygous for the reference allele, 1 is a genotype heterozygous for the reference allele; and 2 is a genotype homozygous for the non-reference allele. We calculated the sample covariance of the matrix that contained the information for all individuals (with the numbers 0, 1, and 2). Finally, we calculated the eigenvector decomposition of the matrix and plotted the PCA using GCTA tools.

We used PLINK to generate the map files necessary for downstream analyses (Purcell *et al.* 2007). Population structure was determined using the program FRAPPE (Tang *et al.* 2005). We increased the pre-defined parameter K from 2 to 5 to cover the maximum numbers of lineages that could be identified in the tree, representing the assumed groups of a simulated population in ancient times. Population genetic structure and individual ancestry proportions were finally inferred using FRAPPE.

### SWEEP analysis

To characterize genome-wide patterns of genetic variation and population differentiation, the nucleotide diversity ( $\pi$ ) and population-differentiation statistic ( $F_{ST}$ ) were calculated using a sliding window approach

(a 40-kb window sliding in 20-kb steps). Population pairwise  $F_{ST}$  based on the mitochondrial *cytb* and nuclear *snx33* gene were calculated using Arlequin software version 3.5.2.2 (Excofer et al. 2005).  $\pi$  was used to estimate the genomic diversity of populations and  $F_{ST}$  was used to evaluate the strength of genomic differentiation between two populations.

The selection signatures of population-specific genomic regions were measured between *S. schlegeli* populations (YS-ECS and SCS). We applied a sliding window approach (a 40kb window sliding with a step size of 20kb) to identify selected regions. Genomic regions with a significantly high  $F_{ST}$  statistic (corresponding to the top 5% level) and  $\theta\pi$  ratio (the  $\theta\pi$  ratio was calculated as  $\log_2(\pi_1) - \log_2(\pi_2)$ , and the top 5% level of  $\theta\pi$  ratio was selected) values were identified as selective sweep regions. Genes were submitted to the Gene Ontology databases for annotation (Ashburner *et al.* 2000; Ogata *et al.* 2000). Gene ontology analyses were performed to annotate the genes to functional ontology terms. Using these candidate genes, we performed a targeted population genetic analysis. We used a higher resolution sliding window analysis in a 10-kb window (compared with a 100-kb window in selection analysis) to calculate  $F_{ST}$  and the  $\theta\pi$  between populations.

### Comparative morphometric analysis of populations

We examined the cranial morphology of 163 individuals (included the 69 individuals involved in the RADseq) of *S. schlegeli* using landmark-based geometric morphometry analysis. Digital images of the right lateral side of each specimen were taken using a Canon camera (EOS Rebel T4i; 18 mega pixels). All the morphometric analyses were performed using the comprehensive programs of Thin-plate spline (tps) (available at <http://life.bio.sunysb.edu/morph/>). Digital images were sorted according to geographic population and converted to tps files using tpsUtil. On the head images, we digitized 26 landmarks using tpsDig2, and the tps files with landmarks were then processed in tpsRelw to enable relative warp analysis and to obtained X and Y coordinates for further analysis. ANOVAs were performed on the RW scores of all *S. schlegeli* populations.

### Ecological niche analyses

We collected 149 presence records of *S. schlegeli* from the Global Biodiversity Information Facility (GBIF, 2019), field investigations, and published literature (Table S1). To avoid clustered records, we used only one presence record per 5 arcminute grid cell, approx. 10 by 10 km, to match the spatial resolution of marine environmental predictors. As a result, we retained 131 records for subsequent analyses (Figure S2). Presence records of *S. schlegeli* along the coastal waters of mainland China were assigned to three distinct clades (YS, ECS, and SCS) based on results of genetic clustering and previous phylogeographic evidence from related taxa in the same region (e.g. Dong et al. 2012; Liu et al. 2007; Ni et al. 2012; Xu et al. 2009).

Marine environmental predictors at a spatial resolution of 5 arcmin were retrieved from the MARSPEC database (<http://marspec.weebly.com>) (Sbrocco & Barber 2013). We initially considered bathymetry (i.e. depth of the seafloor), and 10 marine environmental predictors, including annual mean, range, variance, and extreme values of sea surface temperature (SST) and sea surface salinity (SSS) (Figure S3). Bathymetry was considered since seaweed pipefish prefers shallow waters (Figure S1) (Chen *et al.* 2017; Zhang *et al.* 2017). We calculated the pairwise Pearson's correlation coefficients ( $r$ ) among predictors and selected only one among highly correlated predictors (i.e.  $|r| > 0.7$ ) (Dormann et al. 2013). Accordingly, we retained five predictors for model development: bathymetry, mean annual SSS (bio8), annual range in SSS (bio11), mean annual SST (bio13), and annual range in SST (bio16) (Figure S3).

We used  $n$ -dimensional hypervolumes (Blonder *et al.* 2014) to explore niche differentiation among the three clades. We delineated hypervolumes using the five predictors selected for SDM and a gaussian kernel density estimator (Blonder *et al.* 2018). We studied niche overlap using both a similarity and a distance metric (Mammola 2019), namely the hypervolumes' distance between centroids and pairwise overall differentiation ( $\beta$  diversity) (Carvalho & Cardoso 2018; Mammola & Cardoso 2020). The latter approach decomposes overall niche differentiation ( $\beta_{total}$ ) in two process: niche shifts ( $\beta_{replacement}$ ) and niche expansion/contraction

( $\beta_{\text{richness}}$ ) (Carvalho & Cardoso 2018).  $\beta_{\text{total}}$  values range from 0 (when niches are identical) and 1 (fully dissimilar niches), holding true the equality  $\beta_{\text{total}} = \beta_{\text{replacement}} + \beta_{\text{richness}}$ .

Furthermore, we explored habitat suitability of *S. schlegelii* during present-day and the Last Glacial Maximum (LGM, about 21 kya) using species distribution model (SDM). SDMs represent a widely used tool to quantify a species' habitat suitability by exploring the relationships between distribution data and environmental predictors (Araujo et al., 2019; Thuiller et al. 2019). The MARSPEC database contains paleo marine environmental layers (Sbrocco 2014). For simplicity, Sbrocco (2014) assumed that sea level at the LGM was 120 meters lower than today. Marine environmental predictors at the LGM can be derived from general circulation models (GCMs). We considered paleoclimate simulations from three GCMs (i.e. CCSM3, ECBilt-Clio, and HadCM3M2) to reduce uncertainties derived from choices of GCMs (Thuiller et al. 2019).

We developed SDM using all distribution data (i.e. 131 records) and the five non-collinear predictors. For the purpose of reducing uncertainties associated with choices of SDM algorithms, we used an ensemble modeling approach (Thuiller et al. 2019). We fitted and tested the performance of ten algorithms available in *biomod2* R package (Figure S4), using default settings (Thuiller et al. 2020). In addition to species presence data, we randomly generated 10,000 background points within the study area as pseudo-absences. We evaluated the predictive abilities of the ten algorithms by a five-fold cross-validation approach. True skill statistic (TSS) and the area under the receiver operating characteristic curve (AUC) were used to measure SDM predictive performance; algorithms with TSS over 0.75 and AUC over 0.9 were considered to exhibit excellent predictive capacity (Landis & Koch 1977; Omri Allouche et al. 2006; Swets 1988). All distribution data and selected algorithms were used to predict habitat suitability of *S. schlegelii* under present-day and LGM climate conditions. Predictions of single algorithms were ensembled by a TSS weighted average strategy (Thuiller et al. 2020; Thuiller et al. 2019). The continuous habitat suitability results (ranging continuously from 0 to 1) were transformed into binary suitable/unsuitable maps by using the 10th percentile presence probability threshold.

## Results

### Analysis of morphological variation

The shape of the specimens was described using relative warp analysis (Table S2). The first relative warp (RW1) shows the length of the snout and gill cover, and the second relative warp (RW2) explains the angle of the snout (Table 2). The first three relative warps explained 65.91% of the variance in RWs, and the ANOVA results showed that there are significant differences between populations in terms of RW3 (the third relative warp), which represents eye size (Table 2). Specimens plotted on the positive axis of RW3 tend to have larger eye, whereas specimens plotted on the negative axis have smaller eyes (Figure 1b,c).

The morphological difference in different pipefish populations shown by the RW scatter plot reveals that the YS population generally has a higher RW3, which indicates that the fish in this population have larger eyes. In comparison, individuals in the SCS population have lower RW3 values than those in the other two populations, with ECS population fish having intermediate RW3 values (Figure 1d). We found that seven of the total 26 landmarks were related to morphological variation in eye size, which accounts for 51.4% of the variation in information determined by RW analysis (Figure 1e).

### RAD sequencing and population structure analysis

We performed RAD-sequencing for 69 *S. schlegelii* individuals, obtaining a total of 120.08 Gb raw sequence data, of which 118.13 Gb clean data remained after filtering (Table S3). The sequencing reads were aligned to the reference genome (gulf pipefish *Syngnathus scovelli*) to evaluate the mapping rate, sequencing depth, and coverage. The results shows that the average mapping rate of the sequencing reads was 85.28%, the average sequencing depth was 10 fold, and the average coverage of the data was 20.70% (coverage at least 1×) and 10.82% (coverage at least 4×) (Table S4). We identified a total of 1,607,529 SNPs from the *S. schlegelii* populations, and retained 93,845 high-quality SNP markers after filtering (Table S5).

We reconstructed phylogenetic relationships among the 69 sequenced *S. schlegelii* individuals based on the

SNPs data using the neighbor-joining method. The result showed three distinct clusters that represent three geographical lineages (Figure 2a). Principal components analysis (PCA) recovered the same clusters revealed by the phylogenetic analysis (Figure 2b), revealed that the YS population is closely related to the ECS population, whereas the SCS population formed a separate cluster. Multi-level ( $K = 2, 3, 4$ , and  $5$ ) population structures showed that SCS pipefish comprise a relatively stable and independent group, with a change in  $K$  value from 2 to 5, indicating that the SCS population has low genetic exchange with other populations. In contrast, the YS and ECS populations were well connected with appreciable gene exchange (Figure 2c).

### Screening for selective sweeps

The pairwise  $F_{ST}$  revealed by the *snx 33* gene indicated significant genetic differentiation among different *S. schlegeli* populations (SCS vs. YS,  $F_{ST} = 0.057, p < 0.001$ ; SCS vs. ECS,  $F_{ST} = 0.058, p < 0.05$ ) (Table S6). We calculated the genome-wide nucleotide diversity ( $\pi$ ) and population-differentiation statistic ( $F_{ST}$ ) for each population using a 50kb sliding window with a 10kb step size. Selected genomic regions invariably show specific patterns of variation, such as high population-differentiation statistic ( $F_{ST}$ ) and lower levels of nucleotide diversity ( $\pi$ ) (Ellegren & Sheldon 2008; Sabeti *et al.* 2006; Wang *et al.* 2015). We calculated the distribution of  $\pi$  ratio and  $F_{ST}$  values to identify the selected regions for the north (YS-ECS) and SCS populations, and we inferred the selected genome regions from high  $\pi$  log-ratios and an extreme divergence of allele frequencies (Figure 3a,b). Most of the genome regions shows a low level of differentiation (Figure 3b), and the differentiated genome regions harbored 143 annotated protein-coding genes, including cold shock domain-containing protein E1 (*csde1*), paired box protein 7 (*pax7*), eyes shut homolog gene (*eys*), retinal structural protein (*rx3*), and retinal guanylate cyclase (*retgc-2*), *et al* (Table S7), which we assumed to be subject to selection. The Gene Ontology annotations on these genes were conducted to further understand their regulatory functions (Figure 3c). The selection sweep analysis of the candidate gene *eys* revealed high  $\pi$  value difference and an extreme divergence of allele frequencies. (Figure 3d).

### Results of ecological niche analyses

Hypervolume analysis indicated that all clades diverge in their niche, having adapted to largely distinct environmental conditions across their distribution range (Figure 4). Clade YS had a relatively narrower ecological niche compared with clade ECS and SCS. The greater niche differentiation occurs between clade YS and SCS (Table 3), and this is due to two niche-based operating simultaneously: the most preponderant is a niche shift, with clade SCS experiencing a wider range of depths and climatic conditions, and a parallel niche expansion of SCS (or a contraction of clade YS). The niche of clade ECS, on the other hand, although disjunct, was more similar to that of the other two clades, as emphasize by a small distance between centroids. The differentiation of clade ECS with respect to clades YS and SCS was only due to niche expansion/contraction processes (Table 3).

TSS and AUC scores showed that all algorithms but surface range envelop had excellent performance (Figure S4); thus, these nine algorithms were used to develop the ensemble models. The ensemble SDMs had high predictive capacity with mean ( $\pm$  standard error) TSS = 0.911 ( $\pm$  0.003) and mean AUC = 0.987 ( $\pm$  0.001). Projections of ensemble SDM suggest that a large part of coastal areas of China, Japan, and Korea are suitable for *S. schlegeli*, which generally matches the current known distribution range of this species (Figure 5a, Figure S5a). Habitat suitability projections for this species derived from all three GCMs showed a clear geographical separation between south and north populations during the LGM (Figure 5b-d, Figure S5b-d). In addition, SDM projections suggest that suitable area of *S. schlegeli* was greatly reduced during the LGM: the projected distribution area during the LGM was 10–15% with respect to current depending on the different GCMs.

### Discussion

The integration of genetic, morphological and ecological data, provide support to the hypothesis that even the barrier-free sea environment maintains a vast potential for ecological specialization. We delimited three independent clades of *S. schlegeli* by genomic data, and observed that this differentiation was mirrored in the

morphological divergence of local populations. A similar phylogeographical pattern have been documented in the northwestern Pacific using fishes, molluscs, and crustaceans (Liu *et al.* 2007; Ni *et al.* 2012; Wilson *et al.* 2020; Xu *et al.* 2009).

In marine environments, worldwide glaciation is confirmed to be the most efficient way in generating intraspecific genetic splits (Hewitt 2000), as well as shaping the present-day phylogeographical pattern of marine species (Dong *et al.* 2012; Wilson *et al.* 2020). Example of this have been found in different biogeographic realms, from the Indo-West Pacific to the Atlantic and Mediterranean basins (Cunha *et al.* 2008; Ni *et al.* 2014). Pleistocene sea level fluctuation resulted in the isolation of marginal seas; for instance, sea level during the LGM decreased in the marginal seas of northwestern Pacific (approximately 100–120 m drop in the SCS, and 130–150 m drop in the ECS) (Wang & Sun 1994). Therefore, along with the sea level dropping, the ECS was reduced to an elongated trough, the Okinawa Trough, while the SCS became a semi-enclosed sac-shaped gulf connected with the Pacific mainly through the Bashi Strait (Wang & Sun 1994). These sea basins served as separate refuges for different geographic populations of local marine species. Clearly, the present genetic data suggests a relatively deep genetic split between the ECS-YS and SCS lineages, which has also been reported in other marine fishes (Liu *et al.* 2007). A few genomic regions in *S. schlegeli* showed high differentiation while the rest of the genome was very similar, as indicated by  $F_{ST}$  analysis between North and SCS populations. Such a heterogeneous genome divergence has been commonly described in situations where differentiated genomes have experienced differential introgression following secondary contact (Xu *et al.* 2009). Postglacial range expansions of marine species may bring formerly isolated populations into secondary contact at suture zones between seas including ECS/SCS (Ni *et al.* 2014). In fact, paleo habitat suitability projections showed a clear geographical separation between south and north populations during the LGM, which provide ecological support for the observed differentiation of ECS-YS and SCS lineages of *S. schlegeli*. We hypothesize that Taiwan Strait land bridge may have served as a barrier to the dispersal of coastal marine species during the glacial periods, favouring inter-population genetic divergences.

Gene exchange among YS and ECS populations of *S. schlegeli* appears to be relatively frequent, as revealed by the genetic ancestry analyses, despite the fact that ECS and YS populations are clearly separated based on the PCA and phylogenetic tree clustering. During glacial periods, the northern regions of the YS were covered by land, and the ECS populations of *S. schlegeli* may have experienced expansion concomitant with a rise in sea level during interglacial periods, leading to genetic exchange between the ECS and YS populations (Chen *et al.* 2006; Yu & Zhou 2001). In parallel, genetic breaks may be attributed to freshwater outflow from rivers, as demonstrated, for example, in the Amazon basin (Rocha *et al.* 2002). As the third largest river in the world, Yangtze (Changjiang) River pours into the ECS with an average annual discharge of 924 billion m<sup>3</sup> (Ni *et al.* 2012). While the barrier effect of the huge freshwater outflow on gene flow of coastal species is still a controversial issue, we suggest that the freshwater outflow may have affected the gene change of marine species in a more recent historical scale following potential changes in the estuary and variable river discharges. Given the fact that the freshwater outflow seems to affect mostly intertidal species with a strong habitat specificity (Dong *et al.* 2012), and SDM projections suggest that a large part of coastal areas of China, Japan, and Korea are suitable for *S. schlegeli*, we are inclined to believe that the phylogeographical pattern of *S. schlegeli* is largely explained by the historical glaciation.

It is worth noting that pipefishes are unique among fishes in that the males have a brood pouch and are ovoviviparous. Moreover, they are generally considered to be slow swimmers. These life history traits render them sensitive to environmental change (Vincent *et al.* 2011; Wilson *et al.* 2003; Zhang *et al.* 2017). Compared with other marine fishes, environmental heterogeneity is considered more likely to affect the *S. schlegeli* population structure due to local adaptation and the limited gene flow between different geographical populations. Hypervolume analysis indicated that *S. schlegeli* from different geographic regions diverge in their realized niche, implying that the three clades have adapted to largely distinct environmental conditions including temperature and salinity. A recent study of *Syngnathus typhle* based on geometric morphometrics, prey availability, and dietary composition, demonstrated a pronounced variation in snout morphology across its distribution range, which may contribute to its adaptation in novel environments (Wilson *et al.* 2020). In the present study, a north-to-south variation of cranial morphology was detected in different *S. schlegeli*

populations, which indicates that the individuals in the south of the distributional range (SCS) have smaller eye, and the opposite patterns are observed in the north (ECS and YS). The differences detected in the eye phenotype of pipefish from different populations may be related to differences in various environmental conditions, such as sea water turbidity, and food type and availability. Divergent selection resulted from the diverse environmental conditions could constitute barriers to gene flow (Tobler *et al.* 2008), and identification of loci or genes under natural selection is important for determining the genetic basis of local adaptation affecting fitness traits (Zardi *et al.* 2013). A total number of 143 genes were annotated from the selected genome regions revealed by genome scan method, among which many key genes are involved in growth (*pax7*) (Akolkar *et al.* 2016), cold adaptation (*csde1*) (Yang *et al.* 2012), and eye development (*eyes*, *rx3*) (Loosli *et al.* 2003; Yuet *et al.* 2016). Accordingly, local populations of *S. schlegeli* may benefit from the adaptive variation to improve fitness of this species to local environments.

Despite the fact that marine fish generally show high genetic diversity and shallow population structure (Takahashi *et al.* 2015; Utter & Seeb 2010), the results of this study confirm that *S. schlegeli* has obvious geographical population structures along its distribution range in northwestern Pacific. The divergence among the different geographical populations may be related to the geological events, as well as adaptation to the habitat heterogeneity across the latitudinal gradients.

## Conclusion

We disentangled the morphological, molecular, and ecological mechanisms underlying adaptive evolution against a background of high-level gene flow in a seemingly barrier-free environment. We showed that the effect of historical processes during the last glacial maximum, in concert with niche-driven ecological differentiation, lead to the establishment of three distinct clades across a widely distributed marine pipefish. Ultimately, our study shows that even the sea environment maintains a vast potential for adaptive radiation and ecological specialization. In fact, when multiple lines of evidence and temporal scales are taken into consideration, it turns out that ‘marine speciation’ is actually far from being a ‘paradox’.

## Acknowledgements

We are grateful to Prof. Liu Jinxian and Wang Zhongming for sample collection.

## Funding

This research was supported by the National Natural Science Foundation of China (41825013, 41890850), the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0407), the Guangdong Basic and Applied Basic Research Foundation (2019A1515110199), and the Youth Talent Program Supported by Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao) (2018-MFS-T04).

## Availability of data and materials

All mitochondrial and nuclear gene sequences were deposited into GenBank database under the accession numbers MT467667–MT468179. The sequencing reads of RAD-seq analyzed in the study have been deposited to the CNGB (China National Genebank) with accession CNP0000953 (<https://db.cngb.org/cnsa/>).

## Authors’ contributions

XW, QL, and AH were involved in designing the study; YHZ and GQ were involved in sampling; XW, AH, ZXZ, and SM analyzed the data; XW and ZXZ drafted the manuscript; XW, QL, ZXZ, and SM contributed to the revisions and the edits of the manuscript. All authors approved the final version of the manuscript.

## Ethics approval and consent to participate

We confirm that the collection of fish samples for this study were in accordance with relevant animal ethics guidelines and approved by the Ethical Committee of South China Sea Institute of Oceanology, Chinese Academy of Sciences.



## Consent for publication

Not applicable

## Competing interests

The authors declare that there are no conflicts of interest.

## References

- Akolkar KS, Asaduzzaman M, Kinoshita S, Asakawa S, Watabe S (2016) Characterization of Pax3 and Pax7 genes and their expression patterns during different development and growth stages of Japanese pufferfish *Takifugu rubripes*. *Gene*, **575**, 21-28.
- Araújo MB, Anderson RP, Barbosa AM, *et al.* (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, **5**, eaat4858.
- Ashburner MM, Ball CAC, Blake JA, *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, **25**, 25-29.
- Bierne N, Bonhomme F, David P (2003) Habitat preference and the marine-speciation paradox. *Proc Biol Sci*, **270**, 1399-1406.
- Blonder B, Lamanna C, Violle C, Enquist BJ (2014) The n-dimensional hypervolume. *Global Ecology and Biogeography*, **23**, 595-609.
- Blonder B, Morrow CB, Maitner B, Harris DJ (2018) New approaches for delineating n-dimensional hypervolumes. *Methods in Ecology and Evolution*, **9**, 305-319.
- Bolnick DI, Stutz WE (2017) Frequency dependence limits divergent evolution by favouring rare immigrants over residents. *Nature*, **546**, 285-288.
- Carvalho JC, Cardoso P (2018) Decomposing the causes for niche differentiation between species using hypervolumes. *bioRxiv*.
- Chen Y, Chen Y, He D, Sui X (2006) Current Status and Perspective of Biogeography Study in Taiwan Strait, China. *Chinese Journal of Zoology*, **41**, 118-122.
- Chen Z, Zhang Y, Han ZQ, Song N, Gao TX (2017) Morphological characters and DNA barcoding of *Syngnathus schlegeli* in the coastal waters of China. *Chinese Journal of Oceanology & Limnology*, 1-11.
- Cunha RL, Tenorio MJ, Afonso C, Castilho R, Zardoya R (2008) Replaying the tape: Recurring biogeographical patterns in Cape Verde Conus after 12 million years. *Molecular Ecology*, **17**, 885-901.
- Devictor V, Clavel J, Julliard R, *et al.* (2010) Defining and measuring ecological specialization. *Journal of Applied Ecology*, **47**, 15-25.
- Dong YW, Wang HS, Han GD, *et al.* (2012) The impact of Yangtze River discharge, ocean currents and historical events on the biogeographic pattern of *Cellana toreuma* along the China coast. *PLoS ONE*, **7**, e36178.
- Dormann CF, Elith J, Bacher S, *et al.* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27-46.
- Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169-175.
- Excofer L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47-50.
- Fox LR, Morrow PA (1981) Specialization: Species Property or Local Phenomenon? *Science*, **211**, 887-893.

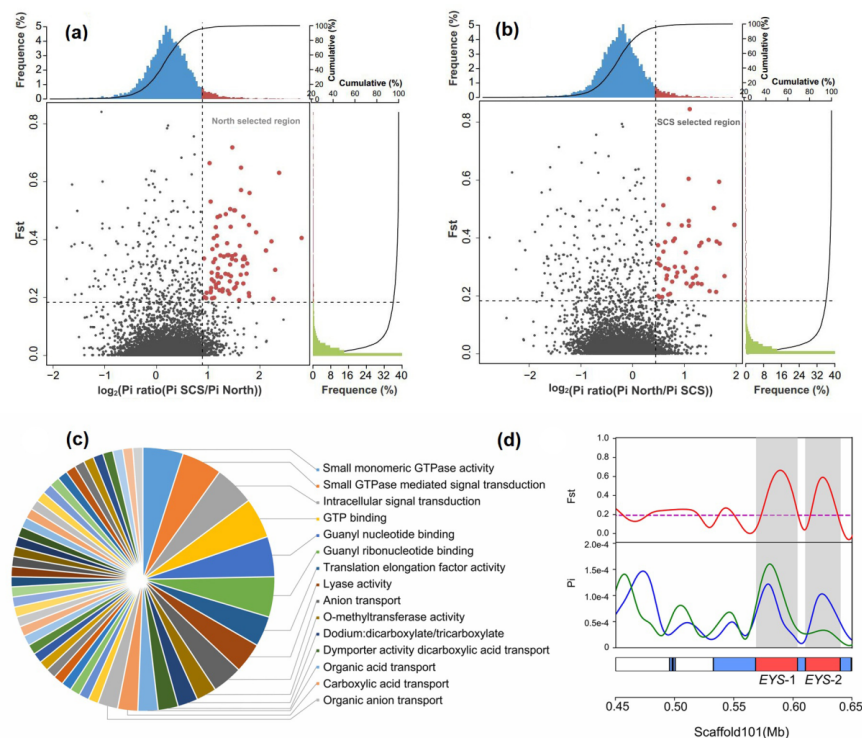
- Garant D, Kruuk LE, Wilkin TA, Mccleery RH, Sheldon BC (2005) Evolution driven by differential dispersal within a wild bird population. *Nature*, **433** , 60-65.
- GBIF.org (13 December 2019) GBIF Occurrence Download <https://doi.org/10.15468/dl.ec6hrx>
- Gillespie RG, Bennett GM, De Meester L, *et al.* (2020) Comparing Adaptive Radiations Across Space, Time, and Taxa. *Journal of Heredity*, **111** , 1-20.
- Hendry AP, Wenburg JK, Bentzen P, Volk EC, Quinn TP (2000) Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science*, **290** , 516-519.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405** , 907-913.
- Hewitt GM (2004) Genetic Consequences of Climatic Oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, **359** , 183-195.
- Kawecki T, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7** , 1225-1241.
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33** , 159-174.
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27** , 2987-2993.
- Li H, Durbin R (2009) *Fast and accurate short read alignment with Burrows-Wheeler transform* Oxford University Press.
- Liu J, Gao T, Shifang WU, Zhang Y (2007) Pleistocene isolation in the Northwestern Pacific marginal seas and limited dispersal in a marine fish, *Chelon haematocheilus* (Temminck & Schlegel, 1845). *Molecular Ecology*, **16** , 275-288.
- Loosli F, Staub W, Finger-Baier KC, *et al.* (2003) Loss of eyes in zebrafish caused by mutation of *chokh / rx3* . *EMBO Journal*, **4** , 894-899.
- Loxdale HD, Gugs L, Harvey JA (2011) The evolutionary improbability of ‘generalism’ in nature, with special reference to insects. *Biological Journal of the Linnean Society*, **103** , 1-18.
- Mammola S (2019) Assessing similarity of  $n$  -dimensional hypervolumes: Which metric to use? *Journal of Biogeography*, **46** , 1-12.
- Mammola S, Arnedo MA, Fišer C, *et al.* (2020) Environmental filtering and convergent evolution determine the ecological specialization of subterranean spiders. *Functional Ecology*, **34** , 1064-1077.
- Mammola S, Cardoso P (2020) Functional diversity metrics using kernel density  $n$ -dimensional hypervolumes. *Methods in Ecology and Evolution*, early view.
- Ni G, Li Q, Kong L, Yu H (2014) Comparative phylogeography in marginal seas of the northwestern Pacific. *Molecular Ecology*, **23** , 534-548.
- Ni G, Li Q, Kong L, Zheng X, Crandall KA (2012) Phylogeography of Bivalve *Cyclina sinensis* : Testing the Historical Glaciations and Changjiang River Outflow Hypotheses in Northwestern Pacific. *PLoS ONE*, **7** , e49487.
- Ogata H, Goto S, Sato K, *et al.* (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **27** , 29-34.
- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43** , 1223-1232.
- Palumbi SR (1994) Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology & Systematics*, **25** , 547-572.

- Purcell S, Neale B, Todd-Brown K, *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81** , 559-575.
- Rocha LA, Bass AL, Robertson DR, Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology*, **11** , 243-251.
- Rocha LA, Robertson DR, Roman J, Bowen BW (2005) Ecological speciation in tropical reef fishes. *Proc Biol Sci*, **272** , 573-579.
- Sabeti PC, Schaffner SF, Fry B, *et al.* (2006) Positive natural selection in the human lineage. *Science*, **312** , 1614-1620.
- Sbrocco EJ (2014) Paleo-MARSPEC: gridded ocean climate layers for the mid-Holocene and Last Glacial Maximum. *Ecology*, **95** , 1710.
- Sbrocco EJ, Barber PH (2013) MARSPEC: ocean climate layers for marine spatial ecology: *Ecological Archives* E094-086.
- Shipley LA, Forbey JS, Moore BD (2009) Revisiting the dietary niche: When is a mammalian herbivore a specialist? *Integrative & Comparative Biology*, **3**.
- Smith TB, Wayne RK, Girman DJ, Bruford MW (1997) A Role for Ecotones in Generating Rainforest Biodiversity. *Science*, **276** , 1855-1857.
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science*, **240** , 1285-1293.
- Takahashi H, Takeshita N, Tanoue H, *et al.* (2015) Severely depleted genetic diversity and population structure of a large predatory marine fish (*Lates japonicus*) endemic to Japan. *Conservation Genetics*, **16** , 1155-1165.
- Tamaki K, Honza E (1991) Global tectonics and formation of marginal basins: Role of the Western Pacific. *Translated World Seismology*, **14** , 224-230.
- Tang H, Peng J, Wang P, Risch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. *Genetic Epidemiology*, **28** , 289-301.
- Thuiller W, Georges D, Engler R, Breiner F (2020) *biomod2: Ensemble Platform for Species Distribution Modeling*. R package version 3.3-7.1. <https://CRAN.R-project.org/package=biomod2>
- Thuiller W, Guéguen M, Renaud J, Karger DN (2019) Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, **10** , 1-9.
- Tobler M, Dewitt TJ, Schlupp I, *et al.* (2008) Toxic hydrogen sulfide and dark caves: phenotypic and genetic divergence across two abiotic environmental gradients in *Poecilia mexicana*. *Evolution*, **62** , 2643-2659.
- Todd Streelman J, Danley PD (2003) The stages of vertebrate evolutionary radiation. *Trends in Ecology & Evolution*, **18** , 126-131.
- Utter F, Seeb J (2010) A perspective on positive relationships between genetic diversity and abundance in fishes. *Molecular Ecology*, **19** , 4831-4833.
- Vilella A, Severin JV, A, Heng L, Durbin R, Birney E (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research*, **19** , 327-335.
- Vincent ACJ, Foster SJ, Koldewey HJ (2011) Conservation and management of seahorses and other Synbranchidae. *Journal of Fish Biology*, **78** , 1681-1724.
- Wainwright PC, Bellwood DR, Westneat MW, Grubich JR, Hoey AS (2004) A functional morphospace for the skull of labrid fishes: patterns of diversity in a complex biomechanical system. *Biological Journal of the Linnean Society*, **82** , 1-25.

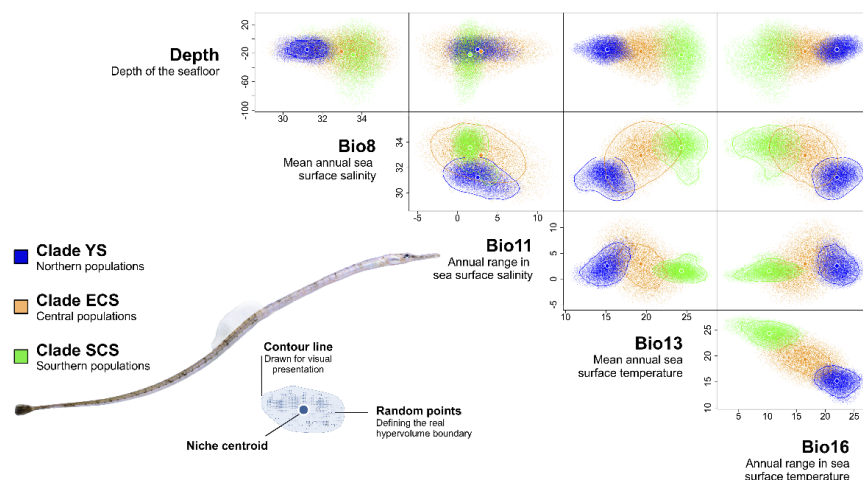
- Wainwright PC, McGee MD, Longo SJ, Hernandez LP (2015) Origins, Innovations, and Diversification of Suction Feeding in Vertebrates. *Integrative and Comparative Biology*, **55** , 134-145.
- Wang MS, Li Y, Peng MS, *et al.* (2015) Genomic Analyses Reveal Potential Independent Adaptation to High Altitude in Tibetan Chickens. *Molecular Biology & Evolution*, **32** , 1880-1889.
- Wang P (1999) Response of Western Pacific marginal seas to glacial cycles: Paleoceanographic and sedimentological features. *Marine Geology*,**156** , 5-39.
- Wang PX, Sun XJ (1994) Last glacial maximum in China: comparison between land and sea. *Catena*, **23** , 341-353.
- Watanabe S, Watanabe Y (2001) Brooding season, sex ratio, and brood pouch development in the seaweed pipefish, *Syngnathus schlegeli* , in Otsuchi Bay, Japan. *Ichthyological Research*, **48** , 155-160.
- Wilson AB, Ahnesjö I, Vincent ACJ, Meyer A (2003) The dynamics of male brooding, mating patterns, and sex roles in pipefishes and seahorses (family Syngnathidae). *Evolution*, **57** , 1374-1386.
- Wilson AB, Wegmann A, Ahnesjö I, Goncalves JMS (2020) The evolution of ecological specialization across the range of a broadly distributed marine species. *Evolution*, **74-3** , 629-643.
- Xu J, Chan TY, Tsang LM, Chu KH (2009) Phylogeography of the mitten crab *Eriocheir sensu stricto* in East Asia: Pleistocene isolation, population expansion and secondary contact. *Molecular Phylogenetics and Evolution*,**52** , 45-56.
- Yang C, Wang L, Siva VS, *et al.* (2012) A novel cold-regulated cold shock domain containing protein from scallop *Chlamys farreri* with nucleic acid-binding activity. *Plos One*, **7** , e32012.
- Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*,**88** , 76-82.
- Yu HX, Zhou YW (2001) Geomorphology and geological features of the continental shelf in northern and western Taiwan, China. *Science in China (Series D)*, **31** , 486-495.
- Yu M, Liu Y, Li J, *et al.* (2016) Eyes shut homolog is required for maintaining the ciliary pocket and survival of photoreceptors in zebrafish. *Biology Open*, **5** , 1662-1673.
- Zardi GI, Nicastro KR, Costa JF, Serrão EA, Pearson GA (2013) Broad scale agreement between intertidal habitats and adaptive traits on a basis of contrasting population genetic structure. *Estuarine Coastal & Shelf Science*, **131** , 140-148.
- Zhang BD, Xue DX, Wang J, *et al.* (2016) Development and preliminary evaluation of a genomewide single nucleotide polymorphisms resource generated by RAD-seq for the small yellow croaker (*Larimichthys polyactis* ). *Mol Ecol Resour*, **16** , 755-768.
- Zhang YH, Qin G, Zhang HX, Wang X, Lin Q (2017) DNA barcoding reflects the diversity and variety of brooding traits of fish species in the family Syngnathidae along China's coast. *Fisheries Research*,**185** , 137-144.



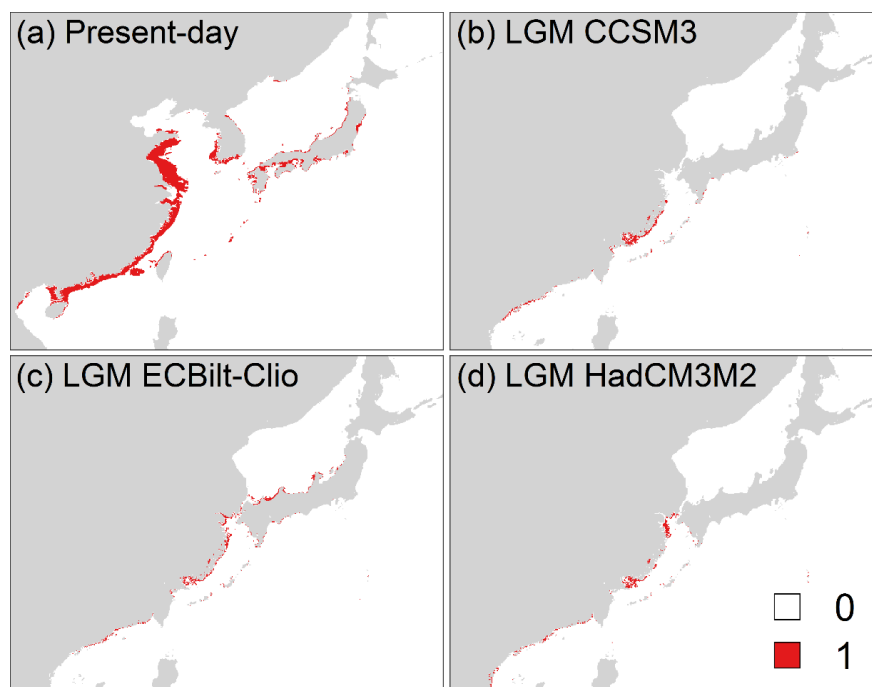
**Figure 2** Genetic relationships and population structure of 69 sequenced pipefish based on SNPs. (a) Neighbor-joining tree of the 69 pipefish. The scale bar represents proportional similarity (p distance); (b) Principal component analysis plots based on the first two principal components (PC); (c) Genetic structure and individual ancestry as derived from FRAPPE. The length of each colored segment represents the proportion of the individual's genome from  $K = 2-5$  ancestral populations.



**Figure 3** Selective sweep analysis of the populations from different regions. (a, b) Distribution of Pi ratios (North/SCS, SCS/North) and  $F_{ST}$  values calculated using the sliding window approach. Data points located to the right of the vertical dashed line (corresponding to the 5% right tails of the empirical Pi ratio distribution, where the  $\pi$  ratios are 0.887 and 0.446, respectively), and above the horizontal dashed line (upper 5% tail of the  $F_{ST}$  distribution, where  $F_{ST}$  is 0.183) were identified as selected regions for North populations and SCS populations (red points), respectively; (c) The distribution of selected genes (North vs. SCS) related to Gene Ontology terms; (d) Pi ratios and  $F_{ST}$  values plot using a 10-kb sliding window of the genes (North vs. SCS) with strong selective sweep signals, and the *EYS* gene regions are marked in red



**Figure 4** Pairplots representing the estimated ecological niche for the three genetic lineages. For each clade, colored points are 10,000 points stochastically sampled from each hypervolume, and represents its real shape and boundaries



**Figure 5** Binary predictions of habitat suitability for *Syngnathus schlegeli* under present-day (a) and Last Glacial Maximum (LGM) (b, c, d) climatic conditions. Gray color represents land, white color (0) indicates unsuitable and red color (1) suitable marine range. The threshold to convert continuous projections into binary results was set at 0.69. CCSM3, ECBilt-Clio, and HadCM3M2 represent different general circulation models used to simulate paleoclimates

**Table 1.** Sampling information of *Syngnathus schlegeli*

Locations	Locations	Geographic Coordinates	Sample Number	Sample Number	Sample Number
				Morphology	Mitochondrial g
<b>YS</b>	QD1	E 121.977, N 36.611	15	15	14
	QD2	E 121.839, N 36.573	25	18	25
	HZW	E 119.386, N 34.927	48	30	43
<b>ECS</b>	ZS1	E 122.476, N 30.848	15	11	15
	ZS2	E 122.477, N 30.093	12	9	11
	ZS3	E 122.175, N 29.544	25	14	23
<b>SCS</b>	ZH	E 113.729, N 21.874	40	28	38
	HK1	E 110.645, N 20.203	55	20	52
	HK2	E 110.241, N 20.114	45	18	44
Total number	Total number	Total number	<b>280</b>	<b>163</b>	<b>265</b>

Yellow Sea (YS), East China Sea (ECS), South China Sea (SCS), Qingdao (QD), Haizhouwan (HZW), Zhoushan (ZS), Zhuhai (ZH), Haikou (HK)

**Table 2.** Summary of relative warp analysis and ANOVA results of RWs

RW	ANOVA	ANOVA	ANOVA	SV	PVE(%)	Description of the variation
	<i>d.f.</i>	<i>F</i>	<i>P</i>			
RW1	2	0.811	0.446	0.36157	41.81	length of snout and gill cover
RW2	2	0.238	0.789	0.22652	16.41	angle of the snout
RW3	2	56.937	<b>&lt;0.001</b>	0.15502	7.69	eye size

RW, Relative warps; SV, Singular values; PVE, percent variance explained

**Table 3** Pairwise niche differentiation among hypervolumes, as estimated through a distance metric (lower panel; Distance between centroids) and an overlap metrics (upper panel;  $\beta$  total =  $\beta$  replacement +  $\beta$  richness). The diagonal of the matrix is the volume of each hypervolume.

	Clade YS	Clade ECS	Clade SCS
<b>Clade YS</b>	15747.58 [v]	$0.97 = 0.04 + 0.93 [\beta]$	$1.00 = 0.60 + 0.40 [\beta]$
<b>Clade ECS</b>	7.58 [dc]	296963.75 [v]	$0.94 = 0.12 + 0.82 [\beta]$
<b>Clade SCS</b>	16.75 [dc]	9.35 [dc]	36572.67 [v]

$\beta$  = Beta diversity; dc = distance between hypervolume centroid; v = volume

### Supporting information

**Figure S1** Sampling sites of *Syngnathus schlegeli* along the coastline of China (QD=Qingdao; HZW=Haizhouwan; ZS=Zhoushan; ZH=Zhuhai; HK=Haikou). Background gradient of color indicate water depth.

**Figure S2** Presence records of *Syngnathus schlegeli* used in this study. All distribution data (black points, red points, blue points, and green points) were used to develop species distribution model. Records in red represent Yellow Sea clade (56 presence records), records in blue correspond to East China Sea clade (23 records), and records in green indicate South China Sea clade (29 records). Presence records of *S. schlegeli* along the coastal waters of mainland China were divided into three distinct clades based on the present genetic cluster results and previous phylogeographic studies conducted in the same region (Dong *et al.* 2012;



Liu *et al.* 2007; Ni *et al.* 2012; Xuet *et al.* 2009).

**Figure S3** Multi-collinearity analysis results between eleven marine predictor variables. bathymetry: depth of the seafloor; bio8: mean annual SSS (psu); bio9: SSS of the freshest month (psu); bio10: SSS of the saltiest month (psu); bio11: annual range in SSS (psu); bio12: annual variance in SSS (psu); bio13: mean annual SST (°C); bio14: SST of the coldest month (°C); bio15: SST of the warmest month (°C); bio16: annual range in SST (°C); bio17: annual variance in SST (°C). SSS presents sea surface salinity and SST means sea surface temperature. Predictors in same box are highly correlated (i.e. pairwise Pearson’s correlation coefficients  $|r| > 0.7$ ) and only one was chosen for subsequent analyses.

**Figure S4** True skill statistic (TSS) (a) and the area under the receiver operating characteristic curve (AUC) (b) values of ten modelling algorithms. Dashed lines represent cutoff values for TSS (0.75) and AUC (0.90). ANN: artificial neural network; CTA: classification tree analysis; FDA: flexible discriminant analysis; GAM: generalized additive model; GBM: generalized boosting model; GLM: generalized linear model; MARS: multiple adaptive regression splines; Maxent: maximum entropy; RF: random forest; SRE: surface range envelop. Each algorithm was run ten times and results are expressed as mean  $\pm$  standard error.

**Figure S5** Continuous predictions of habitat suitability for *Syngnathus schlegeli* under present-day (a) and Last Glacial Maximum (LGM) (b, c, d) climate conditions. CCSM3, ECBilt-Clio, and HadCM3M2 indicate different general circulation models used to simulate paleoclimates.

**Table S1** References of distribution range of *Syngnathus schlegeli*

**Table S2** Singular values and percent explained for relative warps based on the geometric morphometric data

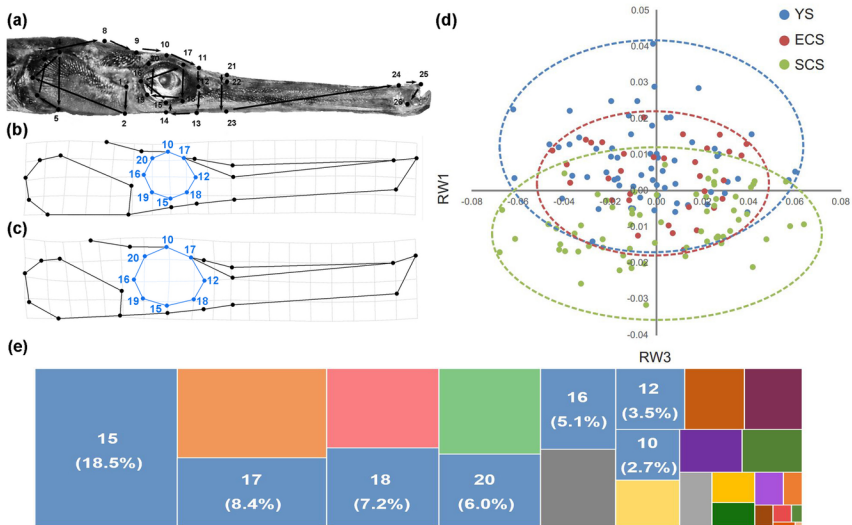
**Table S3** Statistics of properties of the RAD sequences after filtering

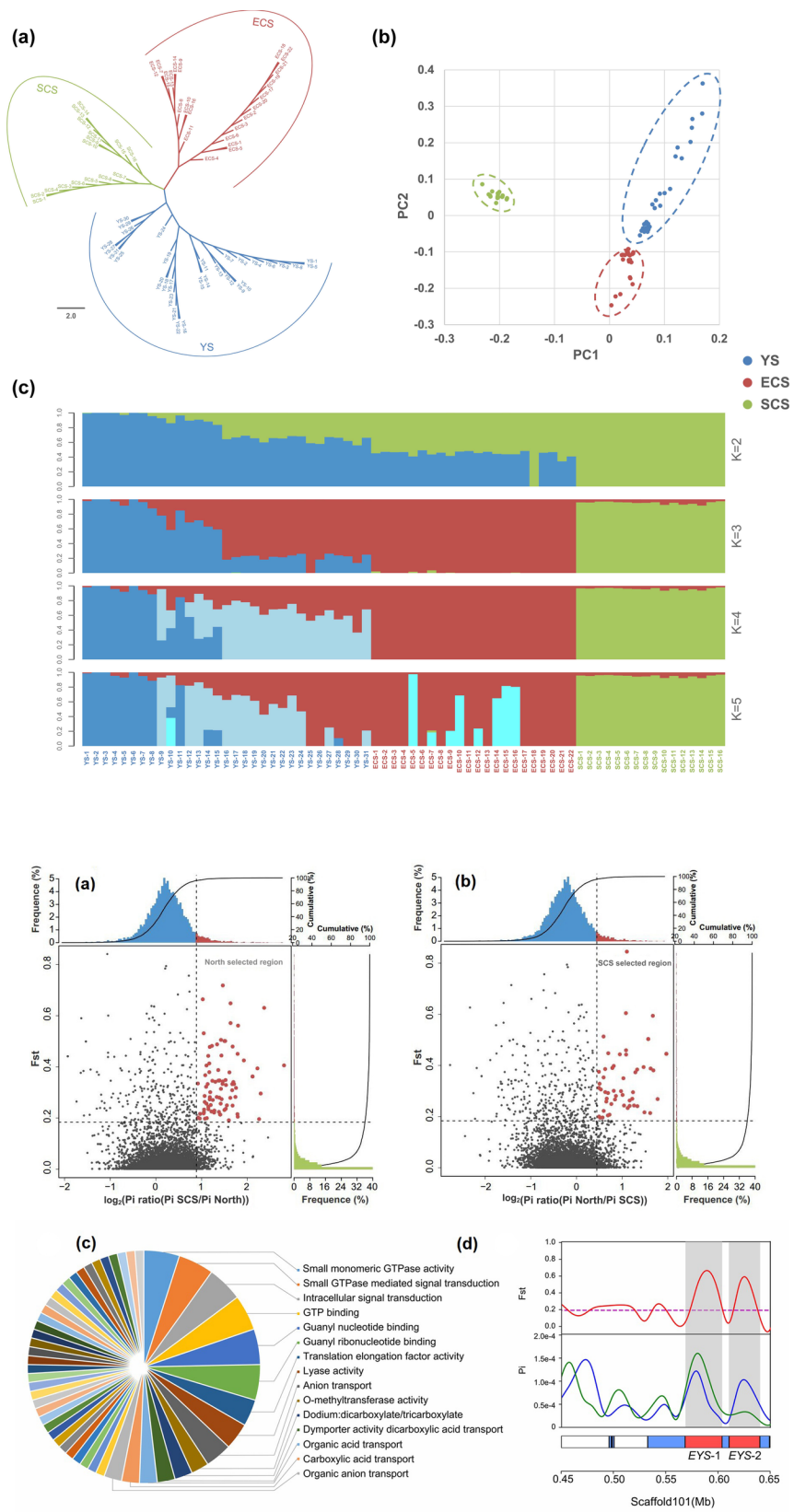
**Table S4** Statistics of the depth and coverage of reads after mapping to the reference genome

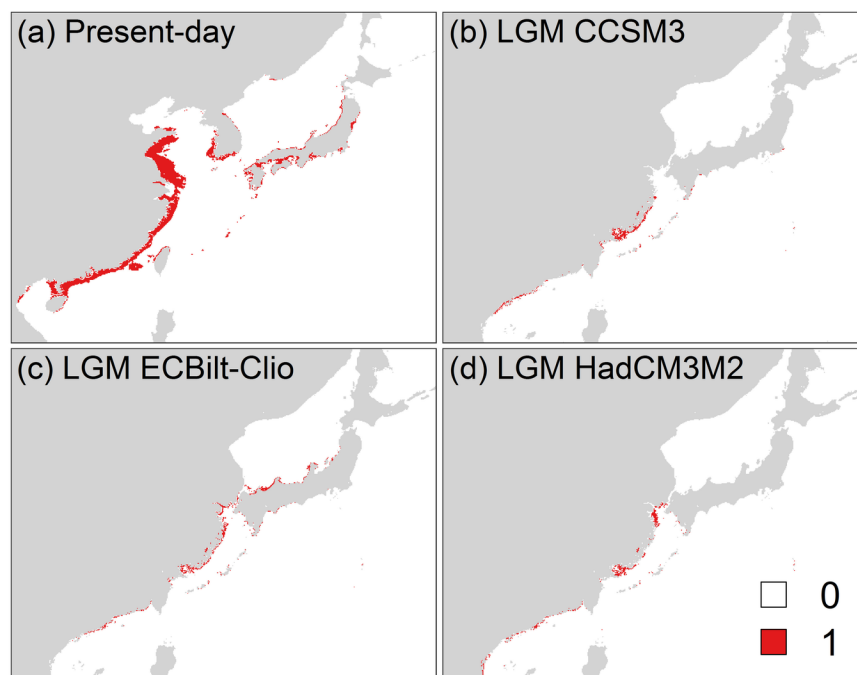
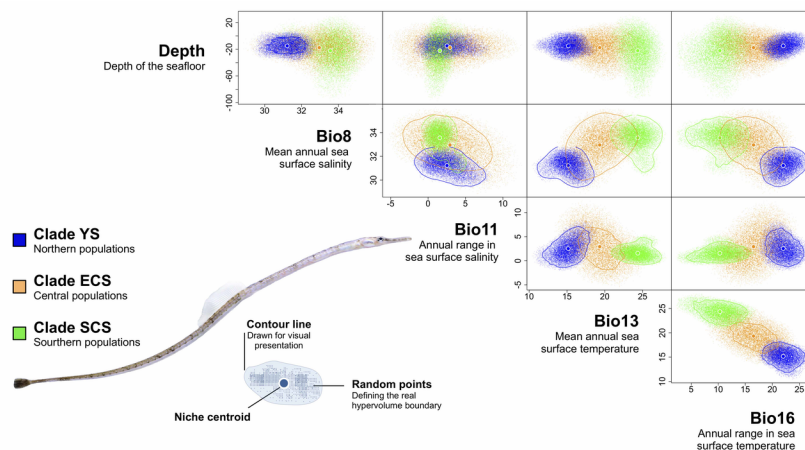
**Table S5** Statistics of SNPs after filtering

**Table S6** Population pairwise  $F_{ST}$  between different *S. schlegeli* populaitons base on the *cytb* and *snx33* gene

**Table S7** Statistics of selected genes detected by selection sweep







## Hosted file

Table 1 Sampling information of *Syngnathus schlegelii*.docx available at <https://authorea.com/users/329247/articles/456279-revisiting-the-marine-speciation-paradox-using-genomic-morphometric-and-ecological-evidence-in-pipefish>

## Hosted file

Table 2 RW analysis and ANOVA results.docx available at <https://authorea.com/users/329247/articles/456279-revisiting-the-marine-speciation-paradox-using-genomic-morphometric-and-ecological-evidence-in-pipefish>

## Hosted file

Table 3 Niche differentiation among hypervolumes.docx available at <https://authorea.com/users/329247/articles/456279-revisiting-the-marine-speciation-paradox-using-genomic-morphometric-and-ecological-evidence-in-pipefish>