

Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers

Le Wang^{1,2}  | Shufang Liu³  | Yang Yang¹  | Zining Meng^{1,4}  | Zhimeng Zhuang³ 

¹State Key Laboratory of Biocontrol, Institute of Aquatic Economic Animals and the Guangdong Province Key Laboratory for Aquatic Economic Animals, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

²Molecular Population Genetics Group, Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore City, Singapore

³Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences & Function Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

⁴Southern Laboratory of Ocean Science and Engineering, Zhuhai, China

Correspondence

Zining Meng, State Key Laboratory of Biocontrol, Institute of Aquatic Economic Animals and the Guangdong Province Key Laboratory for Aquatic Economic Animals, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China.
Email: mengzn@mail.sysu.edu.cn

Funding information

Central Public-interest Scientific Institution Basal Research Found, Grant/Award Number: 20603022020005; The Marine S & T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology, Grant/Award Number: 2018SDKJ0501-2; National Natural Science Foundation of China, Grant/Award Number: 31172411

Handling Editor: Cynthia Riginos

Abstract

Understanding the mechanisms underlying heterogeneous genomic divergence is of particular interest in evolutionary biology. Highly differentiated genomic regions, known as genomic islands, often evolve between diverging lineages. These genomic islands may be related to selection promoting adaptation or reproductive isolation. Based on whole genome assembly and genome-wide RAD sequencing in a pair of yellow croakers (genus: *Larimichthys*), we investigated the evolutionary processes shaping genomic landscapes of divergence. Demographic modelling indicated that the two species diverged following a secondary contact scenario, where differential introgression and linked selection were suggested to be involved in heterogeneous genomic divergence. We identified reduced recombination rate in genomic islands and a relatively good conservation of both genetic diversity and recombination landscapes between species, which highlight the roles of linked selection and recombination rate variation in promoting heterogeneous divergence in the common ancestral lineage of the two species. In addition, we found a positive correlation between differentiation (F_{ST}) and absolute sequence divergence (D_{xy}), and elevated D_{xy} in genomic islands, indicating that the genomic landscape of divergence was not shaped by linked selection alone. Restricted gene flow in highly differentiated regions has probably remodelled the landscape of heterogeneous genomic divergence. This study highlights that highly differentiated genomic regions can also arise from a combination of linked selection and differential gene flow in interaction with varying recombination rates.

KEYWORDS

differential introgression, genomic islands, *Larimichthys*, linked selection, low recombination, secondary contact

Le Wang and Shufang Liu authors contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Evolutionary mechanisms can shape the genomic architecture that may be related to adaptation and speciation (Cruickshank & Hahn, 2014; Wolf & Ellegren, 2017). Understanding the genomic architecture can help dissect the evolutionary mechanisms and reconstruct the processes between recently evolved species (De La Torre et al., 2019). Recent studies, particularly with the advancement of chromosomal level genome assembly and high throughput sequencing, have uncovered the genomic landscapes of divergence in a wide range of species (Ellegren et al., 2012; Malinsky et al., 2015; Renaut et al., 2013). It has been established that genomic divergence is usually highly heterogeneous across the genome, with some regions being more differentiated (Feder et al., 2012; Wolf & Ellegren, 2017). These regions of elevated genetic differentiation are widely scattered across the whole genome and preferably enriched in some specific regions, thus termed as "genomic islands" (Harr, 2006; Nosil et al., 2009; Turner et al., 2005). Genomic islands were hypothesized to resist gene flow and thus be involved in adaptation and reproductive isolation (Turner et al., 2005). With more and more studies focused on this evolutionary question, a variety of additional evolutionary processes have been discovered to be responsible for such genomic islands (Han et al., 2017; Ravinet et al., 2017; Wolf & Ellegren, 2017). Thus, the potential roles of genomic islands in speciation have been put into controversy (Cruickshank & Hahn, 2014; Nachman & Payseur, 2012; Turner & Hahn, 2010).

Under the hypothesis of speciation-with-gene-flow, genomic islands evolve as speciation islands, as a result of selection acting on beneficial variants (Feder et al., 2012). These regions are expected to spread with divergence time through divergence hitchhiking, while gene flow homogenizes the remaining parts of the genome (Nosil, 2008; Via, 2012). Similarly, genomic islands can also arise in the secondary contact model (Bierne et al., 2013; Harrison & Larson, 2014). Upon secondary contact, genomic differentiation accumulated in allopatry is expected to be homogenized by gene flow in regions that are not involved in reproductive isolation, while the remaining are kept differentiated (Bierne et al., 2013; Harrison & Larson, 2014). However, genomic islands might also evolve as consequences of lineage sorting in the early stage of speciation (Guerrero & Hahn, 2017; Ma et al., 2018), local selection (Tavares et al., 2018) and/or linked selection, such as background selection and selective sweeps in the ancestral lineage and/or acting independently in the divergent populations (Burri, 2017; Burri et al., 2015), regardless of recent gene flow. Especially, when these evolutionary forces interactively act on low-recombining regions, the effects become more profound (Burri et al., 2015; Ma et al., 2018).

It is challenging to answer whether genomic islands indicate the existence of local genetic barriers to gene flow. First, because in the absence of gene flow or under limited gene flow, selection mechanisms that are not directly related to speciation can cause variation in the rate of lineage sorting along the genome, and hence, incidental islands of F_{ST} in low-recombining regions arise through linked selection (Burri et al., 2015; Cruickshank & Hahn, 2014; Vijay

et al., 2017). These are usually associated with "valleys" of absolute divergence (D_{xy}), because ancestral variation was also probably reduced in low-recombining regions in the ancestral population (Burri et al., 2015; Cruickshank & Hahn, 2014). Second, in the presence of substantial gene flow, genomic islands may have formed due to the effect of local selection in the face of gene flow (i.e. primary divergence), or alternatively due to differential erosion of divergence during secondary contact (Cruickshank & Hahn, 2014; Han et al., 2017; Wolf & Ellegren, 2017). However, D_{xy} is expected to be increased in both primary divergence with gene flow and differential introgression after secondary contact, but reduced in pure allopatric (or post-speciation) divergence because of linked selection removing diversity in low recombining regions in the ancestral population even before divergence starts (Burri et al., 2015; Cruickshank & Hahn, 2014). Therefore, it is necessary to determine the conditions under which genomic islands have formed with respect to gene flow (i.e., no gene flow, continuous gene flow or secondary contact). Under the secondary contact scenario, the overall heterogeneous divergence pattern may result from a mix of heterogeneous divergence during the allopatric phase due to linked selection, followed by heterogeneous introgression during secondary contact due to the presence of genetic barriers (i.e., local adaptations or genetic incompatibilities) that locally reduce gene flow.

The small and large yellow croakers (*Larimichthys polyactis* and *L. crocea*) are the only two sister species within the genus *Larimichthys* of family Sciaenidae. The two species are critically important fishery resources endemic to the Northwest Pacific. *L. polyactis* distribute from the East China Sea northward to the Bohai Bay, while *L. crocea* from the South China Sea northward to the East China Sea, with their distribution ranges overlapping in the East China Sea (Wang, Liu, et al., 2013; Wang, Shi, et al., 2013). The two species differ in body size, spawning migration routes and overwintering aggregation grounds, but slightly overlap in spawning time and spawning grounds and nursery areas in the northern East China Sea (Ikeda, 1964; Liu, 1964). Previous studies have suggested that *L. polyactis* and *L. crocea* were isolated mainly in two different refuges during Pleistocene glacial times, located in the East and South China Sea, respectively (Wang, Shi, et al., 2013; Xiao et al., 2009). Following recent glacial retreat, both species have experienced rapid expansions and partial contact, and adapted to diversified environmental conditions, for example, spawning temperature and salinity (Liu et al., 2016; Wang, Liu, et al., 2013; Wang, Shi, et al., 2013). In the overlapping distribution region, we observed population mixture of the two species, implying recent interspecific gene flow although we could not exclude the possibility that there was absence of gene flow (Wang et al., 2012). Hence, this system might follow a speciation model of secondary contact. Therefore, the two species provide a valuable system for examining how differential evolutionary factors including gene flow, genetic barriers, selection and recombination landscapes, could have shaped the patterns of genomic divergence in the mode of secondary contact.

Here, we first sequenced the genome of *L. polyactis* and obtained genome-wide variants of the two sister species in their

distribution range, by RAD (restriction-site associated DNA) sequencing. We then examined the demographic history of the two species taking into consideration both heterogeneous introgression and linked selection in the genome, and tested whether the divergence conforms to the secondary contact scenario. Under this scenario, we expected restricted gene flow in genomic islands in comparison to the rest of the genome. In addition, we analysed the genome-wide patterns of genetic diversity to examine the hypothesis of “genomic islands” for interspecific divergence. In particular, by comparing the genetic diversity measure between genomic islands and the remaining regions, we infer the roles of gene flow, selection and recombination in promoting heterogeneous genomic divergence during two successive phases of speciation. Furthermore, by comparing the genomic patterns between interspecific and intraspecific divergence, we tested if the mechanisms that influence the genomic patterns of interspecific divergence also affect that of intraspecific differentiation. This study provides not only valuable genomic resources for the two sister species, but also novel insights into the evolutionary factors underlying genomic islands in the studied system.

2 | MATERIALS AND METHODS

2.1 | Ethics statement

The studied species are not protected by Chinese law or by any of the countries where the sampling was performed. It is a commercially harvested species in Northeast Asian countries. The samples were collected by trawling at the designed sampling sites and were already dead when sampling.

2.2 | Samples for genome, transcriptome and RAD sequencing

One female *L. polyactis* was selected for genome sequencing. One 20-kb library was constructed and sequenced on PacBio Sequel II (Pacific Biosciences) by BGI (Shenzhen). Both paired-end (PE) (270 and 550 bp) and mate pair (MP) (5 and 10 kb) libraries were also constructed from the same sample (Table S1). Three mRNA libraries were separately constructed for three individuals (two females and one male at 1 year old), by pooling total RNA from multiple tissues: brain, eye, gill, intestine, spleen, liver, heart, kidney, gonad, skin and muscle, with equal amounts from each tissue. PE, MP and mRNA libraries were sequenced for 2×150 bp by BGI (Shenzhen), using Illumina HiSeq 4000 (Illumina).

One hundred and five *L. polyactis* from five locations and 55 *L. crocea* from three locations were collected across their distribution ranges (Table S2 and Figure 1a). DNA was isolated from fin using standard phenol-chloroform extraction protocol. RAD libraries of 500-bp inserts were constructed according to a previous protocol (Baird et al., 2008), using restriction enzyme PstI-HF (New England

Biolabs). The libraries were sequenced on Illumina HiSeq 3000 platform (Illumina) in single end with read length of 150 bp.

2.3 | Genome assembly and annotation

Both PE (~55×) and MP reads (~40×) were first cleaned using `process_shortreads` in Stacks version 1.45 (Catchen et al., 2011). ALLPATHS-LG version 52,488 (Gnerre et al., 2011) was then used to estimate genome size based on k-mer frequencies and de novo assemble the genome. PacBio reads of ~80× coverage were assembled using Canu version 1.8 (Koren et al., 2017), with an estimated genome size of 650Mb (Figure S1). The assembly was then polished with Pilon version 1.22 (Walker et al., 2014), using both cleaned PE and MP libraries. There are no genomic resources available for chromosome-level genome assembly of *L. polyactis*. Due to a recent divergence at <5 million years ago (Ma) (Liu et al., 2010), *L. polyactis* and *L. crocea* probably have a high level of genome synteny. Comparison of long scaffolds between the two species found no evidence of chromosome rearrangements, except for six small-scale inversions (see below). We therefore anchored the scaffolds of *L. polyactis* to the high-density genetic maps of *L. crocea* (24 linkage groups [LGs], 2.7 cM/Mb and an average marker resolution of 0.36 cM) (Kong et al., 2019) to achieve a chromosome-level assembly using ALLMAPs (Tang et al., 2015) with default parameters.

Completeness of genome was assessed by BUSCO version 3.0.1 (Simão et al., 2015). MAKER2 (Cantarel et al., 2008) was used to annotate the genome according to the description in Appendix S1. We further predicted conserved noncoding elements (CNEs), according to a previous method (Brawand et al., 2014) (Appendix S1).

2.4 | Comparative genomic analysis between *L. polyactis* and *L. crocea*

Genomic synteny between *L. polyactis* and stickleback (Jones et al., 2012) was studied based on one-to-one orthologous genes, which were identified by pairwise blast search using Orthologuefinder (Horiike et al., 2016) based on OrthoMCL database version 5 (Chen et al., 2006). We further identified chromosome rearrangements between the two sister species. Since the pseudochromosome sequences of *L. polyactis* were dependent on the genetic maps of *L. crocea*, we could not examine rearrangements by pairwise chromosome alignment. Instead, we screened fragment rearrangements by pairwise alignment of scaffolds, which were directly assembled using PacBio reads for both species (Chen et al., 2019). Putative rearrangements were then verified by examining the linkage and/or recombination rates around the breaking points. If there was a fragment rearrangement in one species, for example, *L. polyactis*, markers flanking the breakpoints would be in tight linkage when using the genomic coordinates of *L. polyactis* for the markers. In contrast, the linkage between markers flanking the breakpoint would be distorted when using the genomic coordinates of *L. crocea* for the markers.

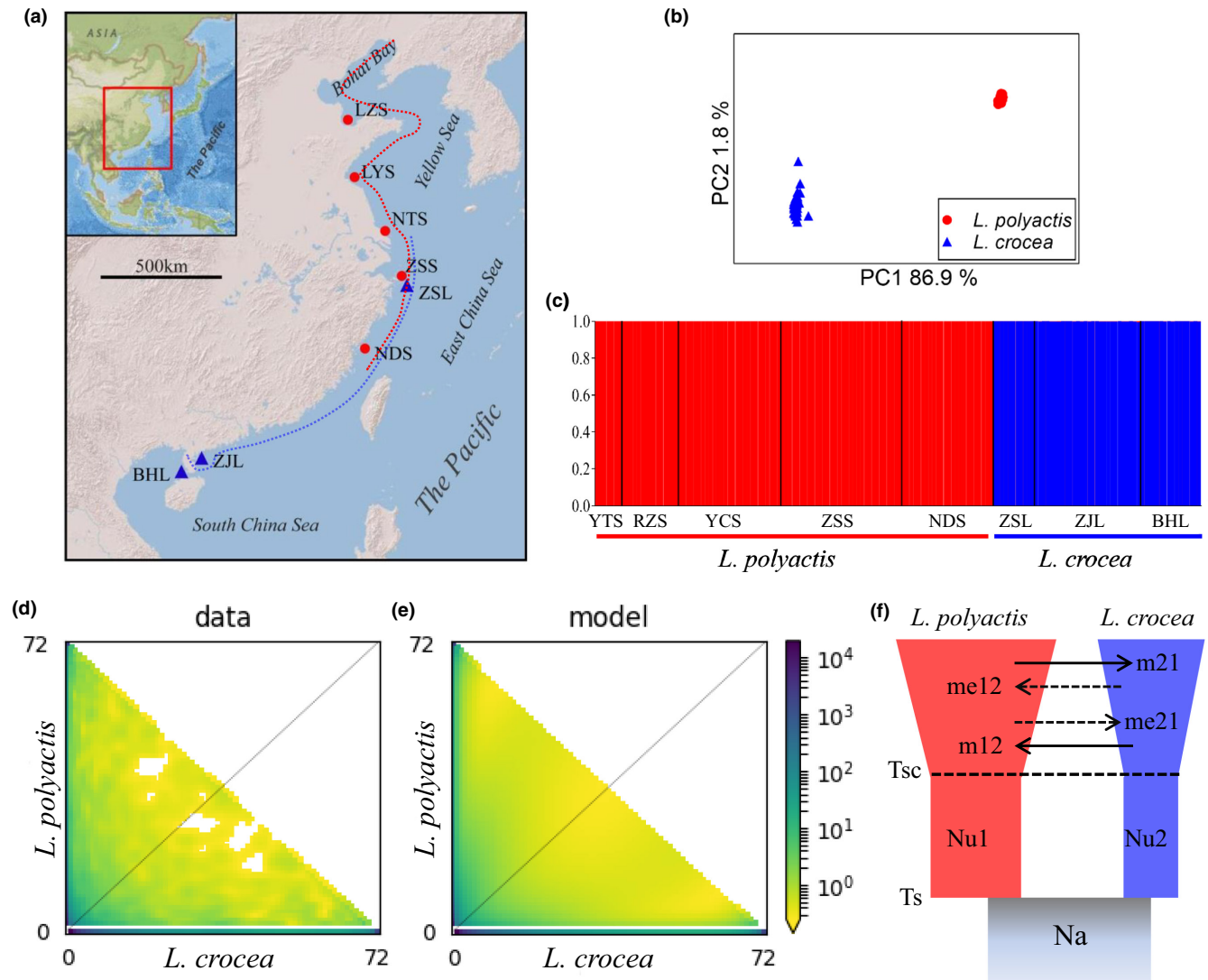


FIGURE 1 Population structure and demographic history between *L. polyactis* and *L. crocea*. (a) Sampling locations of *L. polyactis* (five, red circles) and *L. crocea* (three, blue triangles) along coastal waters of the Northwest Pacific, where the distribution range of *L. polyactis* and *L. crocea* are indicated with red and blue dashed lines, respectively. (b) Principal component analysis (PCA) of population structure between *L. polyactis* and *L. crocea*, based on common SNP data set. (c) Genetic clusters showing population structure between *L. polyactis* and *L. crocea*, based on common SNP data; (d) and (e) showing the folded joint site frequency spectrum (SFS) and the expected folded joint SFS under the most fitting demographic scenario (SC2N2mG), respectively, between *L. polyactis* and *L. crocea*. (f) The SC2N2mG demographic scenario between *L. polyactis* and *L. crocea*. Na indicates effective population size of ancestral population; Nu1 and Nu2 indicate the effective population size of the *L. polyactis* and *L. crocea* lineages after split from their common ancestral population, respectively; m12 and m21 indicate the homozygous gene flow, while me12 and me21 indicate heterozygous gene flow from one species to the counterpart; T_s denotes the split time of ancestral population and T_{sc} is the time when ancient isolation stops.

2.5 | Genome-wide SNP discovery and genotyping

Raw RAD sequencing reads were filtered using process_radtags (-r -c -q -t 135) in Stacks version 1.45 (Catchen et al., 2011). An average number of 6.2 M clean reads was obtained per sample. We found that 98.6% and 98.2% of reads of *L. polyactis* and *L. crocea*, respectively, were mapped to the genome of *L. polyactis*, while 96.9% and 98.4% of reads of *L. polyactis* and *L. crocea*, respectively, were mapped to the genome sequences of *L. crocea*. The genome sequence of *L. polyactis* was used as reference due to the slightly

higher mapping rates. Clean reads were aligned to the reference genome using BWA-MEM (Li & Durbin, 2009) with default parameters. Uniquely aligned reads were used for SNPs discovery and genotyping using Stacks version 1.45 (Catchen et al., 2011) with default parameters. We further filtered out the variants that had more than two alleles, deviated from Hardy-Weinberg equilibrium (HWE) in a single population at the significance threshold of 0.001 and showed heterozygosity of >0.5. Only SNPs that were present in >70% of individuals in each location, were retained for further analysis.

2.6 | Genetic diversity and population structure

Genetic diversity in terms of nucleotide diversity (π) was estimated by taking monomorphic sites into consideration using VCFtools version 0.1.15 (Danecek et al., 2011). The program ARLEQUIN version 3.5 (Excoffier & Lischer, 2010) was used to estimate pairwise F_{ST} among locations and examine the significance of differentiation using Fisher's exact test. Principal components analysis between and within species, was performed using Plink2 (Purcell et al., 2007). Population structure in the form of genetic clusters was analysed using Admixture version 1.3.0 (Alexander & Lange, 2011), with the number of genetic clusters ranging from 2 to 10. For these analyses, we filtered the SNPs that show high linkage disequilibrium: only one SNP was retained from each RAD tag and those with neighbouring SNPs on different RAD tags and showing elevated LD ($R^2 > 0.1$) within 100kb window were excluded. Patterns of isolation-by-distance were examined using Mantel test (Smouse et al., 1986).

2.7 | Modelling demographic history

The demographic history of divergence between species was studied by modelling the alternative two-population demographic scenarios based on joint site frequency spectrum (SFS), with the program DaDi version 1.6.3 (Gutenkunst et al., 2009). Due to very weak intraspecific differentiation ($F_{ST} < 0.01$), we pooled samples separately for each species. We selected 45 individuals separately from *L. polyactis* and *L. crocea* that had the least missing data. With missing data of <10%, 86,854 SNPs were retained to generate the folded joint SFS. We further projected down the folded SFS to a phase, in order to minimize the effect of missing data and meanwhile maximize the effective number of samples. We first fitted the four standard alternative scenarios of historical divergence according to our previous study (Wang et al., 2016), including the strict isolation model (SI), isolation-with-migration model (IM), ancient migration model (AM) and secondary contact model (SC), where the gene flow was assumed to be shared among loci. We further fitted the folded SFS to three modified models: IM2m, AM2m and SC2m, corresponding to the above IM, AM and SC models, respectively. The modified models allow heterogeneous gene flow across the genome using two categories of effective migration rates among loci ($-2m$) (Tine et al., 2014). Three additional modified alternative scenarios: IM2N, AM2N and SC2N that take into consideration the Hill–Robertson effect (Hill & Robertson, 1966) or effect of linked selection, were further modelled to fit the spectrum by considering two categories of loci ($-2N$) occurring in the genome (Rougeux et al., 2017). We also modified the four standard alternative models by allowing population expansion ($-G$) (SIG, IMG, AMG and SCG). The best fitting models within each category (standard, $-2m$, $-2N$ and $-G$), were inferred from Akaike information criterion (AIC), and were then optimized by simulation of historical population growth, allowing heterogeneous gene flow and/or considering the effect of linked selection ($-2mG$, $-2NG$ and $-2N2mG$), to capture additional residuals of the folded

SFS. The parameter estimates from each alternative scenario were examined for convergence by running the program for 20 replicates with each using independent starting parameters, and the confidence intervals of each parameter for the best fit model were obtained by bootstrapping over loci for 100 times.

2.8 | Detection of outlier loci influenced by selection

Footprints of local selection within each species were first detected using FDIST2 (Beaumont & Balding, 2004), which assumes that loci under putative diversifying selection would show elevated population differential condition on their heterozygosity. Loci that were distributed above 99% quantiles of the null distribution and were significant at 0.01 level after FDR corrections were considered as outliers. We further employed a Bayesian simulation-based test to identify signatures of local selection using BAYESCAN version 2.1 (Foll & Gaggiotti, 2008), which works by decomposing locus-specific F_{ST} into a population and a locus specific component, and evaluates the necessity to include the locus specific component to account for selection. Posterior odds (PO) were estimated for each locus and loci that were significant at 0.01 level were considered as outliers. For these analyses, only SNPs with minor allele frequency > 0.05 were used, in order to simulate the null distribution and reduce false positives in the background of high gene flow. Outlier SNPs that were located in the genomic regions of protein coding genes and within its 1-kb flanking regions were considered for functional enrichment analysis. Genes under positive selection and related to outliers were extracted from genome annotation. GO IDs of these genes were obtained by blast search against the zebrafish protein database (Ensembl database release 96) using blastp with an E-value cutoff $< 1E-10$. Gene ontology (GO) enrichment analysis was carried out by comparing the GO IDs of the outlier genes with that of all genes of zebrafish, using the program Metascape (Zhou et al., 2019) with default parameters.

2.9 | Identifying genomic regions of elevated differentiation

In order to identify genomic islands, we partitioned the genome sequences into 150-kb nonoverlapping windows and estimated both interspecific and intraspecific F_{ST} for each window. Unweighted F_{ST} was estimated with VCFtools version 0.1.15 (Danecek et al., 2011). Windows of less than five SNPs were removed, which kept 3706 windows for further analysis. We estimated the distribution of F_{ST} across windows and those within the upper 5% quantile of the F_{ST} value distribution were considered as genomic islands, according to previous studies (Le Moan et al., 2021; Liu et al., 2020). Adjacent outlier windows were merged as one. We compared the other window sizes of 100, 200 and 500kb to infer evolutionary mechanisms and found little difference among the results. We also selected

windows within the upper 1% quantile of the F_{ST} value distribution, which did not lead to different results with the windows within the upper 5% quantile. For intraspecific divergence, 3411 windows were retained for inferring genomic islands. Intraspecific genomic islands were defined using the same criterion as the above. Intraspecific F_{ST} was estimated across all populations using hierarchical island model with the R package “hierfstat” (Goudet, 2005).

2.10 | Inferring factors shaping heterogeneous genomic divergence

To dissect the factors promoting heterogeneous genomic divergence, we estimated the genetic diversity parameters of genomic islands in comparison to the rest of the genome. Absolute sequence divergence, D_{xy} , between *L. polyactis* and *L. crocea*, was calculated for individual windows, using a perl script (Han et al., 2017). Window-sized estimation of nucleotide diversity (π) and Tajima's D were calculated using VCFtools version 0.1.15 (Danecek et al., 2011). Tajima's D is often used to detect selective sweeps. Negative Tajima's D values indicate an excess of low frequency polymorphisms relative to equilibrium expectation, indicating recent selective sweep or population expansion after recent bottleneck. Population-scaled recombination rate along individual chromosomes, was measured as $\rho = 4Ner$ per kb and estimated using LDhat version 2.1 (McVean et al., 2004) with 150-kb nonoverlapping sliding windows, with parameters: -bpen 10, -its 10,000,000, -theta 0.05. The differences of genetic diversity measures between genomic islands and the rest of the genome were examined using Mann-Whitney U test. The correlations of genomic landscapes between genetic diversity measures were assessed using Pearson's correlation coefficient.

3 | RESULTS

3.1 | Genome assembly and annotation

Genome assembly of *L. polyactis* was ~653Mb, with contig N50 of 2.47Mb. Approximately 95.7% of total sequences were anchored onto high density linkage maps of *L. crocea*, with N50 length of pseudomolecules of 37.31Mb (Table S3). Assessment of genome completeness showed that only 1.7 and 0.0% of the BUSCO genes, respectively, were missing and fragmented, indicating the high quality of the genome assembly (Table S4). Repeated sequences accounted for ~21.1% of the genome, in which 83.7% were transposable elements (Table S5). A total of 24,625 protein coding genes were predicted in *L. polyactis*, showing little differences with that of *L. crocea* (23,657 and 25,401 for two different assemblies, respectively) (Ao et al., 2015; Wu et al., 2014). We predicted 235,237 CNEs with an average length of 202 bp.

Overall GC content in the genome of *L. polyactis* was 42.3%. However, it was not randomly distributed throughout the genome. For some chromosomes, for example, chr1 and chr24, GC content

was higher towards both ends, while for some other chromosomes, for example, chr4 and chr5, GC rich sequences were located in the middle of the chromosomes (Figure S2). In addition to GC content, repeated sequences, recombination rates, as well as CNEs were also observed not to be randomly distributed across the whole genome (Figure S2). Interestingly, we found that GC content was positively correlated with recombination rates ($R = 0.596$, $p < 10^{-15}$) and content of repeated sequences ($R = 0.435$, $p < 10^{-15}$), but negatively correlated with content of CNEs ($R = -0.309$, $p < 10^{-15}$). Further investigation showed that content of repeated sequences was also positively and negatively correlated with recombination rates and content of CNEs, respectively (Figure S3).

We compared the genomic synteny between *L. polyactis* and stickleback. A high collinearity was observed between any pair of homologous chromosomes (Figure S4), except for some evidence of genome rearrangements (Figure S5), although the two species have diverged since ~100 million years ago (<http://www.timetree.org/>). Six putative chromosome inversions were identified between *L. polyactis* and *L. crocea*, at contig level (Figure S6). With sufficient marker resolution, we estimated the recombination rates around the inverted region of the longest fragment inversion. We found that recombination rates in the regions closely flanking the breaking points were increased for *L. crocea* and *L. polyactis* when referenced to the genome of *L. polyactis* and *L. crocea*, respectively (Figure S7). Such distortion of recombination rates around the breaking points was suggested as a feature of chromosome inversion (Crown et al., 2018).

3.2 | Population structure between and within species

After quality control, a total number of 1,113,736 SNPs were obtained. The SNPs were distributed on 88,300 RAD tags throughout the two sister species, within which 149,201 SNPs distributed on 61,346 RAD tags showed a minor allele frequency of ≥ 0.05 (Table S6). Within each species, 806,471 and 537,052 SNPs distributed on 87,701 and 83,264 RAD tags were obtained, for *L. polyactis* and *L. crocea*, respectively. A total of 129,768 and 114,520 SNPs, distributed on 54,059 and 48,672 RAD tags showed a minor allele frequency of ≥ 0.05 for *L. polyactis* and *L. crocea*, respectively (Table S6).

Overall genetic diversity in terms of nucleotide diversity (π) showed little difference between species (*L. polyactis*: 0.0048–0.0056 vs. *L. crocea*: 0.0040–0.0050) and among locations within species (Table S2). We observed clear divergence between species (F_{ST} , ~0.6), but very weak population differentiation within each species, with F_{ST} ranging from 0.0001 to 0.0068 and from 0.0009 to 0.0014 for *L. polyactis* and *L. crocea*, respectively (Table S7). In particular for *L. polyactis*, none of the pairwise F_{ST} across all studied locations was significant except for one pair between the northernmost (LZS) and the southernmost (NDS) locations (Table S7). Overall population differentiation for *L. polyactis* did not follow a pattern of isolation-by-distance (Mantel test, $R = 0.408$, $p > .05$). Consistently, both PCA and Admixture analysis revealed striking interspecific

genetic divergence (Figure 1b,c). However, we did not observe a significant geographical pattern of divergence within each species (Figures S8–S11).

3.3 | Demographic history of speciation

A total of 17 alternative demographic models were designed to fit the folded joint SFS (Figure S12). Among the four standard alternative models, SC was indicated to have the best support, showing the highest maximum-likelihood value and lowest AIC score (Table S8). In agreement with the above results, SC2m and SC2N were the best supported scenarios in the two categories of custom alternative scenarios that take into consideration the heterogeneous migration ($-2m$) and the effect of linked selection ($-2N$) in the genome, respectively (Table S8). Among the four modified alternative models that allow population expansion ($-G$), SCG was observed as the best model to capture the folded spectrum. Finally, we modified the SC model by allowing population expansion, heterogeneous migration and/or linked selection (SC2mG, SC2NG and SC2N2mG) and observed substantial improvement for these models to capture the folded spectrum. SC2N2mG was the best supported model among all examined alternative scenarios (Figure 1d–f and Table S8), where the folded SFS was fairly well captured and most of the residuals were in a very narrow range from -1 to 1 (Figure S13).

We converted the demographic parameters in units of $2Na$ generations into biologically meaningful units using the formula $Na = \theta / (4\mu * L)$, where μ is the mutation rate fixed at 10^{-8} mutations per site per generation and L is the effective length of the genome explored by RAD sequencing (Rougeux et al., 2017). Under SC2N2mG, $\sim 16.3\%$ of genome was observed to have a reduction in effective migration rates compared to the rest of the genome (Table S9). The effective migration rates of genomic-island genes from *L. polyactis* to *L. crocea* ($1.851E-07$) and from *L. crocea* to *L. polyactis* ($1.481E-07$) were more than 10 times lower than those for neutral migration rates ($3.701E-06$ and $5.120E-06$, respectively) (Table S9). The proportion of the genome with a locally reduced effective size due to the effect of background selection and selective sweeps was estimated to be $\sim 95.4\%$, but the reduction in effective size in these regions was estimated to be only $\sim 6.9\%$ (Table S9). Using an average generation time of 3 years (Liu, 1962), the divergence time between *L. polyactis* and *L. crocea* was dated to ~ 1.19 (95% CI: 1.10–1.28) Ma. The time when the two species had secondary contact was estimated to be relatively ancient near ~ 0.81 (95% CI: 0.77–0.86) Ma. The population sizes for *L. polyactis* and *L. crocea* were revealed to have dramatically expanded by ~ 10 and ~ 110 times, respectively.

3.4 | Adaptive divergence within species

Outlier tests identified 779 and 206 SNPs under putative local selection for *L. polyactis* and *L. crocea*, respectively (Table S10). Only two outliers were overlapping between species. Annotation

against the reference genome revealed that 457 (58.7%) and 125 (60.7%) outliers were associated with 440 and 125 unique protein coding genes for *L. polyactis* and *L. crocea*, respectively, which was higher than the ratio under the expectation of random distribution that $\sim 45\%$ of outliers are located in genic regions. We further observed that 23.2% (181) and 27.7% (57) of outliers were located in coding sequences for *L. polyactis* and *L. crocea*, respectively, which were also higher than the ratio under the expectation of random distribution that $\sim 7\%$ of outliers hit the coding sequences in the genome. Only 10 outlier genes were overlapping between species. Gene ontology analysis showed that there was no overlapping for the enriched terms between *L. polyactis* and *L. crocea* (Figure S14).

Population differentiation based on outliers were statistically significant and much higher than that of neutral differentiation (F_{ST} ranging from 0.03 to 0.28 and from 0.07 to 0.14 for *L. polyactis* and *L. crocea*, respectively), and presented clear geographical patterns, for both species (Table S11–12). PCA analysis showed that *L. polyactis* were differentiated into three populations, corresponding to the Bohai Bay, the Yellow Sea and the East China Sea, respectively, while *L. crocea* were also clustered into three populations: one population in the East China Sea and two populations in the South China Sea that were separated by the Leizhou Strait (Figure 2a,b). Admixture analysis revealed that five locations of *L. polyactis* and three locations of *L. crocea* were distinctively differentiated with each other within species (Figures 2c,d, S15 and 16). Interestingly, the overall population differentiation at outlier loci presented a significant pattern of isolation-by-distance for *L. polyactis* (Mantel test, $R = 0.738$, $p < .05$), contradicting with that of neutral differentiation (Figure S17).

3.5 | Heterogeneous genomic divergence between species

We identified 80 regions with elevated differentiation (F_{ST} , 0.947 ± 0.020) in comparison to the rest of the genome (F_{ST} , 0.627 ± 0.148 ; Mann–Whitney U test, $p < 10^{-15}$). These genomic islands were distributed across almost all chromosomes, with most of the peaks emerging towards the two ends of the chromosomes while the others were in the middle of the chromosomes (Figure 3). Interestingly, out of the six putative chromosomal inversions identified between species, four located in the genomic islands at chr2, chr3, chr7 and chr19, respectively (Figure 3). These genomic islands showed significantly reduced nucleotide diversity (π) and skews towards an excess of rare alleles (as examined in terms of reduced Tajima's D) in each species (Table S13).

To explore what factors have contributed to the heterogeneous genomic divergence, we examined the distribution pattern of D_{xy} throughout the genome. Consistent with the patterns of F_{ST} , D_{xy} was also observed to be elevated in genomic islands where gene flow was restricted as revealed by demographic history modeling, in comparison to the rest of the genome (D_{xy} , 0.030 vs. 0.016;

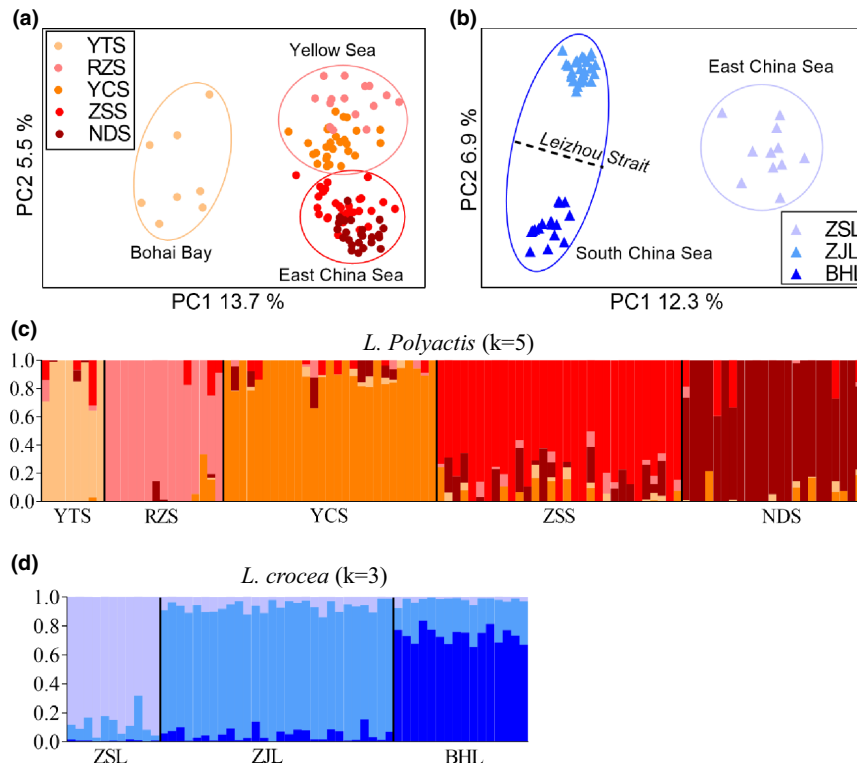


FIGURE 2 Adaptive population differentiation within *L. polyactis* and *L. crocea*. (a) Population structure revealed by PCA among five locations of *L. polyactis*, based on outlier loci influenced by putative local selection, where samples from the Bohai Sea, the Yellow Sea and the East China Sea are highlighted with circles. (b) Population structure revealed by PCA among three locations of *L. crocea*, based on outlier loci influenced by putative local selection, where samples from the each China Sea and the South China Sea are highlighted with circles. (c) Genetic clusters inferred by Admixture among five locations of *L. polyactis* at the most likely K value of 5, based on outlier loci, where the colours of genetic clusters correspond to those of locations in (a), and (d) genetic clusters among three locations of *L. crocea* at the most likely K value of 3, based on outlier loci under putative local selection, where the colours of genetic clusters are corresponding to those of locations in (b).

Mann–Whitney U test, $p < 10^{-15}$). In addition, D_{xy} showed a genome-wide positive correlation with F_{ST} (Pearson's correlation: $R = 0.463$, $p < 10^{-15}$) (Figure S18).

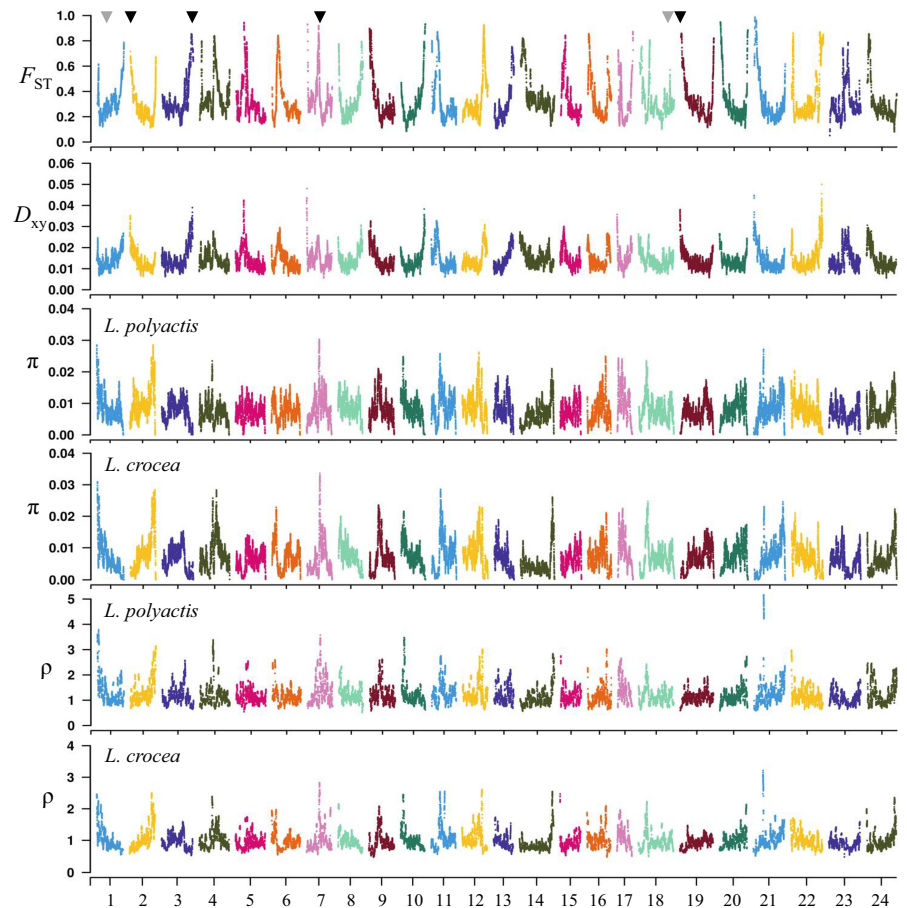
To uncover additional factors leading to heterogeneous divergence, we further examined the distribution pattern of recombination rate (ρ) throughout the genome. In agreement with the distribution pattern of π , we observed that ρ was also significantly reduced in genomic islands compared to the rest of the genome in both species (Figure 4; Table S13). Interestingly, we observed that π and ρ were not only positively correlated with each other within each species (Pearson's correlation: $R = 0.770$, $p < 10^{-15}$ and $R = 0.782$, $p < 10^{-15}$ for *L. polyactis* and *L. crocea*, respectively; Figure 5a,b), but also separately positively correlated between *L. polyactis* and *L. crocea* (π : $R = 0.798$, $p < 10^{-15}$ and ρ : $R = 0.706$, $p < 10^{-15}$) (Figure 5c,d), supporting a relatively good conservation of both nucleotide diversity and recombination landscapes between species. The overall distribution of π and ρ across chromosomes presented negative relationships with F_{ST} for both species (Figures 5e and S19). These observations suggest that linked selection in low recombining regions including genomic islands played crucial roles in the ancestral lineage before divergence of the two species. Interestingly, we

found a global pattern of positive correlations between π and D_{xy} , and between ρ and D_{xy} , for both species (Figure 5e).

3.6 | Intraspecific versus interspecific heterogeneous divergence

In high gene flow background, we observed evidently elevated genomic regions across most of the chromosomes for both species, although the genomic islands were much less pronounced than those of interspecific divergence (Figure 6a,b). A total of 144 and 142 genomic islands were identified in *L. polyactis* and *L. crocea*, respectively. These genomic islands showed a slight reduction in both nucleotide diversity and recombination rate in comparison to the rest of the genome in *L. polyactis*, but not in *L. crocea* (Table S14). We identified 11 (~8%) intraspecific “genomic island” overlapping between species (Figure 6a,b). There was a weak positive correlation for the distribution of lineage-specific F_{ST} , between *L. polyactis* and *L. crocea* (Pearson's correlation: $R = 0.135$, $p < 10^{-14}$). Interestingly, we observed that 5.0% (4) of the interspecific “genomic islands” coincided with intraspecific ones for either *L. polyactis* (2) or *L. crocea* (2) (Figure 6a,b), slightly lower than that

FIGURE 3 Genome-wide heterogeneous divergence between and within *L. polyactis* and *L. crocea*. Genome-wide patterns of interspecific divergence (F_{ST}) and absolute sequence divergence (D_{xy}) between *L. polyactis* and *L. crocea*, intraspecific nucleotide diversity (π) and recombination rate (ρ) separately for each sister species *L. polyactis* and *L. crocea*, were estimated along individual chromosomes based on 150 kb-windows. Six putative chromosomal inversions are indicated with triangles, among which four overlapped with genomic islands at chr2, chr3, chr7 and chr19, respectively, are highlighted with black.



of the expectation under random distribution of $\sim 7.3\%$. We mapped outlier loci putatively influenced by local selection in either species to the genome and found that these loci were not enriched in interspecific genomic islands. Only 12 out of 984 (1.2%) were located in these regions, less than that of the expectation under random distribution of 4.2%. Interestingly, we observed significantly negative correlations between interspecific F_{ST} and intraspecific F_{ST} for both species (Pearson's correlation: $R = -0.312$, $p < 10^{-15}$ and $R = -0.299$, $p < 10^{-15}$ for *L. polyactis* and *L. crocea*, respectively).

4 | DISCUSSION

In this study, we assembled a chromosome-level genome sequence of *L. polyactis*. Based on the genome, we modelled the demographic history and analysed the genome-wide patterns of genetic diversity between and within two sister species: *L. polyactis* and *L. crocea*. Our data support that the two species diverged following a secondary contact scenario, and the genome-wide heterogeneous divergence between species is due to linked selection in the two species and restricted gene flow in highly differentiated regions in the phase of secondary contact. Our results also address the crucial roles of recombination rate variation in facilitating evolution of highly differentiated regions throughout the whole genome.

4.1 | Important genomic resources for *Larimichthys*

In this study, we assembled and annotated a high-quality genome sequence of *L. polyactis*. In particular, beyond protein coding genes, we predicted 235,237 CNEs throughout the whole genome. Previous studies have shown that CNEs can affect the expression of neighbouring genes by cis-regulation (Wang et al., 2022, 2021). Therefore, sequence variations in CNEs could also play important roles in bringing about phenotypic innovations. The chromosomal-level genome sequence provides valuable resources for investigation of the genome-wide landscape of divergence. We detected very low genetic differentiation within each species using genome-wide variants, indicating high gene flow. In particular, *L. polyactis* was shown to be nearly panmictic in the distribution range. However, a lack of geographical pattern of population differentiation notably affects management and conservation of the natural resources of both species (Wang et al., 2015). Outlier loci can increase the power of mixed stock analysis between migratory groups of *L. polyactis* (Wang et al., 2015). Here, we separately identified hundreds of outlier loci putatively influenced by local selection and detected clear geographical patterns of adaptive differentiation for both species. Adaptive differentiation within each species not only provides advanced information for designing functional units for conservation and management of the natural resources, but also helps predict

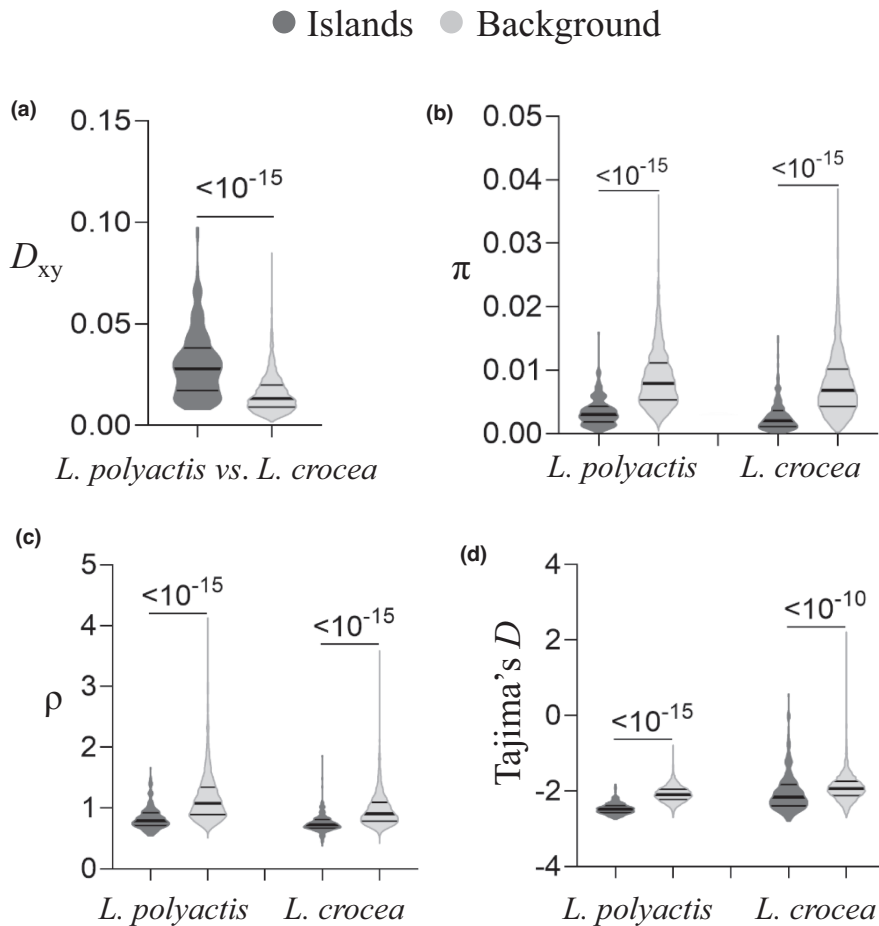


FIGURE 4 Genome-wide heterogeneous divergence between *L. polyactis* and *L. crocea*. (a–d) show the comparisons of absolute sequence divergence (D_{xy}), nucleotide diversity (π), recombination rate (ρ) and Tajima's D between genomic islands and the rest of the genome for interspecific divergence between *L. polyactis* and *L. crocea*, respectively, where p -values for Mann–Whitney U tests are indicated above the plots.

the evolutionary trajectory of the populations against global climate change (Allendorf et al., 2010).

4.2 | Linked selection before secondary contact shapes heterogeneous divergence

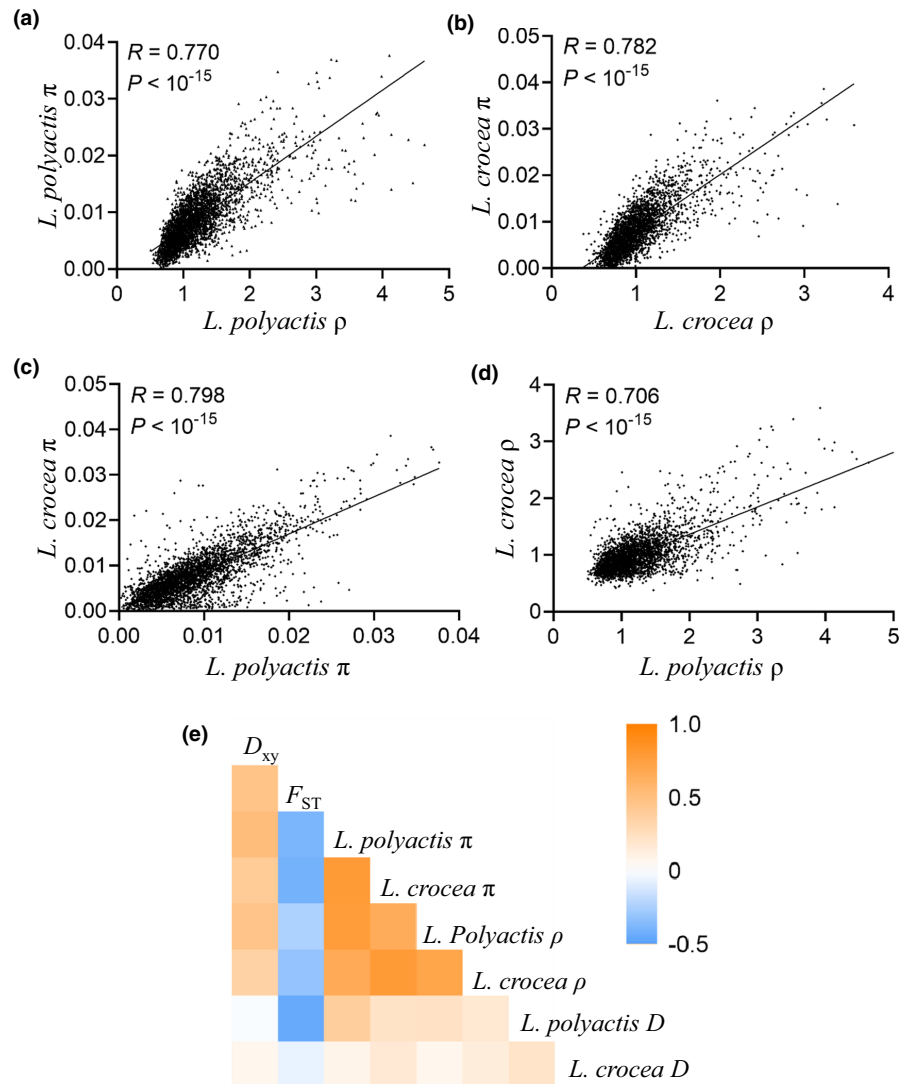
Pleistocene glacial fluctuations and the following postglacial recolonization have played critical but complicated roles in speciation or divergence, by repeated glacial isolation and mix (Duranton et al., 2018; Nevado et al., 2018; Wang et al., 2016). Looking into the demographic history is necessary to disentangle the factors that shape the patterns of both population differentiation and genome-wide divergence (O'Connell et al., 2021). Herein, interspecific divergence could be best explained by the SC2N2mG scenario, where differential introgression and linked selection are suggested to have shaped the pattern of heterogeneous genomic divergence (Rougeux et al., 2017).

Under the SC2N2mG scenario, a majority of the genome was suggested to have reduced effective size compared to the rest of the genome of either or both species, although the degree to which the effective population size locally reduced is not very high (~7%). As expected under the scenario of linked selection, we observed both reduced genetic diversity and recombination rate in genomic islands in comparison to the rest of the genome, and significantly positive correlations between genetic diversity and recombination rate in both species, and a negative correlation between recombination rate and

differentiation (F_{ST}) (Burri et al., 2015; Cruickshank & Hahn, 2014; Duranton et al., 2018). Interestingly, we found a good conservation of both nucleotide diversity and recombination landscapes between species. Such shared patterns of genetic diversity strongly suggest that linked selection in the common ancestral lineage of the two studied species has shaped the observed heterogeneous genomic landscape between species. This is because recombination rate is the key mediator of linked selection, and linked selection based on either background selection or selective sweeps can reduce genetic variation in ancestral population and consequently generate shared patterns of reduced genetic diversity in regions of low recombination (Burri et al., 2015; Cruickshank & Hahn, 2014; Vijay et al., 2017). This result is similar to the observations in *Ficedula* flycatchers (Burri et al., 2015), Darwin's finches (Vijay et al., 2017) and *Primulina* species (Ke et al., 2022), where the shared patterns of heterogeneous genomic landscapes between species or lineages are contributed by linked selection in low recombination regions.

However, under the scenario of linked selection, D_{xy} is expected to be unchanged or reduced in highly differentiated regions due to the effects of recurrent background selection or selective sweeps, which reduce sequence diversity in these regions (Burri et al., 2015; Cruickshank & Hahn, 2014; Vijay et al., 2017). Contrary to this expectation, we observed elevated D_{xy} in genomic islands in comparison to the rest of the genome and a positive correlation between D_{xy} and F_{ST} . These data suggest that additional evolutionary factors that can increase absolute divergence have probably remodelled

FIGURE 5 Pairwise correlations between genetic diversity measures between and within species. (a) and (b) show the overall Pearson's correlation between recombination rate and nucleotide diversity for *L. polyactis* and *L. crocea*, respectively. (c) and (d) show the overall Pearson's correlation of nucleotide diversity and recombination rate, respectively, between *L. polyactis* and *L. crocea*. (e) a heatmap showing the overall Pearson's correlation among D_{xy} , F_{ST} , π , ρ and Tajima's D between and within *L. polyactis* and *L. crocea*.



the heterogeneous genomic divergence landscapes that have been previously shaped by linked selection in ancestral lineage. Under the scenarios of divergent sorting of ancient polymorphisms and speciation-with-gene-flow or secondary contact, D_{xy} is expected to elevate in highly differentiated regions, leading to a positive correlation between D_{xy} and F_{ST} (Cruickshank & Hahn, 2014; Han et al., 2017; Ma et al., 2018). In particular, linked selection can accelerate lineage sorting in low recombination regions by increasing the chance to fix alternative alleles in these regions, generating heterogeneous genomic divergence between diverging populations at the initial stage of geographic isolation (Duranton et al., 2018). Divergent sorting of ancient haplotypes after the onset of speciation is common and predicted to elevate D_{xy} and have reduced lineage-specific genetic measures (ρ , π and Tajima's D) in genomic islands (Han et al., 2017; Ke et al., 2022; Ma et al., 2018). Under this scenario, global patterns of positive correlations between F_{ST} and D_{xy} , between genetic diversity and D_{xy} , and between recombination rate and D_{xy} , are also expected (Ke et al., 2022; Ma et al., 2018; Wang et al., 2019). Here, we observed the same global patterns of positive correlations for these genetic diversity measures in both species,

suggesting that divergent sorting of ancient polymorphisms probably with the aid of linked selection, divergent local selection and genetic drift has also played important roles in promoting heterogeneous genomic divergence (Duranton et al., 2018; Han et al., 2017; Ma et al., 2018). In addition, recombination landscape per se can not only influence linked selection but also divergent sorting of ancient polymorphisms, leading to a similar pattern of D_{xy} as observed in this study (Duranton et al., 2018). Other than these factors, the differential introgression following secondary contact scenario as revealed by demographic modelling also generates the same pattern of D_{xy} in highly differentiated regions, where gene flow is restricted and thus absolute sequence divergence is elevated (Cruickshank & Hahn, 2014; Han et al., 2017; Ma et al., 2018).

4.3 | Differential introgression following secondary contact promotes heterogeneous genomic divergence

The inferred demographic model suggests that differential introgression following secondary contact is also responsible for the

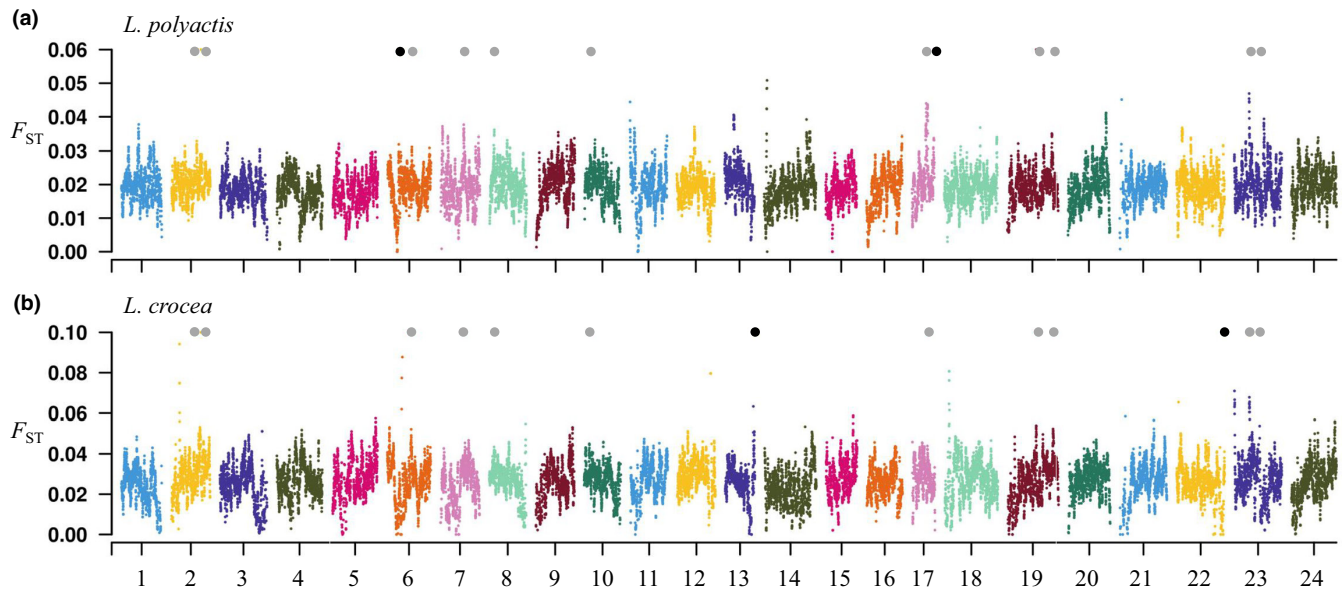


FIGURE 6 Genome-wide patterns of heterogeneous divergence within *L. polyactis* and *L. crocea*. (a) and (b) genome-wide patterns of genetic divergence within *L. polyactis* and *L. crocea*, respectively, estimated based on 150-kb sliding windows. Overlapping genomic islands between the two species are indicated with grey solid dots, while overlapping genomic islands between intraspecific and interspecific divergence are highlighted with black solid dots.

heterogeneous genomic divergence. We found that both F_{ST} and D_{xy} were increased, while all lineage-specific measures (ρ , π and Tajima's D) were reduced in genomic islands, which are in accordance with the secondary contact scenario (Burri et al., 2015; Cruickshank & Hahn, 2014; Duranton et al., 2018). Under this scenario, the split between the two sister species was estimated to be ~1.2 Ma, overlapping with the Calabrian stage of Pleistocene, when recurrent divergence events of marine animals were revealed to occur in the West Pacific (Jiang et al., 2021; Shahdadi et al., 2022; Zhai et al., 2012). The following postglacial contact was suggested to start recently at ~0.8 Ma. This implies that *L. polyactis* and *L. crocea* had evolved in allopatry for ~0.4 million years. The estimation of the time of the split between species was noticeably more recent than that reported in a previous study (Liu et al., 2010). This is probably due to the estimation taking the demographic models with heterogeneous population size and migration rates into consideration (Guirao-Rico et al., 2017; Liu et al., 2020). It is also probably due to the underestimation of generation time, which can be longer in the past colder environment.

Under the best supported model, we observed reduced gene flow in genomic islands by more than one order of magnitude in comparison to the rest of the genome. It has been shown that heterogeneous migration rates are common at intermediate stages of speciation in various animal and plant systems (Liu et al., 2020; Rougeux et al., 2017; Tine et al., 2014). The overall effective gene flow throughout the genome islands was estimated to not be enough to counteract the effects of genetic drift ($N_m < 1$). In contrast, gene flow estimated based on the genome that is free of genomic islands was effective to counteract the effects of genetic drift ($N_m > 1$). These data suggest that genetic barriers to gene flow could have

been established in genomic islands (Duranton et al., 2018; Wang et al., 2016). Upon secondary contact following glacial retreat, genetic differentiation in regions that are not related to adaptive differentiation or reproductive isolation was eroded by gene flow (Cruickshank & Hahn, 2014; Harrison & Larson, 2016). Therefore, restricted gene flow has resulted in highly differentiated regions and elevated D_{xy} in these regions in comparison to the rest of the genome.

The rising of genomic islands can be strengthened through postzygotic isolation mechanisms after secondary contact, when intermediate hybrids are selected against (Butlin & Smadja, 2018; Duranton et al., 2018). This mechanism can further strengthen the effect of genetic barriers to gene flow in genomic islands (Butlin & Smadja, 2018). In addition, chromosomal rearrangements can reduce gene flow and affect linked isolation genes by altering local recombination landscape, prompting the evolution of reproductive barriers (Crown et al., 2018; Rieseberg, 2001). Here, evidence for chromosomal inversions were observed to coincide with genomic islands between the two species, addressing the important roles of genomic islands in speciation in the studied system. Moreover, genomic islands generated by linked selection can be involved in reproductive isolation during early allopatric speciation, although some of them are incidental and can be eroded by gene flow (Duranton et al., 2018). Furthermore, highly differentiated regions generated due to divergent sorting of ancient polymorphisms during the early divergence period, can also act as barriers to gene flow upon secondary contact (Duranton et al., 2018). Finally, we found reduced recombination rate in genomic islands compared to the rest of the genome. Low recombination rate helps cumulating the effect of multiple mutations contributing to reproductive

isolation, such that they resist more to gene flow when linked together on a same haplotype (Duranton et al., 2018). In mouse and butterflies, genetic barriers to introgression are suggested to be shaped by recombination rate variation, with a positive correlation between introgression and recombination rate (Janoušek et al., 2015; Martin et al., 2019). Above all, differential gene flow is a major factor promoting genomic islands of interspecific divergence in the second phase of divergence under secondary contact scenario. The reduced introgression in low-recombining regions is probably related to the effects of linked selection, divergent sorting of ancient polymorphisms, genetic drift, accumulation of different mutations and selection against introgressed alleles from the alternated species due to for example, genetic incompatibilities (Jahner et al., 2021; Martin & Jiggins, 2017; Pazhenkova & Lukhtanov, 2021).

4.4 | Lack of correlated pattern between interspecific and intraspecific divergence

Previous studies in birds have shown that the levels of genetic diversity and differentiation are highly correlated among population-, species- and even higher-order taxa (Burri et al., 2015; Van Doren et al., 2017; Vijay et al., 2017). It is suggested that similar processes (e.g., linked selection in specific genomic regions of low recombination) have probably been underlying such genome-wide heterogeneous landscapes of divergence (Burri et al., 2015; Van Doren et al., 2017; Vijay et al., 2017). In this study, we first observed both conserved nucleotide diversity and in particular recombination landscapes between the two species, implying correlated variations in the intensity of linked selection between intraspecific and interspecific genetic diversity. However, we found that recombination rates were reduced in intraspecific genomic islands in *L. polyactis* but not in *L. crocea*. In particular, ~8% of intraspecific genomic islands were revealed to overlap between the two species, and ~5% interspecific genomic islands were overlapping with those of intraspecific divergence. However, these ratios of overlapping between interspecific and intraspecific genomic islands are lower than those of the expectation under random distribution. These data show little support to the hypothesis that linked selection in ancient lineage of the two diverging species and local introgression of divergent alleles from the alternated species, respectively, have generated similar patterns of genomic islands for intraspecific divergence and between intraspecific and interspecific divergence. One explanation is that high gene flow within each species has probably overwhelmed the effect of linked selection and low recombination rate, and eroded the highly differentiated regions. It is also possible that incomplete lineage sorting at the early stage of speciation between the two diverging species has reshaped the genome-wide landscapes of divergence within each species (Hobolth et al., 2011). We can also not exclude the possibility that divergent alleles that are locally introgressed from alternated species are so strongly selected against by natural

selection that they are eliminated from each species. The slight overlapping of intraspecific genomic islands might be due to local selection from similar selective agents, although this would require continuous or frequent occurrences of selective sweeps at the same genomic regions between species (Vijay et al., 2017). Taken into consideration the overlapping of their distribution ranges, it is likely that the two species have evolved in parallelism under similar selective pressure, which explains the slight overlapping of intraspecific genomic islands.

5 | CONCLUSIONS

Integration of genome-wide variations with demographic history pave a way to disentangle the evolutionary processes shaping heterogeneous genome-wide divergence. We observed both extensive interspecific and intraspecific genomic islands across the studied species. While a proportion of the genomic islands have evolved due to linked selection shaped by variation in recombination rate across the genome before secondary contact, such pattern of heterogeneous genomic divergence has been reshaped by differential introgression following secondary contact. Discrimination of the factors underlying these genomic islands would help understand the genomic basis of speciation in organisms.

AUTHOR CONTRIBUTIONS

Le Wang and Zining Meng conceived and designed the study. Zining Meng supervised the whole study. Shufang Liu, Zhimeng Zhuang and Le Wang collected samples. Zining Meng, Shufang Liu, Le Wang and Yang Yang performed genome sequencing, RNA sequencing and RAD sequencing. Le Wang, Zining Meng, and Yang Yang annotated the genome and analysed the data. Le Wang, Zining Meng and Shufang Liu wrote the manuscript with input from the other authors. All authors discussed the results and approved the final version of the study.

ACKNOWLEDGEMENTS

We thank many anonymous fishermen for their help in the collection of samples. We are grateful to anonymous reviewers for their insightful comments and suggestions, which have significantly improved this study. This work was supported by the Marine S & T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (2018SDKJ0501-2), the National Natural Science Foundation of China (31172411), and Central Public-interest Scientific Institution Basal Research Found (20603022020005).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The genome sequences and all raw sequencing reads for this study have been archived in the National Genomics Data Center of China (NGDC) and China National Gene Bank (CNGB) database with

Bioproject accession numbers of PRJCA003706 and CNP0001482, respectively. Genome sequences and SNP datasets would be archived in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.kwh70rz6j>).

ORCID

Le Wang  <https://orcid.org/0000-0002-3363-5753>

Shufang Liu  <https://orcid.org/0000-0003-3766-2932>

Yang Yang  <https://orcid.org/0000-0001-6904-0404>

Zining Meng  <https://orcid.org/0000-0002-5170-9613>

Zhimeng Zhuang  <https://orcid.org/0000-0002-7108-0824>

REFERENCES

- Alexander, D. H., & Lange, K. (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics*, 12, 246.
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11, 697–709.
- Ao, J., Mu, Y., Xiang, L.-X., Fan, D., Feng, M., Zhang, S., Shi, Q., Zhu, L.-Y., Li, T., Ding, Y., Nie, L., Li, Q., Dong, W., Jiang, L., Sun, B., Zhang, X., Li, M., Zhang, H.-Q., Xie, S., ... Chen, X. (2015). Genome sequencing of the perciform fish *Larimichthys crocea* provides insights into molecular and genetic mechanisms of stress adaptation. *PLoS Genetics*, 11, e1005118.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 3, e3376.
- Beaumont, M. A., & Balding, D. J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969–980.
- Bierne, N., Gagnaire, P.-A., & David, P. (2013). The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Current Zoology*, 59, 72–86.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H. J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., ... di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513, 375–381.
- Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, 1, 118–131.
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružič, M., Sæther, S. A., Sætre, G. P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Research*, 25, 1656–1665.
- Butlin, R. K., & Smadja, C. M. (2018). Coupling, reinforcement, and speciation. *The American Naturalist*, 191, 155–172.
- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, 18, 188–196.
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, 1, 171–182.
- Chen, B., Zhou, Z., Ke, Q., Wu, Y., Bai, H., Pu, F., & Xu, P. (2019). The sequencing and de novo assembly of the *Larimichthys crocea* genome using PacBio and hi-C technologies. *Scientific Data*, 6, 188.
- Chen, F., Mackey, A. J., Stoeckert, C. J., Jr., & Roos, D. S. (2006). OrthoMCL-DB: Querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research*, 34, D363–D368.
- Crown, K. N., Miller, D. E., Sekelsky, J., & Hawley, R. S. (2018). Local inversion heterozygosity alters recombination throughout the genome. *Current Biology*, 28, 2984–2990.
- Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23, 3133–3157.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- De La Torre, A. R., Puiu, D., Crepeau, M. W., Stevens, K., Salzberg, S. L., Langley, C. H., & Neale, D. B. (2019). Genomic architecture of complex traits in loblolly pine. *New Phytologist*, 221, 1789–1801.
- Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remodeling of genomic islands of differentiation in the European sea bass. *Nature Communications*, 9, 2518.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. (2012). The genomic landscape of species divergence in *Ficedula flycatchers*. *Nature*, 491, 756–760.
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10, 564–567.
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28, 342–350.
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993.
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., & Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, 108, 1513–1518.
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186.
- Guerrero, R. F., & Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 26, 5362–5368.
- Guirao-Rico, S., Sánchez-Gracia, A., & Charlesworth, D. (2017). Sequence diversity patterns suggesting balancing selection in partially sex-linked genes of the plant *Silene latifolia* are not generated by demographic history or gene flow. *Molecular Ecology*, 26, 1357–1370.
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, 5, e1000695.
- Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, 27, 1004–1015.
- Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16, 730–737.
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105, 795–809.
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25, 2454–2466.

- Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8, 269–294.
- Hobolth, A., Dutheil, J. Y., Hawks, J., Schierup, M. H., & Mailund, T. (2011). Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Research*, 21, 349–356.
- Horiike, T., Minai, R., Miyata, D., Nakamura, Y., & Tatenno, Y. (2016). Ortholog-finder: A tool for constructing an ortholog data set. *Genome Biology and Evolution*, 8, 446–457.
- Ikeda, I. (1964). Studies on the fisheries biology of the yellow croaker in the East China and the yellow seas. *Seikai Region Fisheries Research Lab*, 31, 1–81.
- Jahner, J. P., Parchman, T. L., & Matocq, M. D. (2021). Multigenerational backcrossing and introgression between two woodrat species at an abrupt ecological transition. *Molecular Ecology*, 30, 4245–4258.
- Janoušek, V., Munclinger, P., Wang, L., Teeter, K. C., & Tucker, P. K. (2015). Functional organization of the genome may shape the species boundary in the house mouse. *Molecular Biology and Evolution*, 32, 1208–1220.
- Jiang, K., Tong, X., Ding, Y. Q., Wang, Z. W., Miao, L. Y., Xiao, Y. E., Huang, W. C., Hu, Y. H., & Chen, X. Y. (2021). Shifting roles of the East China Sea in the phylogeography of red nanmu in East Asia. *Journal of Biogeography*, 48, 2486–2501.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484, 55–61.
- Ke, F., Vasseur, L., Yi, H., Yang, L., Wei, X., Wang, B., & Kang, M. (2022). Gene flow, linked selection, and divergent sorting of ancient polymorphism shape genomic divergence landscape in a group of edaphic specialists. *Molecular Ecology*, 31, 104–118.
- Kong, S., Ke, Q., Chen, L., Zhou, Z., Pu, F., Zhao, J., Bai, H., Peng, W., & Xu, P. (2019). Constructing a high-density genetic linkage map for large yellow croaker (*Larimichthys crocea*) and mapping resistance trait against ciliate parasite *Cryptocaryon irritans*. *Marine Biotechnology*, 21, 262–275.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, 27, 722–736.
- Le Moan, A., Roby, C., Fraisse, C., Daguin-Thiébaud, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, 30, 6718–6732.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25, 1754–1760.
- Liu, B., Zhang, B., Xue, D., Gao, T., & Liu, J. (2016). Population structure and adaptive divergence in a high gene flow marine fish: The small yellow croaker (*Larimichthys polyactis*). *PLoS One*, 11, e0154020.
- Liu, S., Wu, R., Zhu, L., & Zhuang, Z. (2010). Complete sequence and gene organization of mitochondrial DNA of the small yellow croaker *Larimichthys polyactis*. *Oceanologia et Limnologia Sinica*, 41, 885–894.
- Liu, X. (1962). *The research of small yellow croaker (Larimichthys polyactis) geographic race and gonad*. Paper presented at the Proceedings of the Seventh Meeting of the Western Pacific Fisheries Research Council.
- Liu, X. (1964). *The study on biological measurement of population of small yellow croaker*. The proceedings of marine fisheries resources. Agriculture Press.
- Liu, X., Glémin, S., & Karrenberg, S. (2020). Evolution of putative barrier loci at an intermediate stage of speciation with gene flow in campions (*Silene*). *Molecular Ecology*, 29, 3511–3525.
- Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., Jiang, D., Ahani, H., Abbott, R. J., Lascoux, M., Nevo, E., & Liu, J. (2018). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences*, 115, E236–E243.
- Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an east African crater lake. *Science*, 350, 1493–1498.
- Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biology*, 17, e2006288.
- Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, 47, 69–74.
- McVean, G. A., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., & Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science*, 304, 581–584.
- Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: Theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 409–421.
- Nevado, B., Contreras-Ortiz, N., Hughes, C., & Filatov, D. A. (2018). Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. *New Phytologist*, 219, 779–793.
- Nosil, P. (2008). Speciation with gene flow could be common. *Molecular Ecology*, 17, 2103–2106.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18, 375–402.
- O'Connell, K. A., Prates, I., Scheinberg, L. A., Mulder, K. P., & Bell, R. C. (2021). Speciation and secondary contact in a fossorial Island endemic, the São Tomé caecilian. *Molecular Ecology*, 30, 2859–2871.
- Pazhenkova, E. A., & Lukhtanov, V. A. (2021). Genomic introgression from a distant congener in the levant fritillary butterfly, *Melitaea acentria*. *Molecular Ecology*, 30, 4819–4832.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575.
- Ravinet, M., Faria, R., Butlin, R., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30, 1450–1477.
- Renaut, S., Grassa, C., Yeaman, S., Moyers, B., Lai, Z., Kane, N., Bowers, J. E., Burke, J. M., & Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications*, 4, 1827.
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16, 351–358.
- Rougeux, C., Bernatchez, L., & Gagnaire, P.-A. (2017). Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake whitefish species pairs (*Coregonus clupeaformis*). *Genome Biology and Evolution*, 9, 2057–2074.
- Shahdadi, A., von Wyszczetki, K., Liu, H.-C., Chu, K. H., & Schubart, C. D. (2022). Molecular phylogeography reveals multiple Pleistocene divergence events in estuarine crabs from the tropical West Pacific. *PLoS One*, 17, e0262122.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31, 3210–3212.
- Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the mantel test of matrix correspondence. *Systematic Zoology*, 35, 627–632.

- Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., Schnable, P. S., Lyons, E., & Lu, J. (2015). ALLMAPS: Robust scaffold ordering based on multiple maps. *Genome Biology*, *16*, 3.
- Tavares, H., Whibley, A., Field, D. L., Bradley, D., Couchman, M., Copsey, L., Elleouet, J., Burrus, M., Andalo, C., Li, M., Li, Q., Xue, Y., Rebocho, A. B., Barton, N. H., & Coen, E. (2018). Selection and gene flow shape genomic islands that control floral guides. *Proceedings of the National Academy of Sciences*, *115*, 11006–11011.
- Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., ... Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, *5*, 5770.
- Turner, T. L., & Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, *19*, 848–850.
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*, e285.
- Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, *26*, 3982–3997.
- Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 451–460.
- Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology*, *26*, 4284–4295.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*, e112963.
- Wang, B., Mojica, J. P., Perera, N., Lee, C.-R., Lovell, J. T., Sharma, A., Adam, C., Lipzen, A., Barry, K., Rokhsar, D. S., Schmutz, J., & Mitchell-Olds, T. (2019). Ancient polymorphisms contribute to genome-wide variation by long-term balancing selection and divergent sorting in *Boechera stricta*. *Genome Biology*, *20*, 126.
- Wang, L., Liu, S., Zhuang, Z., Guo, L., Meng, Z., & Lin, H. (2013). Population genetic studies revealed local adaptation in a high gene-flow marine fish, the small yellow croaker (*Larimichthys polyactis*). *PLoS One*, *8*, e83493.
- Wang, L., Liu, S., Zhuang, Z., Lin, H., & Meng, Z. (2015). Mixed-stock analysis of small yellow croaker *Larimichthys polyactis* providing implications for stock conservation and management. *Fisheries Research*, *161*, 86–92.
- Wang, L., Shi, X., Su, Y., Meng, Z., & Lin, H. (2012). Loss of genetic diversity in the cultured stocks of the large yellow croaker, *Larimichthys crocea*, revealed by microsatellites. *International Journal of Molecular Sciences*, *13*, 5584–5597.
- Wang, L., Shi, X., Su, Y., Meng, Z., & Lin, H. (2013). Genetic divergence and historical demography in the endangered large yellow croaker revealed by mtDNA. *Biochemical Systematics and Ecology*, *46*, 137–144.
- Wang, L., Sun, F., Wan, Z. Y., Yang, Z., Tay, Y. X., Lee, M., Ye, B., Wen, Y., Meng, Z., Fan, B., Alfiko, Y., Shen, Y., Piferrer, F., Meyer, A., Scharl, M., & Yue, G. H. (2022). Transposon-induced epigenetic silencing in the X chromosome as a novel form of dmrt1 expression regulation during sex determination in the fighting fish. *BMC Biology*, *20*, 5.
- Wang, L., Sun, F., Wan, Z. Y., Ye, B., Wen, Y., Liu, H., Yang, Z., Pang, H., Meng, Z., Fan, B., Alfiko, Y., Shen, Y., Bai, B., Lee, M. S. Q., Piferrer, F., Scharl, M., Meyer, A., & Yue, G. H. (2021). Genomic basis of striking fin shapes and colors in the fighting fish. *Molecular Biology and Evolution*, *38*, 3383–3396.
- Wang, L., Wan, Z. Y., Lim, H. S., & Yue, G. H. (2016). Genetic variability, local selection and demographic history: Genomic evidence of evolving towards allopatric speciation in Asian seabass. *Molecular Ecology*, *25*, 3605–3621.
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*, 87–100.
- Wu, C., Zhang, D., Kan, M., Lv, Z., Zhu, A., Su, Y., Zhou, D., Zhang, J., Zhang, Z., Xu, M., Jiang, L., Guo, B., Wang, T., Chi, C., Mao, Y., Zhou, J., Yu, X., Wang, H., Weng, X., ... Liu, Y. (2014). The draft genome of the large yellow croaker reveals well-developed innate immunity. *Nature Communications*, *5*, 5227.
- Xiao, Y., Zhang, Y., Gao, T., Yanagimoto, T., Yabe, M., & Sakurai, Y. (2009). Genetic diversity in the mtDNA control region and population structure in the small yellow croaker *Larimichthys polyactis*. *Environmental Biology of Fishes*, *85*, 303–314.
- Zhai, S. N., Comes, H. P., Nakamura, K., Yan, H. F., & Qiu, Y. X. (2012). Late Pleistocene lineage divergence among populations of *Neolitsea sericea* (Lauraceae) across a deep sea-barrier in the Ryukyu Islands. *Journal of Biogeography*, *39*, 1347–1360.
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*, 1523.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, L., Liu, S., Yang, Y., Meng, Z., & Zhuang, Z. (2022). Linked selection, differential introgression and recombination rate variation promote heterogeneous divergence in a pair of yellow croakers. *Molecular Ecology*, *31*, 5729–5744. <https://doi.org/10.1111/mec.16693>