ddRAD-seq Reveals Cryptic Phylogeographic Structure and Adaptive Population Differentiation of Hyporhamphus Intermedius in Heterogeneous Environments from China Mainland

Gongpei Wang
Sun Yat-Sen University

Han Lai
Sun Yat-Sen University

Sheng Bi
Sun Yat-Sen University

Dingli Guo
Sun Yat-Sen University

Xiaopin Zhao
Sun Yat-Sen University

Xiaoli Chen
Sun Yat-Sen University

Shuang Liu
Sun Yat-Sen University

Xuange Liu
Sun Yat-Sen University

Yuqin Su
Sun Yat-Sen University

Huadong Yi
Sun Yat-Sen University

Guifeng Li (liguif@mail.sysu.edu.cn)
Sun Yat-Sen University

Research article

Keywords: Hyporhamphus intermedius, heterogeneous environments, RAD-seq, adaptive differentiation, phylogeographic structure
Abstract

Background

Species differentiation and local adaptation in heterogeneous environments have been getting a large focus but poor knowledge of the mechanism. With widespread in coastal areas and the hyperdiverse freshwater systems in China, Hyporhamphus intermedius is an excellent material for phylogeography and local adaptation research. Here, 156 individuals were sampled in 8 sites from heterogeneous aquatic habitats to comprehend environmental and genetic contributions to phenotypic divergence.

Results

32,744 SNPs among populations and obvious population differentiation was found by ddRAD-seq among groups from the different watersheds. Population genetic structure was partly geographic, successfully classifying all individuals into three genetic subgroups corresponding to the North, the South and the Plateau Lakes with more migrations within region. Due to the transplant event, the strikingly increasing divergence in a short time indicated the tendency of local differentiation for introduced populations in the Plateau Lakes. Dorsal and anal traits varied widely between the Southern and others, which highlighted previously unrecognized lineages. Furthermore, we inferred the whole population is in response to balancing selection from neutrality test with positive value (Tajima's D=2.773), and demographic history underlay two main split events from the northern ancestral population. Subgroup ranges appear to counterparts to geographic regions with heterogeneous hydrological factors, indicating these features are likely important drivers of diversification.

Conclusions

Accordingly, our results revealed that genetic polymorphism and divergence of Hyporhamphus intermedius in heterogeneous environments might be ascribed to a consequence of balancing selection, gene flow at microscales, and the impact of abiotic factors, verifying potentially that genetic differences among populations are the result of both genetic canalization and plastic responses to environmental heterogeneity.

Background

Elucidating the origins of phenotypic diversity and the processes that different forms of dispersal generate spatial patterns and structure of genetic variation remain a challenging endeavor for evolutionary biologists and ecologists (Avise 2000; Saenz-Agudelo 2015). Species occupy a heterogeneous environment on a large geographical scale, which may shape patterns of gene flow among populations across the distributing region and drive population divergence (Slatkin 1987; Riginos and Liggins 2013). Alternatively, when different populations experience different environmental conditions, local selective pressures can modulate the underlying genomic architecture and drive phenotypic divergence, thereby promoting local adaptation and novel natural lineage (George 2000; Via
Local adaption is of critical importance in the divergence and novel speciation, which occurs due to differential pressures of natural selection on populations from different environments result in allele frequency shifts, and advantageous variations accumulate over many generations to maximize fitness in local environment, forming distinctive genetic and phenotypic traits eventually (Savolainen et al. 2013). Studies investigating the genomic basis of adaptation often compare genomic information and environmental variables of populations for the comprehension of the evolutionary forces driving local adaption (Holt and Gaines 1992; Sultan and Spencer 2002; Wang et al 2016). Local selective pressure counterparts to certain geographical and environmental gradients, e.g. latitude, longitude, and altitude. Genome–environmental association analysis expects to manifest the prints of selection with increasing geographical scale owing to larger environmental gradient differences (Fraser et al. 2011). Overall, discerning relative contribution of geographical environment factors and identifying factors such as natural selection, genetic drift, and dispersion that modify patterns of genomic variation is critical in interpreting the mechanism of how organisms adapt to the local environment and initiate evolutionary diversification.

Fish are ideal organisms for studying local adaptive divergence and phenotypic plasticity (Rajkov 2018). Given the complex geomorphic conditions in China with dramatic climate change at high altitude and latitude, the hyperdiverse freshwater systems (e.g., Pearl and Yangtze rivers and Plateau lakes) exhibit remarkably high levels regarding fish diversification and peculiarly prone to phylogenetic endemism and novel species, as it is probable that these independent areas separated by long distance also facilitated differentiation and genetic landscape formation of fishes (Stoddart 1978; Tang et al. 2017). Reported studies have highlighted that fish diversity assessments are based almost entirely on morphological differentiation, so that knowledge of the true diversity of different regions, as well as the underlying mechanisms shaping local adaptation and biogeography within the system, remains rare and limited depth (Wiens 2011).

Population genomic studies of fish in heterogeneous environments have thus far been restricted to model species for evolutionary mechanisms comprehension including salmonid (Narum et al. 2017), African cichlids (Brawand et al. 2014), and three-spined stickleback (Jones et al. 2012). Indeed, the phenomena can be broadly recognized in other similar species. In contrast to well-known the model organism, in few studies, *Hyporhamphus intermedius* (Family Hemiramphidae) is an anadromous halfbeak as forage fish that exhibits a large geographic dispersal within the estuarine area of China Mainland, spanning along the coasts from the East China Sea to the South China Sea (Collette and Su 1986; Li and Zhang 2011). Notably, the south-north divergence of its morphological traits is prominent. For example, the number of dorsal and anal fins, vertebra in the northern population is significantly higher than that in the southern population (Collette and Su 1986), which are notable for putatively adaptive features. Additionally, some halfbeaks can also be found in the upper reaches of the Pearl River, hundreds of miles away from shore, which involves a recent introduction intensively conducted since the 1980s (Zhu and Chen 1989), however, where the introduced populations originated from is unclear. The evolutionary dynamics remains another challenging aspect that Plateau Lakes created a unique set of environmental conditions including an altitudinal gradient of wide temperature range, low dissolved oxygen, probably leading to a
number of rapids-adaptive features of introduced populations (Cui et al. 2008; Sihai et al. 2013). The species tolerate different selective pressures from increasing geographical scale and exhibit tangible evidence of strong adaptive capacities. The whole populations distributed in dissimilar habitats typically display broad phenotypic plasticity and environmental tolerance. Preliminary investigations imply extensive morphological diversification across this range, cryptic lineages within whole groups are incapable to identify (Collette and Su 1986). It is unknown whether the selective pressure resulting from environmental changes can be observed from large-scale sampling with extensive geographical coverage.

It is important to combine genetic, ecological components and phenotypes in interpreting adaptation and phylogeographical patterns we observed from empirical studies (Gamboa & Watanabe 2019). Under the rapid genetic technological development, taking advantage of molecular genetic approaches to access genetic diversity can enhance understanding of the process of underlying speciation in heterogeneous freshwater systems (Nosil et al. 2017). Double digest restriction-site associated DNA (ddRAD-seq) involves genome-wide sequencing reads into thousands of orthologous loci, when no reference genome is available (Andrews et al. 2016, Peterson et al. 2012), and has been reported as the most robust approaches to disentangling the evolutionary history of strong dispersal species. For instance, the high rates of mutation experienced by the whole genome allow more rapid accumulation of mutations, and its richer polymorphism loci could result in a better resolution when trying to differentiate populations and identify close relatives and local adaptation with its potential dispersal events among populations (Maroso 2018; Kyriakis et al. 2019). Entirely based on morphological assessments, our previous estimates of *H. intermedius* diversity within the group demonstrated phenotypic differentiation of intraspecific biological traits and NGS data has been scarcely used (Zhu and Chen 1989; Collette and Su 1986). The increasing integration between morphological and molecular approaches to assess the diversity, evolutionary history and phylogeography of the exceptionally diverse ichthyofauna will foster a better understanding of the heterogeneous freshwater of speciation (Alter 2016; Fišer 2018).

As living within distributed heterogeneous environments, local abiotic parameters may have stronger effects on genetic diversity and drive the spatial distribution of freshwater species. Genetic structure and the relative strength of selection are hypothesized to influence the distribution of genomic variation along with different spatial scales. Here, an integrated approach combining phenotypic, environmental and genomic data was used to elucidate the drivers and processes of local divergence of *H. intermedius*. The aim of this research was therefore to probe the genetic diversity, demographic history and environmental adaptation of *H. intermedius* using the ddRAD approach derived SNPs at a genome-wide scale by. Specifically, we plan to address the following questions: (1) characterizing the spatial population structure of *H. intermedius* and genetic diversity at the genomic level; (2) exploring the historical demography and phylogeography of *H. intermedius* on China Mainland; (3) identifying the drivers of local differentiation of *H. intermedius* associated with environmental factors.

**Results**
SNPs from ddRAD-seq

We established ddRAD library and sequenced (150 bp paired-end) from 156 individual fish of 8 populations. The numbers of raw sequence read (average per sample: 16,016,724, range: 12,794,576 – 24,139,524) was obtained. After removing poor quality sequences, the average number of raw reads (15,046,821, range: 11,394,714 – 22,827,782) was applying for sequence mapping and SNP discerning. A total of 13,684,154 tags were detected and the final 32,744 high-quality SNPs showing minor allele frequency (MAF) of > 0.05 across the entire sequence data set were obtained by filtering and screening original SNPs.

For these loci, a significant difference was detected in genotype frequencies and the number of these SNPs among 8 populations. In general, SNPs statistical results of Dianchi (DC), Fuxian Lake (FXH), Chongming (CM), and Suzhou (SZ) were highly similar, in which ref and heterozygous genotypes were relatively dominant. In contrast, the results of all southern populations (GM, SS and DM) demonstrated a consistent genetic variation with more alt genotype, while Jining (JN) population was at an intermediate level (Fig.S1).

Genetic diversity and population divergence

Amplified from 156 individuals, 61,898,184 bp region of genomics gave rise to 32,744 SNP sites. Based on all these variant SNPs, the evaluations of genetic diversity, including Ho, He, π and $F_{IS}$, were estimated (Table 1). There was still a marginal but distinct difference in genetic diversity among groups for different watershed regions. Among the 8 populations, the DM, GM and SS showed slightly lower nucleotide diversity (0.1937 ~ 0.1954) and expected heterozygosity (0.1850 ~ 0.1891) than the others. High genetic diversity was observed in the northern and the plateau group. In addition, the value of $F_{IS}$ within the plateau populations was lower than other groups (0.1113 ~ 0.1147), which might be connected with random genetic drift and the founder effect after transplantation.

AMOVA tests based on 32,744 SNPs was employed to show that the majority of the molecular variances derived from the differentiation among groups (48.8%) or within populations (50.31%) rather than within groups (Table 2). Furthermore, pairwise $F_{ST}$ values of the southern group from the Pearl River System appeared a high level of differentiation (0.296 ~ 0.322) when compared to the remaining groups, coupling with low gene flow (0.527 ~ 0.595) (Table 3), thus corroborating on a distinct cluster in the molecular phylogenetic tree (Fig. 4). However, genetic differentiation between each pair of populations within groups was lower. Due to the transplant event, the genetic divergence of both the northern and the plateau groups did not reach a significant level in a short time (0.046 ~ 0.067), showing at least mild differentiation ($F_{ST} > 0.05$, Nm < 5; Wright 1984) in comparison with the other populations respectively (Table 3), but its pairwise $F_{ST}$ increased strikingly, associated with trends of genetic divergence. And the plateau group exhibited relatively low $F_{IS}$ values (0.1113 ~ 0.1147), indicating a lower level of inbreeding coefficient of an individual relative to a subpopulation. Although genetic differentiation is significant
among groups, no signs of isolation by distance were detected across all locations ($R^2 = 0.2011$, $p = 0.096 > 0.05$).

**Genetics structure, phylogeography and migration**

Genetic structure and phylogenetic analysis determined that the 8 populations were divided into two clades three independent clusters, respectively. The sampled individuals from the Yangtze and Plateau Lakes Systems were assigned to one admixture by phylogenetic analysis, suggesting the close relationship between the two systems (Fig. 3). The southern group formed a distinct cluster that was well separated from the rest of populations in the phylogenetic relationship, lacking gene flow ($0.527 \sim 0.595$) and promoting genetic isolate. Similar phylogenetic topology was recognized by PCA which clearly divided the populations into three distinct lineages (Fig. 4).

Bayesian clustering analyses with high-density SNP markers inferred the optimal number of subgroups (K) by using Evanno's $\Delta K$ method and the likelihood at each K (Evanno et al. 2005). As the most parsimonious partitioning of individuals within species, the estimates for $\Delta K$ fall into two distinct patterns: the highest $\Delta K$ values at $K = 3$, followed by secondary values $K = 4$ (Fig. 5a, b). Nevertheless, these analyses achieve the lowest likelihood at $K = 3$ and the ancestral barplots showed that the plateau groups and the northern were grouped into one cluster, which was not consistent with previous phylogenetic analysis. The barplot at $K = 4$ reasonably reflected the complex relationship that individuals can be attributed to 4 genetic clusters, identifying to the ancestral components, which they denoted as northern group (orange and pink), plateau group (green and pink), and the southern group (blue and orange) with no admixed genotype (Fig. 5c).

The analysis of population splitting patterns, the direction and magnitude of migration was conducted by TreeMix based on our SNP data (Fig. 6), and the tree rooted using SNPs variants of *H. quoyi*, one sister genus of *H. intermedius*, which revealed two key results: First, within two main clades, GM, SS, and DM populations, comprising of the southern groups, were well separated from the remaining sites, while the Plateau and northern groups were admixtures with strong support. 7 historical migration events were added to the tree sequentially. For all SNPs, the migration events were inferred for several pairs of taxa, including gene flow from the whole five population except southern group into CM population (0.320), from northern and plateau groups into southern group (0.440), and from the northern group into Yangtze River System (SZ and CM; 0.345), as well as another 4 weak migration events (0.004 ~ 0.055). Noteworthily, the total direction of migration was non-random, moving from higher to lower altitude and higher to lower latitude.

**Demographic history of ancestry populations**

To assess deviations from neutrality across whole groups, the results for the overall SNPs dataset of *H. intermedius* showed that the significant positive value (mean Tajima's D = 2.773; Fig. 7), which revealed that the medium-frequency loci were polymorphic and dominant in the genomes, indicating effects of balancing selection or population subdivision, which maintained the higher genetic variance and
contributed to local adaptation. There were fewer negative Tajima's D values and low-frequency loci, proving no strong positive selection.

Demographic history analysis based on posterior probabilities was conducted by ABC analysis to interpret the history of the genetic clusters detected in the phylogeography. Prior to this, we hypothesized that 4 scenarios were consistent with the demography of *H. intermedius* (Fig. 8c). The analysis results indicated that support for scenario 2 reached the highest level with a probability of 0.34 in direct approach (Fig. 8a) and 0.7577 in logistic approach (posterior probability, Table 4, Fig. 8b), whereas other models exhibited lower probability. Based on the demographic model, the scaled model values of effective population sizes varied greatly among regions: the population groups of PL had large population sizes (6.58e + 3), whereas the northern populations (NP) were smaller (6.53e + 3), and the southern population (SP) sizes was only 6.40e + 2(Table.4).

**Morphology and phenotype-environment association**

Analysis distinguished pronounced changes of shape associated with the bases of the dorsal and anal fins. Furthermore, the numbers of dorsal and anal fin rays showed a significant difference between the south groups and the remaining groups (*P* < 0.05), indicating the appearance of dramatic phenotypic difference which coincided with the genomic differentiation.

To explore the relationships between morphological traits and environmental heterogeneity, all the measured phenotypic and environmental parameters were conducted by RDA analysis (the redundancy analysis). After the Monte Carlo tests and the forward selection from RDA, water transparency, water temperature, and daily temperature range were selected as the key environmental variables which significantly influence morphological traits of *H. intermedius* (*P* < 0.05; Fig. 9a, b), but dissolved oxygen content also contribute to certainly influence on the traits with no significant p-value(*p* = 0.094 > 0.05). Ordination diagram of RDA displayed that the first two axes cumulatively explained 86.71% of the variance of the phenotype-environment relationship (Axes 1 and 2 modeled 53.7% and 33% of phenotype data respectively). Of the four variables, only water temperature was highly positively paralleled to the axis 1 of redundancy analysis, whereas daily temperature range, DO (dissolved oxygen) and transparency was negatively correlated with both axes. Axis 1 explained 53.7% of the total variability on the phenotype-environment relationship, while the second axis explained 33% of this variance in morphological traits (Table 5).

One main gradient along axis 1 was observed, linking to morphological data (Fig. 9a) and sites (Fig. 9b) being separated from three groups. The phenotypic traits such as dorsal and anal fins related to DTR and DO. While the UJAW(upper jaw length) and many other traits generally preferred warmer water, BW(body weight) and SL(standard length) were correlative with high transparency (Fig. 9a). In terms of sites, the plateau group performed a strong relationship with dissolved oxygen content, daily temperature range, while the southern group distribution was associated with the water temperature and JN population was admixed with the plateau group.
Discussion

Population structure and phylogeographic relationship

Significant spatial genetic structure was evident in ddRAD-seq analyses and reflected the geographic arrangement of the mainland, intraspecific lineage of species is often reported to be differentiated along latitude (north-south axis) and altitude (plateau-plain axis) (Domrös and Peng 1988; Jiang et al. 2018). Overall, our genetic results accorded with a classic geographical scenario that these lineages among different regions have significantly differentiated genetically. We identified 3 unique genetic groups of *H. intermedius* in whole mainland, which conformed to their present geography and were demonstrated by PCA and phylogenetic analyses. On the basis of these phylogeographic results, we hypothesized and demonstrated migration and demographic history. Strong migrations within groups still existed, although some populations are far apart from each other, such as JN and SZ, SS and DM. These findings imply that *H. intermedius* has a much larger range of gene flow than previously thought. In contrast to marine environments, freshwater fishes are typically isolated by various geographical factors and thus often have lower levels of gene flow to the limited space (Bloom et al. 2013; Puebla 2009). Considering geographical barriers among the river systems and each group separated by > 1000 km, we were supposed to observe that pairwise population genetic differentiation was complicated as might be expected.

It is generally supposed that long-distance dispersal events and the geographical distance isolation promote increases in genetic differentiation and decreases in genetic diversity (Wright 1943; Saastamoinen et al. 2017). However, we conclude that our original hypothesis of distance sequential diversification for differentiation was not supported. In this research, the mantel test results were insignificantly correlated with the geographical distance between the different populations ($R^2 = 0.2011$, $p = 0.096 > 0.05$) and there appears no evidence of the impact of IBD (Isolation by distance model), running counter-response to the conclusions of previous studies. Taken together, the complex phylogeographic relationship consistent with ABC results suggested that *H. intermedius* diversified throughout the system and experienced subsequent range shifts that led to the current biogeographic distribution and complicated demographic history.

Genetic differentiation and cryptic lineages

In general, the levels of genetic diversification were observed in relation to spatial variation, especially along latitude. When compared with genetic variants, we observed that the number of polymorphic loci was present in the northern and plateau populations, significantly higher than the southern from Pearl River Systems. Meanwhile, the degree of polymorphism within populations at the genome level ranged from 0.1937 to 0.2376, in which the southern populations almost exhibited lower levels, in conjunction with the number of null loci in genome library was dominant. Consequently, strong geographical clustering of populations was observed, corresponding with our AMOVA estimation (variance among groups occupying up 48.8%). Furthermore, pairwise $F_{ST}$ analysis revealed significant genetic
differentiation between the native and the introduced populations, but also detected the presence of high
differentiation for the southern group, even regard as cryptic subspecies ($F_{ST} > 0.25$, $Nm < 0.5$), within the
whole mainland. It is proved that the strong divergence of *H. intermedius* has appeared along north-south
axis based on genetic analysis whereas Collette and Su (1986) resolved phylogeographic relationships
within using morphological data. In the light of these causes, and given the presence of river barriers for
populations’ dispersion, we hypothesize that these lineages may represent genetically incipient species.
Such differentiation could potentially be caused by selection or local adaptation and was reported in the
sister taxon (Lim et al. 2016). The south of China mainland has been described as genetically
differentiation and more diversification when compared to northern mega-regions related to the climate
difference in the northern hemisphere (Zhang et al. 2012; Wen et al. 2004), a pattern referred to as
southern richness to northern purity phenomenon (Hewitt 1996; 2000). Such conflict, known as “northern
genetic richness and southern purity”, of genome variation decreasing from north to south alternatively
interpreted as a result of northern origin of *H. intermedius* (Fig. 8c), addressed in several other animals’
investigations (Nuñez et al. 2011).

The heterozygosity of the genome also determines the degree of inbreeding and genetic diversity of the
population, in which the observed heterozygosity reflect the true levels of genetic variation within
populations (Courtnée, et al. 2019). We found that Ho and He among whole populations were fairly low
on a genomewide scale, which agrees with the previous nucleotide diversity. Meanwhile, it was observed
that positive values of population differentiation between individuals among sampled locations ($F_{IS} > 0$) in
*H. intermedius* as the consequences of heterozygote deficit and resulting higher $F_{IS}$ values are expected in
members of the more mobile sex because FIS represents the average deviation of the population's
genotypic proportions from Hardy-Weinberg expectations. A considerable mixture of residents and
immigrant individuals in the population will exhibit heterozygote deficiency, and thus have a positive FIS
(Raymond et al. 1995; Wright. 1950; Lagisz et al. 2010), coping with the detrimental effect of the decrease
in fitness. However, it is worth noting that positive $F_{IS}$ is not equivalent to necessary inbreeding, and it is
presumably related to population structure effect, or allele problems such as loci is under selection
pressure (either natural selection or breeding selection). Genomic variation plays a potentially
complicated role in population adaptive differentiation and genetic polymorphism of *H. intermedius*,
which is known to be an important predictor of plastic and adaptive potential.

**Origins of the early introduced populations**

The phylogenetic pattern observed from numerous SNPs at a genome-wide scale manifested the extreme
hydrology of the Yunnan-Guizhou Plateau appears to have acted a central role in genetic differentiation
and structuring diversification of Plateau group following initial colonization. Specifically, individuals
from the northern group serve as a lineage that is resolved as native populations to the genetically
proximate the introduced group, which was evident by phylogeographic analysis and demographic
inference and demonstrated the history record (Zhu and Chen 1989). Clear evidence of admixture as
reflected in phylogenetic tree is present in whole genomes, although the northern are more distant from
the Plateau Lakes and isolated by various continental barriers. According to the historical records, it was
shown that *H. intermedius* was first introduced into the Plateau Lakes in Yunnan Province under no planned programs in the 1980s, without a clear recording of migration routes during that time (Xiong et al. 2006). The introduction routes of Plateau populations could be inferred from several resource investigation and most of them might originate from Jiangsu provinces (Chen and Zhu 1989; Xiong et al. 2008), which overlap geographically with Yangtze River Basin.

Genetic differentiation among the native populations also provided critical clues to track the origins of the Plateau introduced populations (Cornuet et al. 1999; Waples and Gaggiotti 2006; Shen et al. 2019). Both statistical genetic differentiation and an obvious geographic pattern to the population distribution in the diverse aquatic environment were expectedly identified in a background of migration. Introduced populations in the Plateau Lakes were implemented into an independent genetic cluster, though few individuals were an admixture with the northern group. Noteworthily, we identified the genetic divergence of the Plateau group and the northern group did not reach a high level, which is rather unexpected. The Plateau Lakes is geographically more adjacent to the Pearl River System than the Yangtze River System, which was hypothesized reasonably that exited one or more introgressive hybridization events, but its scenario was the lowest support among 4 competing scenarios by ABC analysis. Both pairwise $F_{ST}$ and phylogenetic analyses exhibited that all introduced populations were expected primarily to shape patterns across recent divergence, which was also supported by the inference of ancestral alleles. Similar conclusions are drawn from other fish researches, such as salmonid (Narum et al. 2017) and grass carp (Shen et al. 2019).

**Local diversification in heterogeneous environments**

The combination of morphological characters and genomic data together with environmental heterogeneity elucidates a scenario of diversification in *H. intermedius*. Besides genetic differentiation at the molecular level, we also found that the phenotypic plasticity of morphological traits: dorsal, anal fins rays and traits relative to the bases of the dorsal and anal fins perform local differences in the southern populations, indicating that local diversification along the north-south axis. This morphological polymorphism is in congruence with phylogenetic patterns that initial position is the southern group outside the remaining clades (Fig. 3) and its relatively old age is correspondent to splitting time and branch length while no large phenotypic difference occurs between plateau colonization and native population. Actually, phenotypic differentiation among watersheds detected in the research appears to be driven by selection pressure related to spatial heterogeneity in transparency, water temperature, daily temperature range and DO, which have been shown to be important determinants of phylogeography structure in many other species (Selleslagh et al. 2008; Guo et al. 2015). Despite, these associations with morphological polymorphism and ecological environment do not confirm causality necessarily, as the relatedness may derive from interaction or some unrecognized factor that is strongly associated with one detected parameter (Guo et al. 2016), suggesting that increased sample sizes or greater numbers of species may be needed to detect the effect. Interestingly, sampling sites within a geographical region were located at different altitudes and the plateau group perform strong relationship with dissolved oxygen content, daily temperature range, which reflect the plateau climate intensity on temperature.
fluctuation and oxygen deficit (MacKinnon et al. 1996; Fan et al. 2010), while the distribution of southern group was associated with the water temperature as result of subtropical weather in the Pearl River (Zhou et al. 2019).

The survey highlights the fact from previous empirical studies that diverse selection pressure stemming from environmental heterogeneity, even species facing considerable migration in limited region and microscales, can introduce local divergence and adaptation (Mathias et al. 2001; Michel et al. 2010; Via 2012). Genetically, selective pressures identification during population introduction and adaptation in heterogeneous environments is pivotal to unveil the mechanism of this evolutionary process. With positive Tajima's D values, it revealed that the medium-frequency loci were polymorphic and dominant in the genomes, indicating strong effects of balancing selection or population subdivision (Tajima 1989; Simonsen et al. 1995), which prompt species maintained the higher genetic variance and have capacity to survive and adapt in diverse ecology eventually. We speculate that significant genetic drift and selection and independent colonization events occurred to southern and plateau groups, as result of higher genetic diversity at specific loci is maintained by putative balancing selection. There were fewer negative Tajima's D values and low-frequency loci, proving no strong positive selection. The results lacking the reference genome restricted herein to the non-model species and selective region can't well be defined. However, local divergence and evolution will emerge when strength of divergent selection overrides random genetic drift and homogenizing effect of gene flow among populations (Leinomen et al. 2013). Nevertheless, direct observation of selective regions across genome for non-model species was conducted hardly and remained speculative; taken together, balancing selection could promote the survival of isolated populations in heterogeneous environments and intensified that lineage-specific patterns of selection may potentially intensify boundaries of local populations, which determined its phenotypic polymorphism, plasticity and wide adaptability as a good material for study.

Conclusions

Our phenotypic and environmental results indicate that hydrological barriers and abiotic factors (water temperature, daily temperature range and transparency) perform a critical function in molding phylogeography and diversification on large scales. Genomewide SNP data broaden insights into the complex spatial genetic structure (two clades and 3 clusters) and demographic history of *H. intermedius* in China Mainland, especially uncovering genetic differentiation from three heterogeneous hydrobiological regions and serving as an essential source of genetic polymorphisms for divergence and adaptation. Furthermore, we also demonstrated complex demography that the Plateau populations of halfbeak originated from the northern groups with strikingly pairwise FST increasing, whereas the significantly high level of divergence early occurred in both genotype and phenotype to the southern group, which was proposed on an isolated lineage that may develop to incipient species. Furthermore, we demonstrated hypothesis that *H. intermedius* underwent balancing selection which is also likely indispensable to survival during adaptation to new habitats, respectively. Although IBD was not detected among remote groups, the latitude and altitude distribution of *H. intermedius* likely has an adaptive genetic basis. We offer robust validation that variable migratory behaviors, genetic diversity, population
structure of *H. intermedius* and their adaptive divergence in heterogeneous environment arisen from the interaction of balancing selection and strong gene flow at microscales, as well as strong impact of abiotic factors.

**Methods**

**Study sites and sample collection**

The total specimens were collected from fish docks or local markets in 8 sites of China Mainland across three regions with different climatic conditions, from May 2017 to December 2018, including Danchi (DC), Fu Xianhu (FXH) in the plateau waters of Yunnan Province, Sanshui(SS), Gaoming(GM), DM(Doumen) in the Pearl River, Chongming(CM) near the estuary, Suzhou(SZ) in the Taihu Lake, Jining(JN) in Weishanhu Lake (Fig. 1). They are separated and came from heterogeneous aquatic habitats (Table.S1). Spatial statistic was performed with ArcGIS 10.0 software (Environmental Systems Research Institute, Inc). A total of 156 living or death *Hyporhamphus intermedius* individuals (approximately 20 per population) were selected for processing to save lives and ensure sufficient sample size, and *Hyporhamphus quoyi* (n = 3) were collected as outgroup. The living samples were transferred into 10L aquaria which dissolved with a high concentration of 150 mg/L MS-222 (Beijing Green Hengxing Biological Technology Co., Beijing, China). All individuals appeared deep anesthesia or death behaviors (lost balance and sank into the bottom of aquaria) during 1–2 min, and the surgery and sampling were made after a respiratory arrest. We measured their standard length (SL), weighed them on an electronic balance (BW), and then took a lateral photograph of the left-hand side of body shape to measure 10 morphological traits and perform geometric-morphometric analyses by Digimizer v.4.5.1 (http://www.digimizer.com) (Fig. 2). All dorsal and anal fin-rays were counted under a dissecting microscope. Then, the fins or muscles were sampled and transferred to the laboratory in liquid nitrogen and then stored at -80 °C for further experiments.

**RAD data processing**

Genomic DNA was extracted from fin or muscles of each individual from each of 8 populations using CTAB (Doyle and Doyle 1987) method. Then Double-digest restriction-associated DNA (ddRAD) was conducted as the protocol described by Peterson et al. (2012) and libraries were prepared using 500 ng of DNA per sample. Genomic DNA was digested with the restriction enzymes *EcoR1* and *BfaI* at 37 °C for 5 h followed by the ligation of the original Illumina adapter sequences and unique 8 bp barcodes that were used for library preparation of each individual to distinguish itself. Pools of individuals run via gel-electrophoresis on 1.5% agarose gels, where 220 ~ 450 bp long fragments were size-selected, then purified using a Zymoclean Gel DNA recovery kit. Each pool was incubated at 14 PCR cycles in 25 µl reactions which contain 5 µl 5 × Reaction buffer, 5 µl 5 × High GC enhancer, 0.25 µl Q5 polymerase, 4 µl of library DNA and a unique indexing primer for each pool corresponding to the standard Illumina multiplexed sequencing protocol.
The ligation products were amplified in PCRs using a Veriti 96-well thermal cycler (Life Technologies) and the protocol consisted of initial denaturation at 98 °C for 30 s, 14 cycles (98 °C for 15 s, 65 °C for 30 s and 72 °C for 30 s), followed by a final step at 72 °C for 5 min. DNA libraries were quantified using a 2100 Bioanalyzer (Agilent Technologies). Finally, pools were combined in equimolar concentration to form a single genomic library and sequenced on an Illumina HiSeq 2500 using 150 bp pair-end reads.

**Population genetic and statistical analyses**

All raw sequences from the Illumina HiSeq lanes were checked for initial quality using FastQC, then quality filtering, assembly, SNP discovery of per individual was conducted for bioinformatic analysis. The *population* script was applied to calculate population genetic diversity parameters such as nucleotide diversity (π), expected and observed heterozygosity (He & Ho), and F-statistics including genetic differentiation (F<sub>ST</sub>) and average genetic differentiation between individuals within their sampling locations (F<sub>IS</sub>) were estimated with Stacks v1.46 (Catchen et al. 2013). We exported SNPs from the population's module in Stacks with the -write_single_snp option. A hierarchical analysis of molecular variance (AMOVA) was used to estimate source of variation at three hierarchical subdivisions (among groups, among populations within a group, and within the populations) by Arlequin v.3.5 (Excoffier and Lischer 2010). Nm (gene flow) was calculated by (1-F<sub>ST</sub>)/4F<sub>ST</sub>. To further evaluate isolation-by-distance (IBD), the pattern of population differentiation was examined using Mantel tests with the R v3.2.3 (Core and Team 2012) using the “vegan” packages (Oksanen et al. 2007). The genetic distance was measured using F<sub>ST</sub>/(1-F<sub>ST</sub>), while the geographical distance was estimated from the geographic coordinates of sampling sites.

Bayesian clustering was performed in STRUCTURE 2.3.4 (Pritchard et al. 2000) with the admixture model and correlated allele frequencies, and this analysis assumes that loci are unlinked. We ran STRUCTURE using a burn-in period of 20,000 generations and 80,000 Markov Chain Monte Carlo (MCMC) steps for different values of K ranging from 1 to 10 with 5 replicates for each K value. The optimal K was determined using STRUCTURE HARVESTER (Jakobsson and Rosenberg 2007; Earl and vonHoldt 2012) using the ΔK method (Evanno et al. 2005).

The relationship between geographic coordinates and genetic structure was also identified in principal component analysis (PCA) using the R package “adegenet” (Jombart 2008, Jiang et al. 2018.), which was defined as the first two components of the PCA. Phylogenetic relationships of *H. intermedius* among locations were constructed using maximum likelihood (Neighbor-Joining tree) methods by bootstrapping over loci for 1000 times. We choose *Hyporhamphus quoyi* (its closely related species as outgroups) to root the tree. The results were performed on iTOL (https://itol.embl.de/). Then, the maximum-likelihood tree to infer gene flow and the historical admixture events amongst populations was generated by TreeMix (Pickrell and Pritchard 2012) on SNPs data. We rooted the population graph with *Hyporhamphus quoyi*, and tested the significance of migration events with various migration rates to identify genetic drift amongst populations.
In order to investigate past demographic expansion, and neutrality tests, Tajima's D values (Tajima 1989) was calculated with the 20000 bp sliding windows. Historical changes in population size on the basis of approximate Bayesian computation (ABC) methods were conducted in DIYABC v2.1.0 from a subset of the genetic data (Cornuet et al. 2014). We eliminated variants with low minor allele frequencies (MAFs < 0.1) or the loci that completely miss in at least one regional population. 400 thousand simulated datasets including four different demographic competing scenarios were drawn from prior distributions with the consequences of population genetic structure. All 8 population were divided in advance into the following three regional groups based on the morphometric and population structure: Pop1—he Southern populations (ST including the SS, GM, DM); Pop2—the Northern populations (NT including the SZ, CM, JN); Pop 3—the Plateau populations (PL including the DC, FXH ). To compare 4 competing scenarios and find the most plausible one, ABC procedure was employed to infer the most probable evolutionary history. We characterized 4 scenarios by demographic and historical parameters that were marked as the number of generations back in time (t1, t2). The effective population sizes of three regional groups and the ancestral population were represented by ST, PL, NT, NA respectively. All uniform was chosen for prior distributions and effective population sizes be constant in time. In short, both DIYABC's direct and logistic regression to assess posterior probabilities were chosen for comparison to enable a ranking of scenarios. Finally, each scenario posterior probability was computed based on 10% simulated data closest to the observed data using a logistic regression procedure. The summary statistics were scaled by the mean effective population size of the present four populations because the mutation rates of the SNPs were unclear.

**Phenotype-environment association**

To evaluate associations between the phenotype and environment in driving spatial genetic differentiation, the morphological data were relevant to environmental variables using redundancy analyses (RDA) (Rao, 1964), which utilize matrices of dependent and independent (explanatory) variables. The dependent matrix consisted of the 13 standardized phenotypic variables (SL, Body Weight, Dorsal fin, Anal fin, PAL, PL, UJAW, ED, HL, CL, CD, BD, CFL), while 15 explanatory abiotic factors were regarded as the explanatory variables in one explanatory matrix to assess the phenotypic variation (Table.S2). Climate data was collected from The China Meteorological Data Service Center (http://data.cma.cn/); Environmental variable measurements were carried out with an YSIImeter (computermodule: 650 MDS, sonde: 6920; YSI Inc, Yellow Springs, OH). RDA results in a series of orthogonal axes, including linear combinations of the explanatory variables that best explain the variation in a response matrix. (Andersen et al. 2011; Andersen et al. 2011). Additionally, repeated analyses of the average data set of 8 populations could acquire quantitative information about the impact on populations. Monte Carlo permutation analysis simulation and interactive forward selection of variables were performed using the significance (p < 0.1) by Canoco v.5.0 (Lepš and Šmilauer 2003).

**Abbreviations**

AMOVA
analysis of molecular variance
DdRAD
double digest restriction-site associated DNA
DO
dissolved oxygen
DTG
daily temperature range
IBD
Isolation by distance model
MAF
minor allele frequency
SNP
single nucleotide polymorphism
RDA
redundancy analyses
UJAW
upper jaw length

Declarations

Ethics approval and consent to participate

This study was performed in strict accordance with the Guide for the Institutional Animal Care and Use Commission, Sun Yat-sen University (IACUC, SYSU), Guangzhou, China. The Affidavit of Approval of Animal Use Protocol within the project “Population Adaptive Differentiation and Genetic Polymorphism of Hyporhamphus intermedius in Heterogeneous Environments” was approved by the committee (IACUC, SYSU). During the sampling, H. intermedius was obtained from living or dead individuals. As H. intermedius is not endangered or protected, we don’t need any permission to collect the animals used in our study. All surgery for living samples was performed under MS-222 anesthesia, and every effort was made to minimize suffering.

Consent for publication

Not applicable.

Availability of data and material

DNA SNP data and other analysis data have been deposited at supplementary data.

Competing interests

The authors declare that they have no competing interests.
Funding

This research was supported by the Guangdong Basic and Applied Basic Research Foundation (No. 2019B1515120065), and by the NSFC-Guangdong Joint Fund (No. U1901209). None of the funding bodies participated in the design of the study, collection, analysis, and interpretation of data, or writing the manuscript.

Authors' contributions

G L and G W conceived the study; G L and G W designed the experiments; G W, H L, S B collected the samples; G W did the experiments; G W & H L, D G prepared the figures and tables; G W & G L drafted the work or revised it critically for important content; X C, X Z, S L, X L, Y S, H Y contribute to the draft and revise of manuscript; G L finally approve the version to be published, all authors have read and approved the manuscript.

Acknowledgments

We would like to thank the Shanghai Personal Biotechnology Co., Ltd for sequence and analyses. We also wanted to thank local fishermen for their help in surveys and fish sampling.

References

1. Alter SE, Munshi-South J, Stiassny MLJ. Genome-wide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River[J]. Mol Ecol. 2016;26(5):1401.
2. Andersen R, et al. Environmental control and spatial structures in peatland vegetation[J]. Journal of Vegetation Science. 2011;22(5):878–90.
3. Andrews KR, et al. Harnessing the power of RAD-seq for ecological and evolutionary genomics[J]. Nat Rev Genet. 2016;17(2):81–92.
4. Avise JC. Phylogeography: The History and Formation of Species[M]. Harvard University Press, 2000.
5. Bloom DD, et al. Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of new world silversides (atherinopsidae)[J]. Evolution. 2013;67(7):2040–57.
6. Brawand D, et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 2014;513:375–81.
7. Catchen, et al. Stacks: an analysis tool set for population genomics[J]. Mol Ecol. 2013;22(11):3124–40.
8. Collette BB. Su. The halfbeaks (Pisces, Beloniformes, Hemiramphidae) of the far East[J]. Proceedings of the Academy of Natural Sciences of Philadelphia. 1986;138(1):250–302.
9. Core R, et al. R: A Language and Environment for Statistical Computing[J]. Computing. 2011;14:12–21.
10. Cornuet J-M, et al. Inferring population history with diy abc: a user-friendly approach to approximate bayesian[J]. Bioinformatics. 2008;24(23):2713–9.
11. Cornuet J-M. New methods employing multi locus genotypes to select or exclude populations as origins of individuals[J]. Genetics. 1999;153:1989–2000.
12. Courtneë, et al. Low genetic diversity and shallow population structure in the endangered vulture, Gyps coprotheres[J]. Scientific reports. 2019;9:5536.
13. Cui YD, Liu XQ, Wang HZ. Macrozoobenthic community of Fuxian Lake, the deepest lake of southwest China. Limnologica-Ecology Management of Inland Waters. 2008;38(2):116–25.
14. Domrös M, Peng G. The climate of China. Berlin Heidelberg: Springer-Verlag; 1988.
15. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue[J]. Phytochem Bull. 1987;19:11–5.
16. Earl DA, Vonholdt BM. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method[J]. Conservation Genetics Resources. 2012;4(2):359–61.
17. Evanno G, et al. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study[J]. Mol Ecol. 2005;14:2611–20.
18. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows[J]. Molecular Ecology Resources. 2010;10(3):564–7.
19. Fan Z, et al. Spatial and temporal temperature trends on the Yunnan Plateau (Southwest China) during 1961–2004[J]. Int J Climatol. 2011;31(14):2078–90.
20. Fraser DJ, et al. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis[J]. Heredity. 2011;106(3):404–20.
21. George FT. The Ecology of Adaptive Radiation[J]. Heredity. 2000;86(6):749–50.
22. Gamboa M, Watanabe K. Genome-wide signatures of local adaptation among seven stoneflies species along a nationwide latitudinal gradient in Japan. BMC Genom. 2019;20,84. https://doi.org/10.1186/s12864-019-5453-3.
23. Guo B, et al. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks[J]. BMC Biol. 2015;13(1):19.
24. Guo B, et al. Population genomic evidence for adaptive differentiation in the Baltic Sea herring[J]. Molecular Ecology. 2016;25(12):2833–52.
25. Hewitt GM. Some genetic consequences of ice ages, and their role in divergence and speciation[J]. Biol J Lin Soc. 1996;58:247–76.
26. Hewitt GM. The genetic legacy of the quaternary ice ages[J]. Nature. 2000;405:907–13.
27. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure[J]. Bioinformatics. 2007;23(14):1801–6.
28. Jiang W, et al. Genome assembly for a Yunnan-Guizhou plateau “3E” fish, Anabarilius grahami (regan), and its evolutionary and genetic applications. *Frontiers in Genetics*. 9:614.

29. Jiang XL, et al. 2018. Land bridges in the Pleistocene contributed to flora assembly on the continental islands of south china: insights from the evolutionary history of *Quercus championii*. *Molecular Phylogenetics & Evolution*. 2018;132: 36–45.

30. Jombart T. Adegenet. A R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403–5.

31. Jones FC, et al. The genomic basis of adaptive evolution in three-spine sticklebacks. *Nature*. 2012;484:55–61.

32. Kyriakis, et al. Scanning of genetic variants and genetic mapping of phenotypic traits in gilthead sea bream through ddRAD sequencing. *Frontiers in Genetics*. 2019;10:675.

33. Lagisz M, et al. Genetic population structure of the ground beetle, *Pterostichus oblongopunctatus*, inhabiting a fragmented and polluted landscape: Evidence for sex-biased dispersal. *Journal of Insect Science*. 2010;10:105.

34. Leinonen T, et al. QST-FST comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nat Rev Genet*. 2013;14:179–90.

35. Lepš J, Šmilauer P. Multivariate Analysis of Ecological Data using CANOCO: Regression methods. Cambridge: Cambridge University Press; 2003.

36. Li S, Zhang C. Chinese zoology. Beijing: Science Press; 2011.

37. Mackinnon J, et al. A biodiversity review of China. Hong Kong: World Wide Fund for Nature; 1996.

38. Maroso F, et al. Species identification of two closely exploited flatfish, turbot (*Scophthalmus maximus*) and brill (*Scophthalmus rhombus*), using a ddRADseq genomic approach. *Aquatic Conservation: Marine Freshwater Ecosystems*. 2018;28:1253–60.

39. Mathias A, Kisdi E, Olivieri I. Divergent Evolution of Dispersal in a Heterogeneous Landscape. *Evolution*. 2001;55(2):246–59.

40. Michel AP, et al. Widespread genomic divergence during sympatric speciation. *Proc Natl Acad Sci USA*. 2010;107:9724–9.

41. Narum, et al. Genomic patterns of diversity and divergence of two introduced salmonid species in Patagonia, South America. *Evol Appl*. 2017;10(4):402–16.

42. Nosil P, et al. Tipping points in the dynamics of speciation. *Nature Ecology Evolution*. 2017;1(2):0001.

43. Nosil P, Funk DJ, Ortiz-Barrientos D. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. 2009;18:375–402.

44. Nuñez JJ, et al. Amphibian phylogeography in the Antipodes: refugia and postglacial colonization explain mitochondrial haplotype distribution in the Patagonian frog *Eupsophus calcaratus* (Cycloramphidae). *Mol Phylogenet Evol*. 2011;58:343–52.
45. Oksanen J, et al. vegan: community ecology package. Available at http:// cran.r-project.org/web/packages/vegan/index.html. 2013.
46. Paetkau DR, Slade Burden M, Estoup A. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power[J]. Mol Ecol. 2004;13:55–65.
47. Peterson BK, et al. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species[J]. PLoS One. 2012;7:E37135.
48. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data[J]. PLoS Genet. 2012;8(11):e1002967.
49. Pritchard JK, et al. Association mapping in structured populations. Am J Hum Genet. 2000;67(1):170–81.
50. Puebla O. Ecological speciation in marine v. freshwater fishes[J]. J Fish Biol. 2009;75(5):960–96.
51. Rajkov J, et al. Adaptive phenotypic plasticity contributes to divergence between lake and river populations of an East African cichlid fish[J]. Ecology Evolution. 2018;8:7323–33.
52. Rao CR. The use and interpretation of principal component analysis in applied research[J]. Sankhya A. 1964;26:329–58.
53. Riginos C, Liggins L. Seascape genetics: populations, individuals, and genes marooned and adrift[J]. Geography Compass. 2013;7(3):197–216.
54. Rousset F, Raymond M. Testing heterozygote excess and deficiency[J]. Genetics. 1995;140(4):1413–9.
55. Saenz-Agudelo P, et al. Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes[J]. Mol Ecol. 2015;24(24):6241–55.
56. Savolainen O, Lascoux, Merilä J. Ecological genomics of local adaptation[J]. Nat Rev Genet. 2013;14(11):807–20.
57. Selleslagh J, et al. Fish composition and assemblage structure in three Eastern English Channel macrotidal estuaries: A comparison with other French estuaries[J]. Estuarine Coastal Shelf Science. 2009;81(2):149–59.
58. Shen Y, et al. Population structure, demographic history and local adaptation of the grass carp[J]. BMC Genomics. 2019;20:467.
59. Simonsen KL, Churchill GA, Aquadro CF. Properties of statistical tests of neutrality for DNA polymorphism data[J]. Genetics. 1995;141(1):413–29.
60. Sihai W, et al. Six decades of changes in vascular hydrophyte and fish species in three plateau lakes in Yunnan, China[J]. 2013;22:3197–221.
61. Slatkin M. Quantitative Genetics of Heterochrony[J]. Evolution. 1987;41(4):799–811.
62. Stoddart DR, et al. Evolution of reefs and islands, northern great barrier reef: synthesis and interpretation[J]. Philosophical Transactions of the Royal Society B Biological Sciences. 1978;284(999):149–59.
63. Sultan SE, Spencer HG. Metapopulation structure favors plasticity over local adaptation[J]. American Naturalist. 2002;160:271–83.

64. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95.

65. Tang Q, et al. Geography of China. In: Tang Q, Ge Q, editors. Atlas of Environmental Risks Facing China Under Climate Change. IHDP/Future Earth-Integrated Risk Governance Project Series. Singapore: Springer; 2018.

66. Via S. Sympatric speciation in animals: the ugly duckling grows up[J]. Trends of Ecology and Evolution. 2001;16.

67. Via S. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow[J]. Philosophical Transactions of the Royal Society B: Biological Sciences. 2012;367:451–60.

68. Wang C, Zöllner S, Rosenberg NA. A quantitative comparison of the similarity between genes and geography in worldwide human populations[J]. PLoS Genetic. 2012;8(8):e1002886.

69. Waples RS, Gaggiotti O. What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology. 2006;15:1419–39.

70. Wiens JJ. The niche, biogeography and species interactions[J]. Philosophical Transactions of the Royal Society B: Biological Sciences. 2011;366(1576):2336–50.

71. Wen B, et al. Genetic evidence supports demic diffusion of Han culture[J]. Nature. 2004;4317006:302–5.

72. Wright S. Isolation by distance[J]. Genetics. 1943;28(2):114.

73. Wright S. Genetic structure of populations[J]. British Medical Journal. 1950;2(4669):36.

74. Xiong F, et al. Status and Changes of Fish Resources in Fuxian Lake, Yunnan [J]. Journal of Lake Science. 2006;(03): 305–311.

75. Xiong F, Li W, Pan J. Present status of alien fishes and analysis of relative problems in Lake Fuxian in Yunnan Province[J]. Acta Agric Jiangxi. 2008;20:92–4.

76. Zhang R, et al. Comparative phylogeography of two widespread magpies: importance of habitat preference and breeding behavior on genetic structure in China[J]. Molecular Phylogenetics and Evolution. 2012;65:562–72.

77. Zhou L, et al. Fish assemblage in the Pearl River Estuary: Spatial-seasonal variation, environmental influence and trends over the past three decades[J]. Journal of Applied Ichthyology. 00:1–12.

78. Zhu XL. Chen YR. Fishes and fishery resources of Yunnan Province[M]. Beijing: Science Press; 1989.

Tables

Table 1 The results of population genetic diversity index
| Population | Num Indv | Ho        | He        | π         | F_{IS}   |
|------------|----------|-----------|-----------|-----------|----------|
| DC         | 16.73    | 0.1644±0.0368 | 0.199±0.0373 | 0.2053±0.0397 | 0.1147   |
| FXH        | 15.36    | 0.1655±0.0368 | 0.199±0.0374 | 0.2059±0.0401 | 0.1113   |
| JN         | 14.31    | 0.1563±0.0304 | 0.2081±0.0359 | 0.2159±0.0386 | 0.1646   |
| SZ         | 17.23    | 0.1676±0.0323 | 0.2169±0.0349 | 0.2237±0.0372 | 0.1594   |
| CM         | 17.33    | 0.1654±0.032   | 0.2126±0.0354 | 0.2192±0.0376 | 0.1513   |
| DM         | 16.2     | 0.1358±0.0285 | 0.185±0.0361  | 0.1912±0.0386 | 0.1547   |
| SS         | 17.08    | 0.1436±0.0302 | 0.1878±0.0361 | 0.1938±0.0385 | 0.1413   |
| GM         | 15.77    | 0.1445±0.0312 | 0.1891±0.0367 | 0.1956±0.0393 | 0.1411   |

$ The population genetic statistics includes Num Indv: average number of nucleotide differences; π: nucleotide diversity; He: expected heterozygosity, Ho: observed heterozygosity, F_{IS}: population differentiation between individuals among sampled locations.

Table.2 AMOVA analysis of 8 populations based on 32,744 SNPs sequences

| Source of variation               | d.f. | Sum of squares | Variance components | Percentage of variation |
|-----------------------------------|------|----------------|---------------------|------------------------|
| Among groups                      | 2    | 234340.808     | 1266.63884          | 0.488                  |
| Among populations within groups   | 5    | 10307.797      | 23.03189            | 0.0089                 |
| Within populations                | 149  | 344451.366     | 1306.03383          | 0.5031                 |
| Total                             | 156  | 589099.972     | 2595.70455          |                        |

Table.3 Pairwise F_{ST} (lower diagonal) and Nm (upper diagonal) among 8 populations based on 32,744 SNPs sequences
| Population | DC   | FXH  | JN    | SZ    | CM   | DM   | SS   | GM   |
|------------|------|------|-------|-------|------|------|------|------|
| DC         | 8.841| 3.681| 4.937 | 5.126 | 0.540| 0.535| 0.538|      |
| FXH        | 0.028| 3.459| 4.586 | 4.883 | 0.531| 0.527| 0.527|      |
| JN         | 0.064*| 0.067*| 7.563 | 6.489 | 0.579| 0.558| 0.564|      |
| SZ         | 0.048*| 0.052*| 0.032 | 11.432| 0.595| 0.588| 0.588|      |
| CM         | 0.047*| 0.049*| 0.037 | 0.021 | 0.571| 0.588| 0.565|      |
| DM         | 0.317*| 0.320*| 0.302*| 0.296*| 0.304*| 11.062| 10.620|      |
| SS         | 0.318*| 0.322*| 0.309*| 0.299*| 0.308*| 0.022 | 12.005|      |
| GM         | 0.317*| 0.322*| 0.307*| 0.298*| 0.307*| 0.023 | 0.020 |      |

* P < 0.05

Table 4. Prior distributions of the parameters used in DIYABC

| Parameter                     | Probability distribution | Minimum | Maximum |
|-------------------------------|--------------------------|---------|---------|
| Effective population size     |                          |         |         |
| NA                            | uniform                  | 10      | 100000  |
| NT                            | uniform                  | 10      | 100000  |
| ST                            | uniform                  | 10      | 100000  |
| PL                            | uniform                  | 10      | 100000  |
| Time scale in generations     |                          |         |         |
| t1                            | uniform                  | 10      | 100000  |
| t2                            | uniform                  | 10      | 100000  |
| Admixture                     |                          |         |         |
| ra                            | uniform                  | 0.01    | 0.5     |

Table 5. Redundancy analysis summary statistics of the correlation between fish assemblages and environmental factors.
|                          | Axis 1 | Axis 2 | Axis 3 | Axis 4 |
|--------------------------|--------|--------|--------|--------|
| Eigenvalues              | 0.5370 | 0.3300 | 0.0512 | 0.0115 |
| Species-environment correlations | 0.9853 | 0.9945 | 0.8697 | 0.8818 |
| Cumulative percentage variance |    |        |        |        |
| of species data          | 27.5   | 30.4   | 32.6   | 33.8   |
| of species-environment relation: | 53.70 | 86.71 | 91.83 | 92.98 |
| environmental variables  |        |        |        |        |
| p-value                  |        |        |        |        |
| water temperature        | 0.024  | 36.7   | 36.7   |        |
| daily temperature range  | 0.018  | 33.1   | 33.1   |        |
| transparency             | 0.028  | 16.3   | 16.3   |        |
| DO                       | 0.094  | 7      | 7      |        |

**Figures**
Figure 1

Location of study area and distribution of fish sampling stations (SS, Sanshui; GM, Gaoming; DC, Dianchi; FXH, Fu Xianhu; CM, Chongming; SZ, Suzhou; JN, Jining; DM: Doumen; GS, Zhuhai) (The map was performed by ARCGIS10.2 software, URL: https://desktop.arcgis.com/zh-cn/).
**Figure 2**

The measurement of morphological traits of *H. intermedius*. Abbreviations shown are: SL (standard length), PAL (preanal length), PL (predorsal length), UJAW (upper jaw length), ED (eye diameter), HL (head length), CL (caudal peduncle length), CD (caudal peduncle depth), BD (maximum body depth), CFL (caudal fork length). Adapted from "The halfbeaks (Pisces, Beloniformes, Hemiramphidae) of the far East" by Collette BB. and Su J. 1986; Proceedings of the Academy of Natural Sciences of Philadelphia, 138, p. 259. Copyright 1986 by Academy of Natural Sciences.
Figure 3

The phylogenetic tree of H. intermedius. H. quoyi (GS) is the outgroup with the bootstrap value $> 90\%$. 
Figure 4

The principal components analysis of H. intermedus in the 8 populations distributed in China.
**Figure 5**

Genetic structure analysis of *H. intermedius* among 8 populations. a. Analysis of appropriate K value b. Analysis of delta K value c. modeled-based clustering analysis by STRUCTURE. Vertical bars represent individuals and the heights of different colors point to probabilities assignment each cluster. Different population and the heights of different colors point to probabilities assignment each cluster. Different population was split by black.
Figure 6

Admixture graph constructed with TreeMix using SNP data, 156 individuals of 8 H.intermedius populations, and three other relative individuals. Branch lengths are proportional to the evolutionary change (the drift parameter) and terminal nodes were labeled with population codes (see Table 1). The scale bar represents 10 times the average standard error (s.e.) of the values in the covariance matrix, and the migration weight represents the fraction of ancestry derived from the migration edge. Migration edges are observed between the DM and SS, although gene flow in the other direction is also observed.
Figure 7

SNPs in the unit sliding window corresponds to Tajima's D value. Blueline represents the average of Tajima's D value of SNPs, Tajima's D = 2.773; Red line represents a dividing line, Tajima's D = 0 observed variation equal to expected variation.
Competing scenarios were designed for inferring the demographic histories for ABC analysis. a. Considering the results of STRUCTURE, and phylogenetic analyses, we classified the 8 population into the following three regional groups: Pop1 represents the Southern population; Pop2 the Northern populations: Pop 3 the Plateau populations. b. Model comparison step using the direct approach (Miller et al. 2005). c. Model comparison using the logistic regression approach (Beaumont et al. 2002).
Figure 9

Redundancy analysis diagrams for morphological traits of H. intermedius and environmental variables in sample sites. a. the biplot of morphological scores and environmental variables. b. biplot of site scores and environmental variables and the populations were divided into 3 groups. The red arrows indicated significant factors. DTR=daily temperature range, Water T.=water temperature, DO= dissolved oxygen, Trans. = water transparency. For site codes, refer to Fig.1; For morphological trait codes, refer to Fig.2.
Supplementary Files

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

- SupplementaryMaterial.rar