


Sympatric genetic divergence between early- and late-season weedy rice populations

Zhi Wang¹, Xingxing Cai¹, Xiao-Qi Jiang¹, Qi-Yu Xia², Lin-Feng Li¹ and Bao-Rong Lu¹ 

¹Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, Department of Ecology and Evolutionary Biology, Fudan University, Songhu Road 2005, Shanghai 200438, China; ²Key Laboratory of Biology and Genetic Resources of Tropical Crops, Hainan Key Laboratory for Biosafety Monitoring and Molecular Breeding in Off-Season Reproduction Regions, Institute of Tropical Bioscience and Biotechnology, CATAS, Haikou 571101, China

Summary

Author for correspondence:
Bao-Rong Lu
Email: brlu@fudan.edu.cn

Received: 26 December 2021
Accepted: 13 May 2022

New Phytologist (2022) **235**: 2066–2080
doi: 10.1111/nph.18288

Key words: divergent selection, ecological speciation, genetic differentiation, *Oryza sativa*, restricted gene flow, sympatry, temporal isolation, weed.

- Sympatric genetic divergence is the most appealing and controversial pattern in the theory of ecological speciation. Examples that support sympatric genetic divergence in plant species are extremely rare. Solid evidence of sympatric genetic divergence will provide deep insights for revealing the underlying mechanisms of ecological speciation.
- We analysed the total genomic DNA sequences of 120 weedy rice (WR; *Oryza sativa* f. *spontanea*) plants, representing three WR population pairs separately from three early- and late-season rice fields, in comparison with those of the co-occurring rice cultivars and other rice materials.
- We detected substantial genetic divergence within the pairs of the sympatric early- and late-season WR populations, although genetic divergence was unevenly distributed across the genomes. Restricted gene flow was determined between the sympatric WR populations, resulting in their distinct genetic structures. We also detected relatively low genetic diversity that was likely to be associated with stronger selection in early-season WR populations.
- Our findings provide strong evidence for sympatric genetic divergence between the WR populations in the same fields but in different seasons. We conclude that temporal isolation plays an important role in creating genetic divergence between sympatric populations/species in plants.

Introduction

Ecological speciation, which denotes a speciation process in which ecologically based divergent selection drives the formation of reproductive barriers between populations/species, plays a very important role in creating novel taxa under the contemporary concept of the evolutionary biology (Nosil, 2012). Field studies and laboratory experiments have provided a considerable amount of evidence to prove the frequent occurrences of ecological speciation worldwide (Schluter, 2009; Peccoud & Simon, 2010; Lassalle *et al.*, 2015; Meyer *et al.*, 2016). It is generally recognised that two key elements are involved in the ecological speciation process, namely, the divergent selection and reinforcement (Schluter, 2001). The former element acts as an initiative that creates genetically divergent populations from an incipient population in contrasting directions under differential environmental conditions, which is the key to start the process of ecological speciation (Nosil *et al.*, 2009b). Generally, differences in the environments or ecological niches, including climates, resources and habitat structures, serve as the common sources of divergent selection between populations in nature (Nosil, 2012). The latter element acts as a force to create and strengthen reproductive isolations between populations through the secondary contacts, which

results in the accomplishment of the ecological speciation process (Servedio & Noor, 2003).

The key point to differentiate the ecological speciation from other patterns of speciation involves the identification of whether the divergent selection process has taken place, which creates the reproductive isolations between populations. Therefore, generating knowledge associated with the underlying mechanisms of genetic divergence will provide deep insights into divergent selection that drives ecological speciation. To date, two main patterns of genetic divergence are recognised in the model of ecological speciation: allopatric and sympatric divergence. Allopatric divergence, denoting genetic divergence between spatially isolated populations/species, is the most common and plausible pattern of genetic divergence (Schluter, 2001; Nosil, 2012). However, sympatric divergence, indicating genetic divergence of an ancestral population/species into two or more divergent populations without geographical or spatial isolation, is a pattern occurring much infrequent in nature (Nosil, 2012). The pattern of sympatric divergence has been challenged because of its controversy in terms of genetic divergence under a certain level of gene flow (Butlin *et al.*, 2008; Fitzpatrick *et al.*, 2008). Therefore, it is very important to carry out reliable investigation with solid evidence to support the controversial pattern of sympatric divergence.

The complication or difficulty to explain the sympatric divergence pattern is that genetic divergence occurs within an incipient population or species under the existence of gene flow (Butlin *et al.*, 2008; Bird *et al.*, 2012; Nosil, 2012). However, examples of sympatric divergence were reported in nature, particularly based on the studies of animal species (Via, 2001; Li *et al.*, 2015; Marques *et al.*, 2016; Kautt *et al.*, 2020), meaning that genetic divergence could occur at the same site where gene flow exists hypothetically. A compelling case is the discovery of cichlid fish in small isolated crater lakes (Barluenga *et al.*, 2006; Malinsky *et al.*, 2015; Fruciano *et al.*, 2016; Kautt *et al.*, 2020), in which the assortative mating between ecotypes arose initially as a simple by-product of divergent natural selection on key ecological traits, such as coloration, lip size, pharyngeal jaw morphology and body shape. Another case is the report in the blind mole rat in the evolution Canyon microsite (Li *et al.*, 2015), where the evolution of soil preference accounted for the sympatric divergence between the abutting basalt and chalk soil populations. By contrast, solid evidence of sympatric divergence is extremely rare in plant species. The reported two *Howea* palm species evolved with divergent flowering time associated with the acidity of the soil on Lord Howe Island is considered as one of the most convincing examples of sympatric divergence in plant species (Savolainen *et al.*, 2006). However, the authenticity of these sympatric species has been challenged, concerning the origins of the palm species (single or multiple origins), which is unclear due to the geographical changes of the islands (Stuessy, 2006). Therefore, it is necessary to obtain solid and reliable evidence in plants to prove the current theory of sympatric divergence by studying the recently diverged populations, ideally occurring in exactly the same spatial locations.

Weedy rice (WR; *Oryza sativa* f. *spontanea*, also referred to as red rice) is a noxious weed infecting rice (*O. sativa*) fields worldwide and causing tremendous losses of rice grain yield and commercial quality (Delouche *et al.*, 2007). As a conspecific taxon, WR evolves rapidly under the influence of its cultivated counterparts and human activities (Jiang *et al.*, 2012; Zhu *et al.*, 2021). In some tropical regions in the world, rice is grown for two seasons per year, namely the early and late seasons, when WR also co-occurs in the two seasons' rice fields (Li *et al.*, 2013; Kong *et al.*, 2021). In southern China, rice production is characterised by intensive farming by small farmer households with a rapid shift from traditional seedling transplanting to direct seeding, in addition to less field management due to the lack of a labour force, which causes serious weed problems, including by WR. Agricultural activities, including seed sowing, weeding, and harvesting, are still largely implemented manually in some areas, although mechanical harvesting has now become popular. For example, mechanical harvesters have gradually been applied in Leizhou of Guangdong Province for rice harvesting over the last 5–10 yr, but the harvesters owned by a community are usually applied within the community for its farmers' rice harvesting, rather than across different communities or large regions. Commonly, small farmer households own an average of <0.5-ha farmland (Supporting Information Fig. S1) and prefer to grow conventional rice varieties rather than hybrid varieties bought

from markets. This is because seeds of the conventional rice varieties can be reused in different seasons and years for reducing the costs, compared with the hybrid varieties, for which seeds need to be purchased with higher prices for every rice-cultivation season. Therefore, in some regions such as Leizhou, where we collected our WR and rice samples, farmers still retain the tradition method of saving their own rice seeds for planting in the two seasons.

Commonly, the environmental conditions between the early- and late-season rice fields are considerably different, including the ambient temperature, rainfall, photoperiod, and plant disease and insect pests (Li *et al.*, 2013). Such differences can serve as important sources for the divergent selection of plant taxa. A previous study indicated detectable increases in genetic diversity associated with the higher ambient temperature in the late-season WR populations collected from Leizhou based on the selected microsatellite fingerprints (Kong *et al.*, 2021). Results obtained from that study suggested potential genetic divergence between the early- and late-season populations of WR co-occurring in the same rice fields. Therefore, the early- and late-season WR populations in the same rice fields provide an ideal system with which to investigate sympatric genetic divergence in plants using a proper analytical method.

The rapid development of genomic sequencing technology has provided more opportunities for us to detect the detailed variation between populations at the genomic DNA level and, therefore, to investigate sympatric genetic divergence in depth (Foote, 2018; Richards *et al.*, 2019). In this study, we conducted comprehensive analyses of the total genomic DNA sequences of 120 WR samples (individuals) and their co-occurring rice varieties collected from three separated early- and late-season rice fields owned by different farmer households in Leizhou of Guangdong Province in China. The objectives of this study were to address the following questions: (1) Are the WR populations co-occurring in the early and late seasons of the same (sympatric) rice field diverged genetically? (2) What is the distribution pattern of such sympatric genetic divergence across the entire genomes of WR? (3) Does gene flow exist between populations of early- and late-season WR, causing the alteration of genetic structures of these populations? (4) Does natural selection serve as the possible driving force for sympatric genetic divergence between the early- and late-season WR populations? The answers to these questions will provide deep insights, allowing us to better understand the underlying mechanisms of sympatric genetic divergence in plant species/populations associated with the theory of ecological speciation.

Materials and Methods

Plant materials

The mature panicles from a total of 120 WR (*O. sativa* L. f. *spontanea* Roshev.; Fig. S2) individuals, representing six WR populations, and their six co-occurring rice varieties (CoCV) cultivated in the early and late seasons of three separate rice fields (each was c. 3000 m² in size) were collected in the Chidou (CD), Hejia

(HJ) and Dongcun (DC) villages in Leizhou, Guangdong Province of China (Table S1). Distances between collected samples in a field were > 10 m to avoid sampling the same genotypes. Spatial distances between the three rice fields in the CD, HJ and DC villages were > 30 km. The WR populations sampled from the early (E) and late (L) seasons in the same rice fields were referred to as population pairs, namely CDE/L, HJE/L and DCE/L population pairs, whereas the co-occurring rice varieties sampled from the early and late seasons were coded as CDE/L-CoCV, HJE/L-CoCV and DCE/L-CoCV (Table S1). In addition, to understand the relationships between the early and late seasons' CoCV, as well as with rice varieties from other sources, the total genomic data of 14 rice varieties currently used in Guangdong Province, the typical *indica* variety (Peiai64) and typical *japonica* variety (Nipponbare), and a common wild rice (*Oryza rufipogon* Griff.) downloaded from the public database were also included as references for analyses (Table S1).

Genomic resequencing and variants calling

Genomic DNA was extracted separately from 30-d-old leaves generated from germinated seeds of the 120 WR individuals and six CoCV samples, using the DNeasy Plant Mini kits (Qiagen) and following the manufacturer's instructions. DNA libraries were constructed by the Novogene Co. (Beijing, China) and sequenced on the Illumina HiSeq 2000 platform using the paired end 150 (PE150) strategy. The whole genome of each individual was sequenced with an average > 15× genome depth and *c.* 90% genome coverage.

Raw reads of the 120 WR individuals and the six CoCV samples were assessed for quality control using FASTQC software (<https://www.bioinformatics.babraham.ac.uk/>). The rice reference genome of Nipponbare (*japonica*; MSU7.0 assembly) was downloaded from the Rice Genome Annotation Project website (<http://rice.plantbiology.msu.edu/>). The clean reads (base quality > 20) were mapped onto the Nipponbare reference genome using BURROWS-WHEELER ALIGNER (BWA) software (Li & Durbin, 2009) with the '--mem' option and default parameters. Raw variants (single nucleotide polymorphisms (SNP) and insertions and deletions (InDels)) were called using SAMTOOLS v.0.1.19 (Li *et al.*, 2009) with parameters 'mpileup --Dsqf' and 'bcftools view --Ncvg', whereas InDels were realigned with the INDELREALIGNER v.2.6 package in the Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). Finally, 6092 323 raw variants (SNPs and InDels) were obtained across the 120 WR individuals and the six CoCV samples; whereas 2051 486 SNPs were obtained after filtering in the VCFTOOLS v.0.1.17 (Danecek *et al.*, 2011) with the general filtering criterion '--minDP 3 --minQ 30 --maf 0.05 --max-missing 0.8 --remove-indel'.

Gene annotation and ontology enrichment

The rice gene annotations of Nipponbare (*japonica*; MSU7.0 release) were downloaded from the Rice Genome Annotation Project website (<http://rice.plantbiology.msu.edu/>). Candidate SNPs in the subsequent analyses were annotated to the

Nipponbare gene annotations using both SNPSIFT v.5.0 (Cingolani *et al.*, 2012) with the option of 'intervals' and the BEDTOOLS v.2.29.0 (Quinlan & Hall, 2010) with the option of 'intersect', based on the positions of SNPs and gene locations. The list of previously cloned rice genes was downloaded from the funRiceGenes database (<https://funricegenes.github.io/>), and annotated gene loci in the subsequent analyses were aligned to the list to identify their detail biological functions.

Gene ontology (GO) enrichment was conducted in the online program AGRIGO v.2.0 (<http://systemsbiology.cau.edu.cn/agriGOv2/index.php>) with the tool for singular enrichment analysis (SEA) in *japonica* rice. The steps of GO enrichment were as follows: candidate gene loci in the downstream analyses with the type of MSU7.0 gene ID (TIGR) were selected as an input file; statistical tests were conducted with Fisher's exact test and multiple-test adjustment method of Yekutieli (FDR under dependency) at the significance level of 0.05; and the minimum number of mapping entries was set as 5. The detail GO information was obtained in the National Rice Data Center (<https://www.ricedata.cn/ontology/>, in Chinese).

Estimation of genetic divergence

The fixation index (F_{ST}), average number of nucleotide substitutions per site between populations (D_{XY}), and the number of private SNP (PSNP) in the early- and late-season WR populations were calculated to estimate genetic divergence between the two-season populations. The F_{ST} values were calculated using the VCFTOOLS with the parameter of '--weir-fst-pop', based on nonoverlapping 50-kbp sliding windows. The D_{XY} values were calculated using PERL scripts by the authors of this study. In addition, differences in SNP sites were compared between the early- and late-season WR populations of the same population pairs. The SNP sites that were only possessed by the early- or late-season WR populations were determined as private SNP (PSNP).

Analyses of highly divergent genomic regions

The highly divergent genomic regions (HDRs) were artificially defined as nonoverlapping 50-kbp sliding windows with the top 5% highest F_{ST} values between the early- and late-season WR populations. SNPs involved in the candidate HDRs were annotated to the rice gene annotations (MSU7.0 release) in the three population pairs, respectively, therefore the putative selected genes (PSGs) involved in the HDRs were identified. Then the PSGs were used in the GO enrichment analyses in the online program AGRIGO v.2.0.

Phylogenetic relationships and genetic structures

To improve the quality and reliability of SNPs used in the analyses, a more stringent conditions of filtering (--minDP 5, --minQ 40, --max-missing 1) was adopted, leading to the reduction of the total number of the reliable SNPs to 629 175.

PERL scripts were used to convert the variant call format (VCF), including all SNP sites of individuals, into the FASTA

format. Homozygous SNP sites in each WR individual and CoCV sample were selected to generate neighbour-joining phylogenetic trees using the maximum-likelihood method in MEGAX software (Kumar *et al.*, 2018), with 1000 replicates for bootstrap values. To determine the phylogenetic relationships of cultivated rice varieties used in the early and late seasons in the same rice fields, the co-occurring six rice varieties grown in the early- and late-season rice fields in Leizhou and 14 currently used rice varieties in other regions of Guangdong Province (coded as GD-C) were included for analysis, using the typical *indica* variety (Peiai64), *japonica* variety (Nipponbare), and the common wild rice (*O. rufipogon* Griff.) as references. To determine the phylogenetic relationships of WR with its cultivars, 40 WR individuals and their CoCV from each of the collected field sites (CD, HJ and DC) were included for analysis, using the typical *indica* variety (Peiai64), *japonica* variety (Nipponbare), and the common wild rice (*O. rufipogon* Griff.) as references.

In addition, the SNP sites of the 120 WR individuals were converted from the VCF into the bed format using PLINK for genetic structure analyses using the program ADMIXTURE (Alexander *et al.*, 2009) in the proper *K*-values. Linkage disequilibrium analysis was conducted in POPLDDECAY v.3.40 software (Zhang *et al.*, 2019), and the maximum genetic distance was set as 500 kbp.

Estimation of gene flow

The TREEMIX v.1.13 software (Pickrell & Pritchard, 2012) was used to reconstruct the maximum-likelihood (ML) tree of WR populations to estimate gene flow between populations. Stratified allele-frequency outputs from PLINK v.1.9 software (<http://www.cog-genomics.org/plink/1.9/>) were converted into proper format using the *plink2treemix.py* script included in the TREEMIX v.1.13 software. Gene flow between branches was evaluated in a stepwise likelihood procedure, searching the tree for an optimal placement of each subsequent migration event. The *plotting_funcs.R* script was used to visualise gene flow. In addition, the residual heatmap was used to display relationships between populations based on the real covariance and estimated covariance. In the heatmap, the relationships were illustrated by colour codes. Furthermore, the ABBA-BABA test (also called *D* statistic) was used to estimate the level of genomic introgression between populations in the software DSUITE v.0.4 (Malinsky *et al.*, 2021) with the function 'Dtrios'. The CoCV collected from the three rice fields were also included in analyses, using the wild *O. rufipogon* as the outgroup.

Analyses of selective sweeps

The selective-sweep analyses were conducted based on the neutrality test and diversity estimation. Tajima's *D*, Fu & Li's *D** and *F** values were used for the neutrality test. Tajima's *D* values were calculated in the VCFTOOLS with the parameter of '--TajimaD' in nonoverlapping 50-kbp sliding windows. The SNP sites in the 120 WR individuals were converted from VCF into PHYLIP format using the VCF2PHYLIP v.2.0 (<https://github.com/edgardomortiz/vcf2phyip>) to calculate the Fu & Li's *D** and *F**

values using the VARISCAN v.2.0 (Vilella *et al.*, 2005), with the parameters 'RunMode = 12, UseMuts = 1, FixNum = 0, and NumNuc = 4' in nonoverlapping sliding windows.

Nucleotide diversity (π), expected heterozygosity (H_E), and Watterson's estimator of genetic diversity (θ_w) were calculated to estimate genetic diversity. Nucleotide diversity (π) was calculated in VCFTOOLS with parameter '--window-pi', based on nonoverlapping 50-kbp sliding windows. SNP site expected heterozygosity (H_E) was calculated using PERL script uploaded by Ye *et al.* (2019), and Watterson's estimator of genetic diversity (θ_w) was calculated using VARISCAN v.2.0 software based on nonoverlapping sliding windows.

To identify possible selective sweeps in the early-season WR populations, the top 10% lowest Tajima's *D* values and the top 10% highest π ratio (π_{WRL}/π_{WRE}) were analysed, involving the nonoverlapping 50-kbp sliding windows across the genomes. In addition, a \log_2 transformation of π ratios was conducted to have proper data display. The candidate swept genomic regions shared by the early-season WR populations were selected for gene annotation and GO enrichment analyses.

Results

Phylogenetic relationships of rice varieties and WR individuals

A relatively close phylogenetic relationship was detected between the CoCV grown in the three sampling rice fields in Leizhou, particularly for the early- and late-season rice varieties grown in the same rice fields, compared with other reference rice varieties from Guangdong Province and wild rice (Fig. 1), based on the SNPs across the whole genomes. The two CoCVs grown in the early and late seasons from the same rice fields showed much closer relationships in the neighbour-joining tree than other varieties. Rice varieties including the typical *indica* and *japonica* varieties from other regions than Leizhou showed relatively distant relationships with the CoCVs from Leizhou. The wild *O. rufipogon* was distantly related to any of the rice varieties from Leizhou (Fig. 1). However, population divergence was found between the early- and late-season WR populations in the same rice fields (Fig. 2a–c). In addition, a relatively distant genetic relationship was detected between the two-season WR populations of the CDE/L and HJE/L population pairs (Fig. 2a,b), although the DCE/L population pair did not show such a distant genetic relationship (Fig. 2c).

Genetic divergence across genomes between early- and late-season WR populations

Substantial genetic divergence, as estimated by F_{ST} (fixation index), was detected between the early- and late-season WR populations occurring at the same sites (in the same field, Table 1), with a much higher level of divergence ($F_{ST} = 0.31$ and 0.36) in the CDE/L and HJE/L population pairs compared with that in the DCE/L population pair ($F_{ST} = 0.01$). The estimation was made based on SNPs in each of the nonoverlapping 50-kbp

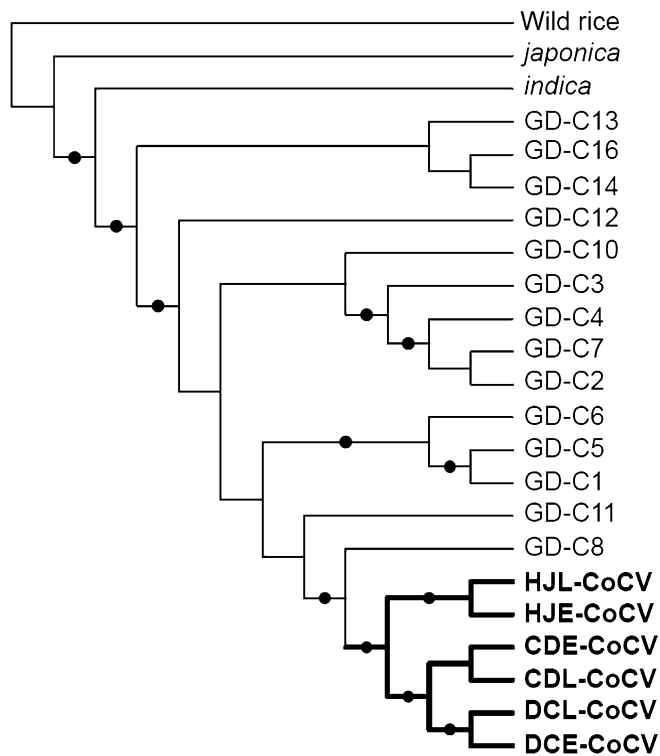


Fig. 1 A neighbour-joining tree indicating phylogenetic relationships among the co-occurring rice varieties (CoCVs) from the three sampling rice fields (as indicated by the bold text), currently available rice varieties in Guangdong Province (GD-C), the typical *indica* variety (Peiai64) and typical *japonica* variety (Nipponbare), and common wild rice (*Oryza rufipogon* Griff.). Black dots on the tree branches indicate the bootstrap values > 0.85 .

sliding windows across the whole WR genomes. In addition, the average number of nucleotide substitutions per site (D_{XY}) detected between populations also suggested greater genetic divergence in the CDE/L and HJE/L population pairs (Table 1) than in the DCE/L population pair. In addition, a large number of private SNP (PSNP) was detected in the early- and late-season WR populations, respectively, and a different number of PSNP was identified between the two-season WR populations in the CDE/L and HJE/L population pairs (Table 1). For example, only 42 170 PSNPs were identified in the CDE population and 233 351 PSNPs were identified in the CDL population. A similar situation was also found between the HJE and HJL populations (100 319 vs 267 846).

Noticeably, the nonrandom distribution pattern of genetic divergence was found across the WR genomes, judging by the unevenly distributed F_{ST} values on different chromosomes (Fig. 3a). In addition, the genomic regions with F_{ST} values > 0.3 were only identified in the CDE/L and HJE/L population pairs, but such highly divergent regions were not found in the DCE/L population pair (Fig. 3a). A similar distribution pattern, as estimated by the D_{XY} values, was also observed in these population pairs (Fig. 3b).

The 50-kbp sliding windows (regions) with the top 5% highest F_{ST} values estimated by SNPs between the two-season WR populations were artificially defined as the HDRs (cut-offs in Fig. 3a) in the subsequent analyses. As a result, 49 394, 61 212, and 72 790 SNPs were identified in the HDRs of the CDE/L, HJE/L, and DCE/L population pairs, respectively (Fig. 4a). In addition, 1478, 2172 and 2147 PSGs were annotated in the three

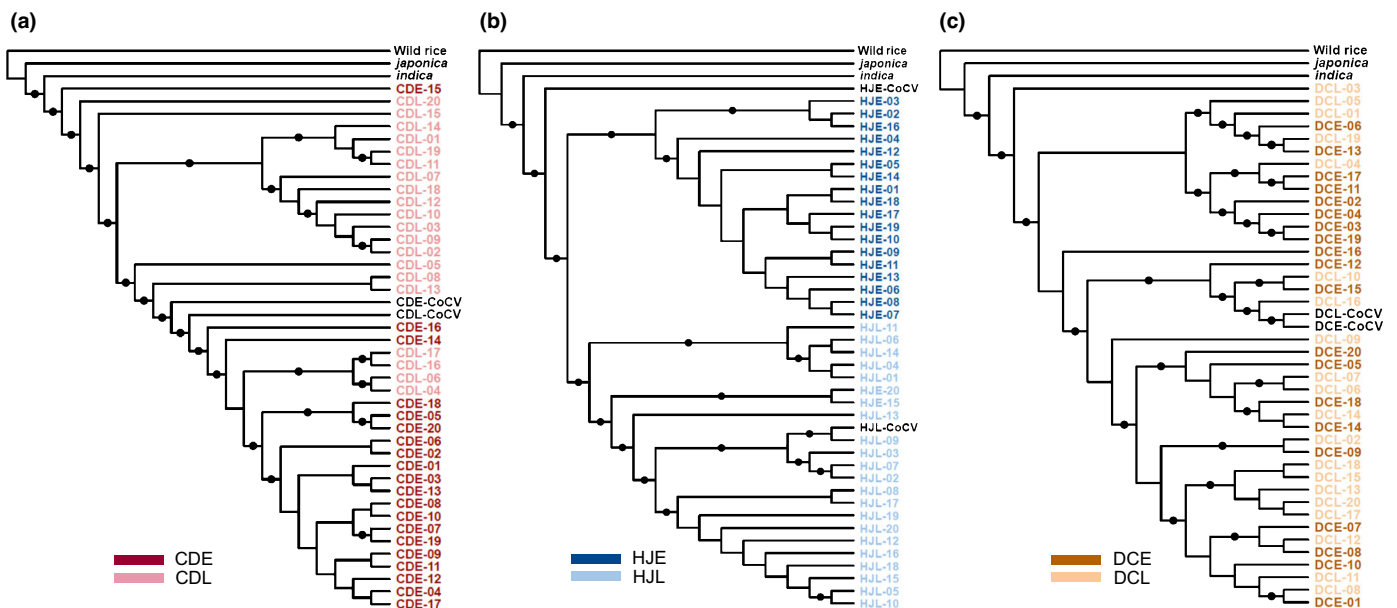


Fig. 2 Neighbour-joining trees indicating phylogenetic relationships among individuals in the early- and late-season weedy rice (WR) populations pairs, their co-occurring rice varieties (CoCV), and the reference group (*indica* variety Peiai64, *japonica* variety Nipponbare and wild *Oryza rufipogon* Griff.). (a) WR individuals (CDE and CDL) and their CoCV (CDE-CoCV and CDL-CoCV) from the Chidou village. (b) WR individuals (HJE and HJL) and their co-occurring rice varieties (HJE-CoCV and HJL-CoCV) from the Hejia village. (c) WR individuals (DCE and HJL) and their co-occurring rice varieties (DCE-CoCV and DCL-CoCV) from Dongcun village. Black dots on the tree branches indicate the bootstrap values > 0.85 .

Table 1 Estimating genetic divergence between the early- and late-season weedy rice populations using F_{ST} , D_{XY} and PSNP.

Population pairs	F_{ST} ¹	D_{XY} ²	PSNP ³
CDE/L ⁴	0.31	0.0014	42 170 233 351
HJE/L	0.36	0.0012	100 319 267 846
DCE/L	0.01	0.0010	92 931 71 175

¹ F_{ST} , F -statistic of genetic differentiation index (Weir & Cockerham, 1984).

² D_{XY} , average number of nucleotide substitutions per site between populations.

³PSNP, number of private SNP.

⁴CDE/L, HJE/L, DCE/L: the early- and late-season WR populations collected from Chidou (CD), Hejia (HJ) and Dongcun (DC), respectively.

population pairs (Fig. 4a), respectively. Interestingly, the 28 most enriched GO accessions were found to be shared among the three population pairs (Fig. 4a), and these PSGs were mainly enriched to functions related to metabolic process, biosynthesis process, and binding (Fig. 4b). However, no such expected HDRs, SNPs, and PSGs were shared among the three population pairs (Fig. S3). Furthermore, several functional genes related to flowering (e.g. *qHd1*, *Ehd3*, *OsFD1*), disease resistance (*OsWRKY30*, *OsLRR-RLK1*, *OsBRR1*), and abiotic stress responses (*OsMYB3R-2*, *OsGA2ox6*, *OsABF1*) were identified in the HDRs of the three population pairs (Table S2). These results indicated a similar pattern in terms of functional differentiation between the early- and late-season populations for these WR population pairs, even though these population pairs showed different levels of genetic

divergence. Noticeably, a much greater number of disease-resistance-associated genes was found in the DCE/L pair (Table S2), suggesting that the possibility of genetic divergence associated with functional differentiation began with disease resistance in the two-season populations.

Gene flow and genetic structures of the WR populations

Two gene flow events (as measured by migrations) were detected between WR populations and their CoCVs, based on the ML tree with zero migration edges and its associated residuals heatmap (Fig. 5a,b). In addition, the ML tree with the best fit migration events (Fig. 5c) and the associated residuals heatmap (Fig. 5d) consistently showed evident gene flow from HJ-CoCVs to the CDL WR population, and from the CDE WR population to the CD-CoCVs. However, no migration event was detected between the early- and late-season WR populations occurring either in the same or different rice fields (Fig. 5c). The ABBA-BABA test also showed significant introgression only between the HJL WR population and HJ-CoCVs and DC-CoCVs (Table S3). These results suggested an extremely limited, if any, gene flow between the early- and late-season WR populations in the same (sympatry) and different (allopatry) rice fields, compared with that between the WR populations and their CoCVs with a certain level of gene flow.

Accordingly, distinct genetic components were detected between the corresponding early- and late-season WR populations (e.g. CDE/L and HJE/L population pairs; Fig. 6) based on ADMIXTURE analyses of ancestry coefficients, when the best fit K -value was determined as 7 (Fig. 6a) and its neighbouring K -values

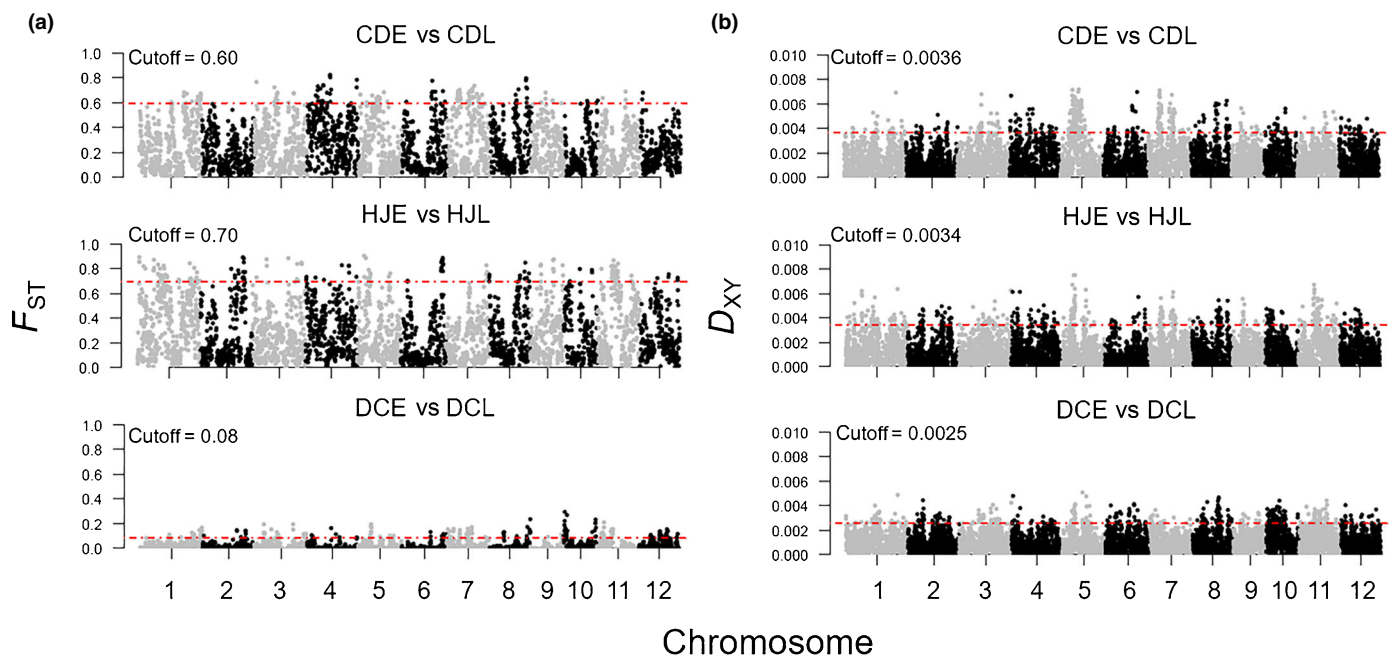


Fig. 3 Genome-wide genetic divergence between the two-season weedy rice populations. (a) Genetic divergence estimated by F_{ST} (Weir & Cockerham, 1984), the horizontal lines indicate the cut-offs of the top 5% highest F_{ST} values in each population pair. (b) Genetic divergence estimated by D_{XY} . D_{XY} , the average number of nucleotide substitutions per site. The genetic divergence was calculated based on 50-kbp nonoverlapping sliding windows across the whole genome, and each dot corresponds to one 50-kbp sliding window.

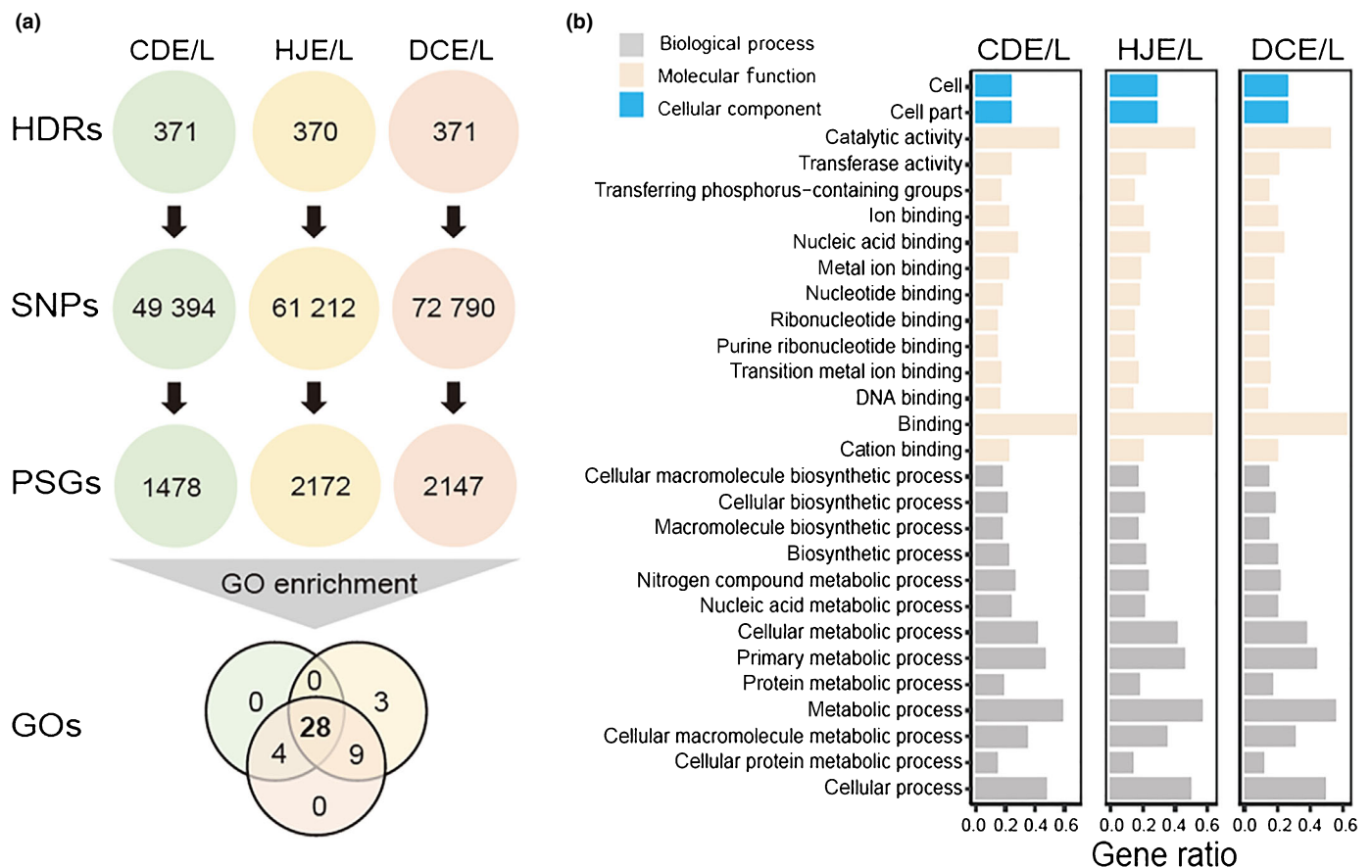


Fig. 4 Analyses of highly divergent genomic regions (HDRs) in weedy rice population pairs. (a) Candidate HDRs, SNPs and putative selected genes (PSGs) based on the top 5% highest F_{ST} values in CDE/L, HJE/L and DCE/L population pairs. (b) Significantly enriched gene ontology (GO) accessions based on PSGs shared by CDE/L, HJE/L and DCE/L population pairs.

as 6 (Fig. 6b) were also included as a comparison. Results showed a major genetic component that dominated the CDE or HJE population, but another genetic component that dominated its corresponding CDL or HJL population, suggesting considerable genetic divergence between the two-season populations. However, different patterns of mixed genetic components were found in the DCE/L population pair, suggesting its less genetic divergence. Furthermore, simpler genetic components were identified in the early-season populations (e.g. CDE and HJE) than those in the late-season populations. These results also indicated a detectable lower level of genetic polymorphisms in the early-season WR populations than their corresponding late-season populations.

Association of more pronounced selection with relatively low genetic diversity in early-season WR populations

The highest level of linkage disequilibrium (as indicated by r^2) values was detected in the two early-season CDE and HJE populations (Fig. 7). This result indicated that the two early-season populations were under a greater linkage disequilibrium selection than the two late-season CDL and HJL populations that showed a low linkage disequilibrium level, although the less

differentiated DCE/L population pair also did not show considerable differences in the disequilibrium level (Fig. 7). In addition, negative values (i.e. -0.02 , -1.92 and -0.18 in CDE; -0.18 , -0.21 and -0.59 in HJE) of the D - and F -statistics were also detected in the two early-season WR populations (CDE and HJE), based on the Tajima's and Fu & Li's neutrality test (Table 2). Clearly, more regions showed negative values across the entire genomes of the two early-season populations (CDE and HJE) compared with those of their corresponding late-season populations (CDL and HJL), based on the density distribution spectrum of the Tajima's D values (Fig. 8a) and Fu & Li's D^* and F^* values (Fig. 8b,c). By contrast, positive values (e.g. 1.58 , 0.68 and 1.45 in CDL; 1.45 , 0.68 and 0.80 in HJL) of D - and F -statistics were obtained in their corresponding late-season WR populations (CDL and HJL), with more genome regions showing positive values of D - and F -statistics. An increased proportion of rare SNPs with a low (close to 0) or high (close to 1) allele frequency was also detected in the CDE and HJE populations (Fig. 8d), based on the estimated SNP site frequency. In the DCE/L population pair, no such differences between the early- and late-season populations were detected. Altogether, the above results indicated a more pronounced natural selection in the early-season populations.

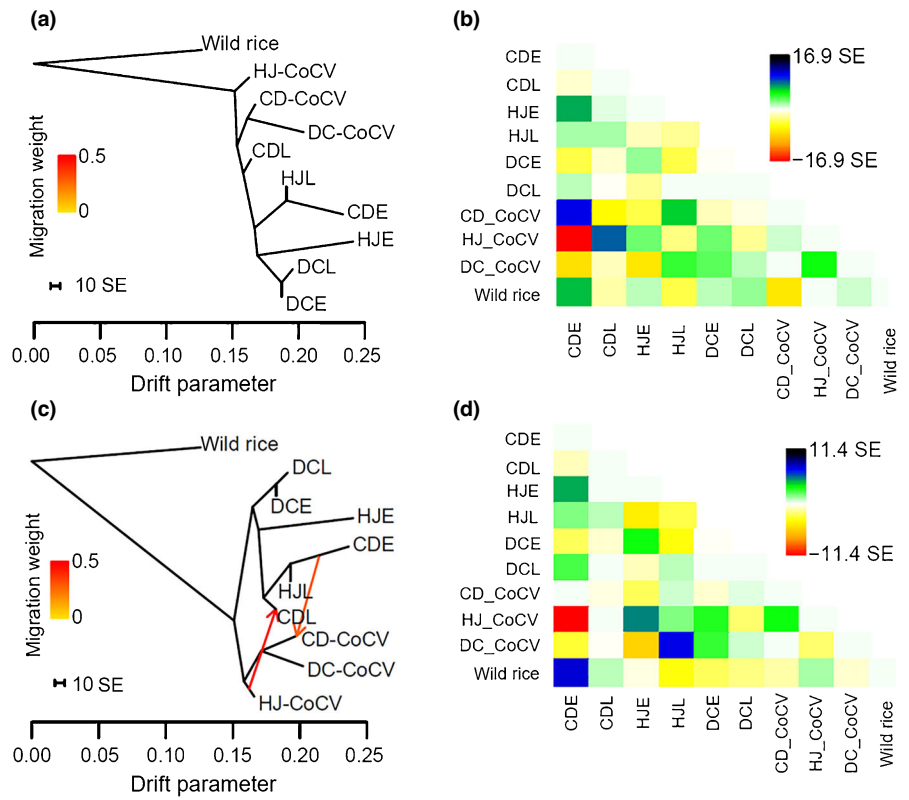


Fig. 5 Inferred migration events by TREEMIX. (a) The maximum-likelihood tree of weedy rice (WR) populations and their co-occurring rice varieties (CoCV) with zero migration event, and (b) its associated residual heatmap. (c) The maximum-likelihood tree of WR populations and their CoCV with two migration events that best fits the data and (d) its associated residual heatmap. The scale bar shows 10 times the average standard error (2SE) of the entries in the sample covariance matrix. Heatmap colours indicate residual covariance between each population divided by the average SE across all populations, and the residuals above the white colour indicate populations more closely related to each other than expected under the maximum-likelihood tree.

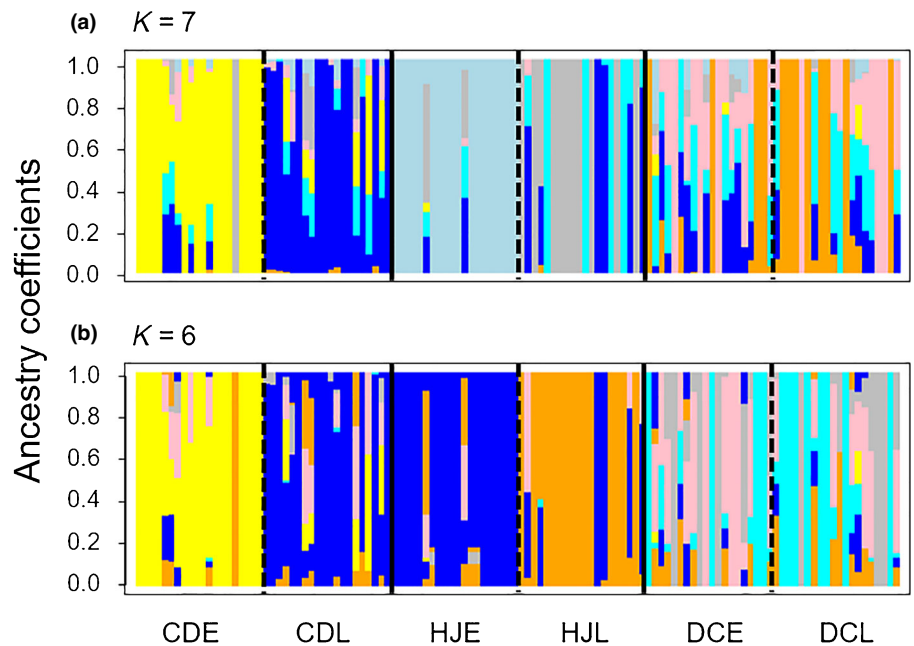


Fig. 6 ADMIXTURE analyses of individual ancestry coefficients in weedy rice populations. (a) The best fit K -value of 7, and (b) its neighbouring K -value of 6. The numbers along the vertical axis represent the individual ancestry coefficients (range from 0 to 1). Population pairs are isolated by black solid lines, and black dotted lines in each population pair separate the early- and late-season weedy rice populations. Different colours represent ancestral genetic components.

Noticeably, a relatively lower level of genetic diversity was also detected in the early-season populations (CDE and HJE) compared with in their corresponding late-season populations (CDL and HJL), as estimated based on the values of nucleotide diversity (Fig. 9a; Table 2), expected heterozygosity (H_E) (Fig. 9b; Table 2) and Watterson's estimator (θ_w) (Fig. 9c; Table 2). All the results demonstrated a relatively low level of genetic diversity in the early-season WR populations. This low genetic diversity

was probably associated with stronger selective sweeps on the certain genomic regions.

Results further indicated that 283 and 335 selective-sweep regions (SSRs) were identified in the CDE and HJE populations, respectively, based on analysis of the genome-wide Tajima's D values and the ratios of nucleotide diversity (π_{WR1}/π_{WR2}) (Fig. 10). In addition, many genes were annotated in the CDE and HJE populations, respectively, based on the SNPs involved

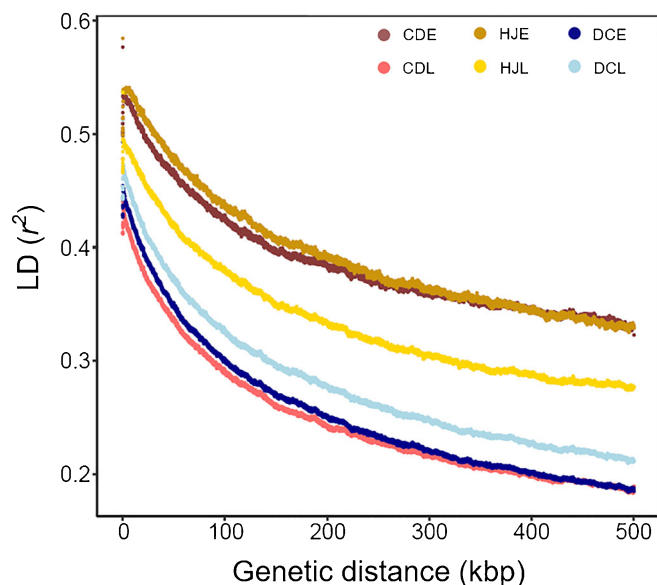


Fig. 7 Linkage disequilibrium (LD) decay in weedy rice populations. The horizontal axis represents genetic distance, and the maximum genetic distance is determined as 500 kbp. The vertical axis represents the level of disequilibrium, as indicated by r^2 values. Weedy rice populations (CDE, CDL, HJE, HJL, DCE and DCL) are indicated by different colours.

in the SSRs, and the 40 most enriched GO accessions were obtained in both CDE and HJE, of which 37 GO accessions were shared between CDE and HJE (Fig. S4a). The shared GO accessions were mainly related to functions related to metabolic process, binding and catalytic activity (Fig. S4b). These results not only indicated the occurrence of historic selective sweep on certain regions of the WR genomes in the early-season WR populations, but also suggested that the genes involved in the regions of selective sweep shared the same functions in the early-season WR populations from different rice fields. For example, several unique functional genes related to heading date regulation were identified in the SSRs of CDE (*Ehd1*, *Ehd3*, *OsWDR5a* and *OsCCT11*) and HJE (*Hd3a*, *Ehd2*, *RFT1*, *OsMADS18* and *OsABF1*). In addition, 21 previously cloned functional genes (Table S4) were identified to be shared by the CDE and HJE populations among the shared 208 PSGs that were associated with different traits, such as rice growth, grain features, cold stress tolerances and other abiotic stresses responses. This result indicated that these genes and the selected regions were possibly involved in the selection of WR growth/development and biotic/abiotic stress responses.

Discussion

Evident sympatric genetic divergence between early- and late-season WR populations

In our study, a relatively close phylogenetic relationship was detected between the early- and late-season rice varieties that co-occurred with their WR populations in the same fields (same sampling sites) in Leizhou, compared with the reference rice varieties from other sources (Fig. 1). Based on the above results and

Table 2 Results of neutrality test and genetic diversity in the early- and late-season weedy rice populations.

Population	Neutrality test			Genetic diversity		
	Tajima's D	Fu & Li's D^*	Fu & Li's F^*	π	H_E	θ_w
CDE	-0.02	-1.92	-2.28	0.0005	0.21	0.15
CDL	1.58	0.68	0.67	0.0010	0.33	0.20
HJE	-0.18	-0.21	-0.59	0.0005	0.23	0.14
HJL	1.45	0.68	0.80	0.0009	0.33	0.21
DCE	1.42	0.93	1.02	0.0010	0.33	0.22
DCL	0.70	0.43	0.43	0.0009	0.30	0.21

Tajima's D , Tajima's neutrality test (Tajima, 1989); Fu & Li's D^*/F^* , Fu and Li's neutrality test (Fu & Li, 1993); π , nucleotide diversity (Nei & Li, 1979); H_E , SNP site expected heterozygosity (Nei, 1978); θ_w , Watterson's estimator of genetic diversity (Watterson, 1975). CDE/CDL, HJE/HJL, DCE/DCL: the early- and late-season weedy rice populations collected from Chidou (CD), Hejia (HJ) and Dongcun (DC), respectively.

field interviews with the Leizhou's local farmers who usually cultivated the same rice variety in the early and late rice planting seasons, we considered that the CoCV at the same sampling field sites were extremely closely related. Also, given the common practices in southern China (Song *et al.*, 2021) that the small farmer households save their own harvested rice seeds for the next rice planting seasons and years, seed-mediated gene flow should be extremely rare between the three sampling field sites. In addition, our study also showed a relatively close phylogenetic relationship between the CoCV and their early- and late-season WR populations sampled from the same rice fields in Leizhou, compared with the reference rice materials, including the typical *indica* and *japonica* rice varieties, and wild *O. rufipogon* (Fig. 2a–c). It is therefore reasonable to compare sympatric genetic divergence between the early- and late-season WR populations from the three sampling fields, using their respective CoCV as a reference. In addition, the results also suggested that WR populations from the sampling field in Leizhou were distantly related to the wild *O. rufipogon* and were *indica* types, as reported previously by Song *et al.* (2015) and Wang *et al.* (2019).

Substantial genetic divergence was detected across the entire genomes between the early- and late-season WR populations occurring in the same rice fields, as estimated by the F_{ST} , D_{XY} and PSNP values. This finding provides strong evidence to support sympatric genetic divergence in plants. Noticeably, more pronounced genetic divergence was detected in the CDE/L and HJE/L population pairs compared with that in the DCE/L population pair, which is probably associated with the history of the WR populations that infested the cultivated rice fields at different time periods. Li *et al.* (2013) and Zhang *et al.* (2012) reported the emergence of WR in rice fields in different years in China, and supported the infestation histories of different WR populations in the studied region. Schluter (2001) proposed the concept of speciation continuum in the ecological speciation process, the different levels of genetic divergence between different pairs of the early- and late-season WR populations detected in our study may exhibit the 'speciation continuum' status of the WR

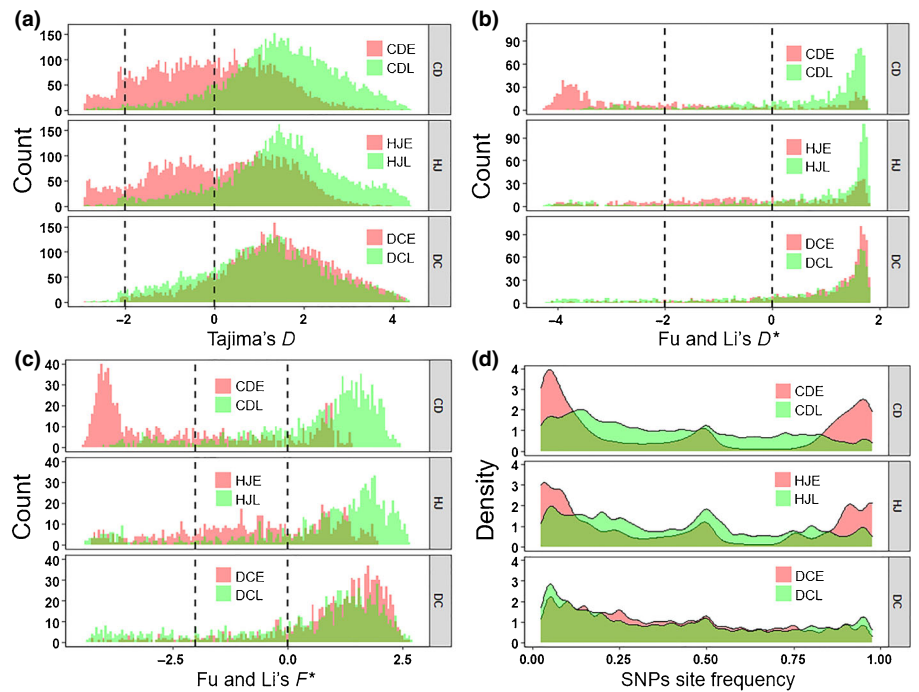


Fig. 8 Neutrality test in weedy rice populations. (a) Density distribution of Tajima's D values across the whole genome based on 50-kbp sliding windows. (b) Frequency density spectrum of Fu & Li's D^* . (c) Frequency density spectrum of Fu & Li's F^* . (d) SNPs site frequency density spectrum. The vertical dotted lines in (a–c) represent $x = 0$ and $x = -2$. The early- and late-season weedy rice populations are indicated by the red and green colours, respectively.

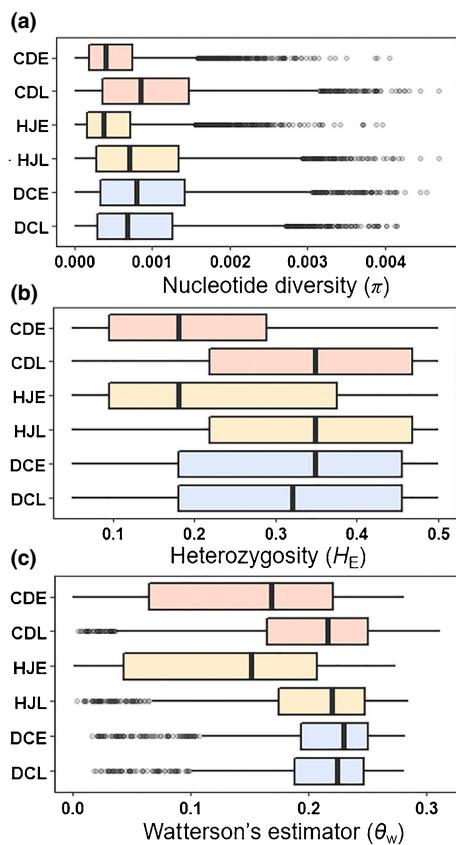


Fig. 9 Genetic diversity of weedy rice populations. (a) Nucleotide diversity, π ; (b) expected heterozygosity, H_E ; (c) Watterson's estimator of genetic diversity, θ_w . The upper and lower quartile values, median, outliers and whiskers are included in the boxplots.

population pairs. Nevertheless, the different levels of genetic divergence reflected that the population pairs of WR occurring in the same rice fields had emerged from the same ancestors at different times. To our knowledge, this is the first report of sympatric genetic divergence between plant populations occurring at exactly the same spatial sites (in the same rice fields with $c. 3000 \text{ m}^2$ for each population pair), although in different seasons.

In addition, genetic divergence between the sympatric WR populations was not evenly distributed across the entire genomes. In other words, some genomic regions of the WR populations were substantially diverged compared with other regions. Similar patterns of such uneven genetic divergence were also reported in other sympatric species (Kautt *et al.*, 2020; Li *et al.*, 2020; Wang *et al.*, 2020). All these findings together supported the reported 'heterogenous genomic divergence' model between sympatric populations or species (Nosil *et al.*, 2009a; Foote, 2018; Richards *et al.*, 2019).

Coincidentally, Kong *et al.* (2021) detected increased genetic diversity in late-season WR populations in Leizhou based on microsatellite (SSR) fingerprints, suggesting genetic differentiation between the two-season WR populations. The authors pointed out that increased genetic diversity in the late-season WR populations was associated with the higher ambient temperature in the late seasons (Kong *et al.*, 2021). In addition, our unpublished results also indicated a close association between phenotypical differentiation of the early-/late-season WR populations in Leizhou and ambient temperature and photoperiod variation. Noticeably, the genomic annotation results of variations from this study also indicated some highly divergent genomic regions/sites that were associated with resistance to stresses, such as diseases and abiotic factors, and change in flowering time

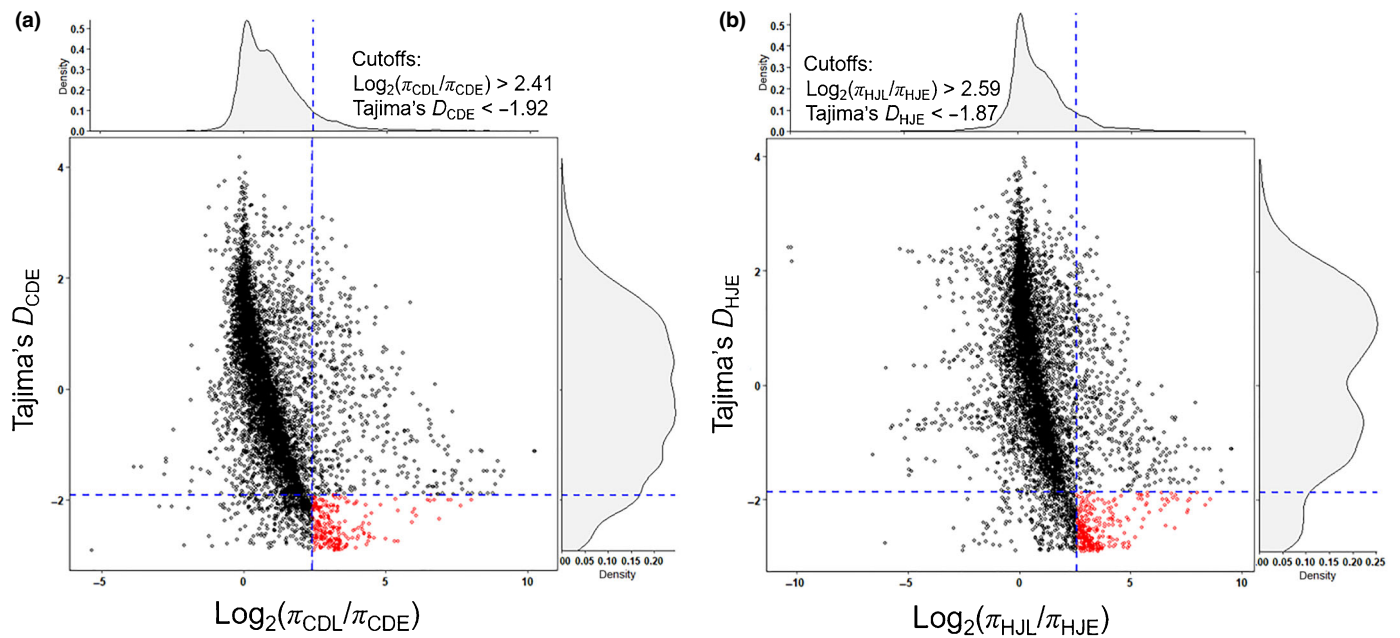


Fig. 10 Selective-sweep analyses in the early-season weedy rice populations. Candidate selective-sweep genomic regions in (a) CDE and (b) HJE populations. Blue dotted horizontal lines indicate the cut-offs of the top 10% lowest Tajima's D values, and blue dotted vertical lines indicate the cut-offs of the top 10% highest π ratio (π_{WRL}/π_{WRE}). The horizontal axis represents the \log_2 transformation of π ratio between the late- and early-season weedy rice populations, and the vertical axis represents the Tajima's D values. Each dot corresponds to one 50-kbp sliding window, and red dots are determined as candidate SSRs.

(Table S2). These findings suggested that ecological factors such as temperature and photoperiod sensitivity between the early- and late-season WR populations may also have played important roles in WR sympatric genetic divergence. Clearly, the detected uneven genetic divergence between the WR population pairs is likely to be due to selection pressure under the differential ecological conditions in the early and late rice planting seasons. This observation may reflect the adaptive genetic variation between plant populations in different environments (Lascoux *et al.*, 2016). Further studies should be conducted to demonstrate the adaptive genomic divergence in the ecological speciation processes in plants.

Examples of genetic divergence in sympatry are found in nature, although the frequency for sympatric genetic divergence is low in plant species. To date, sympatric divergence with relatively solid evidence has only been reported in a few plant species, including palms (Savolainen *et al.*, 2006), *Ricotia lunaria* (Qian *et al.*, 2018), mountain rose (Osborne *et al.*, 2020), wild emmer wheat (Wang *et al.*, 2020) and wild barley (Li *et al.*, 2020), based on analysis of their genome-wide DNA sequences. However, all these examples were found in either oceanic islands or natural canyons where plant populations/species were distributed with a certain spatial distance. Given the reasons causing such sympatric genetic divergence between these populations/species, these authors considered that differences in flowering time were critical for the diverged populations/species, because prezygotic reproductive isolation can generally be mediated by the shift in flowering time in the populations/species (Osborne *et al.*, 2020; Wang *et al.*, 2020). In addition, soil characteristics, such as acidity, water availability, and soil types and temperatures, could be the

major ecological elements that cause genetic divergence between the sympatric populations/species (Savolainen *et al.*, 2006). These findings support the viewpoint that plant species have a great potential to diverge sympatrically in nature by differential ambient ecological conditions.

Altogether, the analytical genomic results generated in this study provide solid and reliable evidence to support sympatric divergence in plant species with an extreme example. Namely, the early- and late-season WR populations represent a unique case – at the same site – to demonstrate sympatric genetic divergence. Many studies have emphasised the importance of spatial/geographical barriers for the development of reproductive isolation that is the key to create genetic divergence between plant populations/species (Schluter, 2001; Butlin *et al.*, 2008; Nosil, 2012). However, our case study using the two-season WR demonstrates the important role of temporal barriers for the development of reproductive isolation and genetic divergence in plant populations. In contrast with other studies of sympatric genetic divergence in plant species, in which ecological speciation happened at the late stages (Li *et al.*, 2020; Osborne *et al.*, 2020; Wang *et al.*, 2020), genetic divergence of the sympatric two-season WR populations used in this study is still at the very initial stages in the ecological speciation process. We believe that if such genetic divergence is continued indefinitely with the temporal isolation under different ecological conditions, these sympatric WR populations can evolve into genetically isolated entities in the long-term ecological speciation processes. Therefore, further investigations should be focused on determining reproductive isolations between the sympatric WR populations to enhance our understanding of the sympatric ecological speciation in plants.

Restricted gene flow between early- and late-season WR populations maintains sympatric genetic divergence

Our analytical results did not indicate evident migration events between the early- and late-season WR populations, even though these populations occurred in the same rice fields. This finding demonstrates extremely restricted gene flow between the early- and late-season WR populations occurring in the same rice fields, which is likely to be due to the different flowering times of WR plants in different rice planting seasons. Findings on the restricted gene flow between WR populations in different seasons in this study are similar to the reported studies in other plant species by shifted flowering time (Li *et al.*, 2020; Osborne *et al.*, 2020; Wang *et al.*, 2020). Altogether, these results indicated that the temporal isolation by different seasons can also play a strong role in influencing gene flow between plant populations.

Previous studies have also found restricted gene flow that resulted from temporal isolation through differential pollination within two grass species *Agrostis tenuis* and *Anthoxanthum odoratum* (McNeilly & Antonovics, 1968). Temporal isolation was formed through heavy metal tolerance- and intolerance-associated differential flowering times for the populations of the two grass species and occurred in the same geographic region. The authors in that study emphasised the development of restricted gene flow between plant populations through shifted flowering times even though these grass species were found in the same regions. By contrast, gene flow can occur between plant populations across a long spatial distance, as long as they are flowering in the same time period. The extreme case of measured long-distance gene flow was between creeping bentgrass (*Agrostis stolonifera*) populations, where pollen-mediated transgene flow was detected in the nontransgenic bentgrass population at the spatial distance of *c.* 21 km apart in a well designed gene flow experiment (Watrud *et al.*, 2004). Based on all the results, we believe that temporal isolation may have a much stronger impact on restricting gene flow, which can well explain the reason for the observed genetic divergence between the sympatric WR populations in the same rice fields with a spatial area of *c.* 3000 m².

Restricted gene flow between the early- and late-season WR populations has its immediate consequences. For example, our results showed distinct genetic structures between the two-season WR populations, particularly in the highly diverged CDE/L and HJE/L pairs that presented distinct genetic components, although this pattern of genetic structure was not observed in the less divergent DCE/L population pair. Therefore, the observed differences in the genetic structure of sympatric populations from different seasons supported the above conclusion that restricted gene flow between the sympatric populations can maintain their distinct genetic structures. In general, restricted gene flow between populations always results in their distinct genetic structures, whereas frequent gene flow between populations will lead to their similar genetic structures. In addition to the results from this study, and studies of other plant species, such as mountain rose (Osborne *et al.*, 2020) and wild apple (Cornille *et al.*, 2013) with limited gene flow, demonstrated that distinct genetic

structures were observed between their populations. By contrast, nearly the same pattern of genetic structures was detected in the closely distributed seagrass (*Posidonia oceanica*) populations due to frequent gene flow (Serra *et al.*, 2010). Therefore, we can conclude that restricted gene flow resulted in distinct genetic structures between the early- and late-season WR populations in this study, supporting our observation of considerable genetic divergence between the sympatric plant populations.

It is still difficult to explain how the early- and late-season WR populations occurring in the same fields maintained such strong barriers for gene flow. We hypothesised that the WR populations retained the ability to grow in their own rice-cultivation seasons due to their induced seed dormancy. In other words, the early-season WR only grows in the early rice planting season when their shattered seeds remain dormancy until next early season, and vice versa. Recently, Jiang *et al.* (2021) reported induced secondary dormancy of WR seeds after soil burial treatment. We also found seed dormancy induced by soil burial treatments in our WR seed materials (data not shown). More investigations should be carried out to fully explore the underlying mechanisms of the phenological features of early- and late-season WR plants that only prefer to grow in their 'own seasons' in the same rice fields.

Lower genetic diversity caused by stronger selective pressure in the early-season WR populations

A relative low level of genetic diversity was detected in the two early-season WR populations (CDE and HJE), as estimated by the parameters of nucleotide diversity (π), expected heterozygosity (H_E), and Watterson's estimator (θ_w), which all showed lower levels of genetic diversity in the early-season populations. This finding clearly indicates the differential genetic diversity between the sympatric WR populations, most likely to be caused both by the divergent natural selection and different temperatures in the two rice-cultivation seasons. Differences in genetic diversity have been reported previously between early- and late-season WR populations occurring in the same rice fields (Kong *et al.*, 2021), or between WR populations occurring along a latitudinal gradient (Wang *et al.*, 2019). The significant role of ambient temperature in shaping genetic diversity of organisms has been commonly recognised (Gaston, 2000). Therefore, we consider that the difference in ambient temperatures is responsible for differential genetic diversity of the sympatric WR populations occurring in the early- and late-season rice fields. Clearly, environmental conditions, such as temperature can considerably affect the genetic diversity of plant populations, creating genetic divergence. However, more studies are still needed to reveal the underlying mechanisms that determine the link between temperature and genetic diversity.

As expected, a much higher level of linkage disequilibrium and negative values of Tajima's *D*, Fu & Li's indices were detected in the early-season (CDE and HJE) compared with in the late-season (CDL and HJL) WR populations. These results indicated that the early-season WR populations were exposed to relatively stronger natural selective pressures than their corresponding late-season populations. Namely, the level of natural selective pressure

was different for the early- and late-season WR populations, supporting our observation that a relative low level of genetic diversity in early-season populations was partially caused by stronger natural selection. Natural selection has a profound effect on genetic diversity for animal and plant populations, and is also a driving force for evolution, promoting adaptation to the environment (Nielsen, 2005; Ellegren & Galtier, 2016). For example, a low level of genetic diversity was reported in the agricultural weed *Ipomoea purpurea* adapting to strong selection under the application of glyphosate herbicides (Kuester *et al.*, 2016). Natural selection created by stressful ecological factors (e.g. water, temperature and geography) played a major role in generating a moderate level of genetic diversity and creating regional genetic divergence in wild emmer wheat (Ren *et al.*, 2013). A similar situation of a relative low levels of genome-wide genetic diversity was also reported between the Lake Michigan steelhead and their ancestral populations (Willoughby *et al.*, 2018). All these studies provided solid evidence for natural selection that affected the genetic diversity of populations/species.

Consistent with the results from the above studies, our findings in this study also indicated that selection resulted in differential genetic diversity at the genome-wide level between the sympatric early- and late-season WR populations. Therefore, we believe that divergent selection can cause differential genetic diversity between sympatric early- and late-season WR populations through adaptation to different ecological conditions. In addition, we consider that the limited gene flow between the sympatric early- and late-season WR populations played an important role in maintaining this accumulated genetic divergence between the sympatric populations in the same rice fields. Therefore, we conclude that the change of genetic diversity caused by the divergent selection under different ecological conditions and restricted gene flow (temporal isolation) together resulted in genetic divergence between the sympatric WR populations occurring in the same rice field, but in different seasons. This type of model may also be found in other plant populations/species with seasonal or temporal isolations.

Conclusions

Very close phylogenetic relationships were detected between the early- and late-season rice varieties in the same fields in Leizhou, compared with rice varieties from other regions. Substantial genetic divergence across genomes was detected between the early- and late-season WR populations occurring in the same sampling rice fields. This sympatric genetic divergence was likely to be associated with restricted gene flow and divergent selection between the two-season WR populations and resulted in relatively distinct genetic structures between the sympatric WR populations. This finding supports the pattern of sympatric divergence in the ecological speciation concept in plants. Furthermore, a relatively low level of genetic diversity was detected in the early-season WR populations, which was supposedly associated with divergent selection and selective sweeps. Together, divergent selection under different ecological conditions and restricted gene flow resulted in genetic divergence between the sympatric WR

populations. Our findings demonstrate that not only spatial/geographic isolations can serve as reproductive barriers, but that temporal isolation can also provide strong barriers to create genetic divergence between sympatric populations/species in plants.

Acknowledgements

This research is supported by the Natural Science Foundation of China (31770409 and 32070245), and the Hainan Provincial Natural Science Foundation of China (319QN304).

Author contributions

B-RL and L-FL conceived and designed this study. ZW performed this study, analysed the data and wrote the manuscript. ZW, XC, X-QJ and Q-YX collected the plant materials. All authors contributed to the manuscript revision.

ORCID

Bao-Rong Lu  <https://orcid.org/0000-0002-0214-2390>

Data availability

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive at the National Genomics Data Center (<https://bigd.big.ac.cn/>), under the bioproject accession number PRJCA004881.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.
- Barluenga M, Stoltzing KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan Crater Lake cichlid fish. *Nature* 439: 719–723.
- Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ. 2012. Sympatric speciation in the post “modern synthesis” era of evolutionary biology. *Evolutionary Biology* 39: 158–180.
- Butlin RK, Galindo J, Grahame JW. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 363: 2997–3007.
- Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SNP-SIFT. *Frontiers in Genetics* 3: 35.
- Cornille A, Gladieux P, Giraud T. 2013. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Application* 6: 737–748.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
- Delouche JC, Burgos NR, Gealy DR, Martín GZ, Labrada R. 2007. *Weedy rice origin, biology, ecology and control*. Rome, Italy: Food and Agriculture Organization of the United Nations.
- Ellegren H, Galtier N. 2016. Determinants of genetic diversity. *Nature Reviews Genetics* 17: 422–433.
- Fitzpatrick BM, Fordyce JA, Gavrillets S. 2008. What, if anything, is sympatric speciation? *Journal of Evolutionary Biology* 21: 1452–1459.
- Footo AD. 2018. Sympatric speciation in the genomic era. *Trends in Ecology & Evolution* 33: 85–95.

- Fruciano C, Franchini P, Kovacova V, Elmer KR, Henning F, Meyer A. 2016. Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nature Communications* 7: 12736.
- Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
- Gaston KJ. 2000. Global patterns in biodiversity. *Nature* 405: 220–227.
- Jiang XQ, Zhu XY, Lu BR. 2021. Induced dormancy in weedy rice seeds by changes of hormone contents under soil burial: implications in adaptive evolution and weed control. *Journal of Systematics and Evolution*. doi: 10.1111/jse.12737.
- Jiang Z, Xia H, Basso B, Lu BR. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theoretical and Applied Genetics* 124: 309–322.
- Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Härer A, Hulsey CD, Franchini P *et al.* 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* 588: 106–111.
- Kong H, Wang Z, Guo JY, Xia QY, Zhao H, Zhang YL, Guo AP, Lu BR. 2021. Increases in genetic diversity of weedy rice associated with ambient temperatures and limited gene flow. *Biology* 10: 71.
- Kuester A, Wilson A, Chang SM, Baucom RS. 2016. A resurrection experiment finds evidence of both reduced genetic diversity and potential adaptive evolution in the agricultural weed *Ipomoea purpurea*. *Molecular Ecology* 25: 4508–4520.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA-X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.
- Lascoux M, Glémin S, Savolainen O. 2016. Local adaptation in plants. In: *eLS*. Chichester, UK: John Wiley & Sons. doi: 10.1002/9780470015902.a0025270.
- Lassalle F, Muller D, Nesme X. 2015. Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. *Research in Microbiology* 166: 729–741.
- Li GJ, He ZX, Wang G, Ke BN, Luo CL, Lu BR. 2013. Occurrence, damage and control strategies of weedy rice in Leizhou, Guangdong Province. *Weed Science* 31: 20–25 [in Chinese].
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMTOOLS. *Bioinformatics* 25: 2078–2079.
- Li K, Hong W, Jiao H, Wang GD, Rodriguez KA, Buffenstein R, Zhao Y, Nevo E, Zhao H. 2015. Sympatric speciation revealed by genome-wide divergence in the blind mole rat *Spalax*. *Proceedings of the National Academy of Sciences, USA* 112: 11905–11910.
- Li K, Zhang S, Song X, Weyrich A, Wang Y, Liu X, Wan N, Liu J, Lovy M, Cui H *et al.* 2020. Genome evolution of blind subterranean mole rats: adaptive peripatric versus sympatric speciation. *Proceedings of the National Academy of Sciences, USA* 117: 32499–32508.
- Malinsky M, Challis RJ, Tyers AM, Schiffels S, Terai Y, Ngatunga BP, Miska EA, Durbin R, Genner MJ, Turner GF. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* 350: 1493–1498.
- Malinsky M, Matschiner M, Svardal H. 2021. DSUITE – Fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resource* 21: 584–595.
- Marques DA, Lucek K, Meier JI, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2016. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genetics* 12: e1005887.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M. 2010. The Genome Analysis Toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.
- McNeilly T, Antonovics J. 1968. Evolution in closely adjacent plant populations IV. Barriers to gene flow. *Heredity* 23: 205–218.
- Meyer JR, Dobias DT, Medina SJ, Servilio L, Gupta A, Lenski RE. 2016. Ecological speciation of bacteriophage lambda in allopatry and sympatry. *Science* 354: 1301–1304.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Nei M, Li WH. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences, USA* 76: 5269–5273.
- Nielsen R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* 39: 197–218.
- Nosil P. 2012. *Ecological speciation*. New York, NY, USA: Oxford University Press.
- Nosil P, Funk DJ, Ortiz-Barrientos D. 2009a. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 18: 375–402.
- Nosil P, Harmon LJ, Seehausen O. 2009b. Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* 24: 145–156.
- Osborne OG, Kaffe T, Brewer T, Dobrevá MP, Hutton I, Savolainen V. 2020. Sympatric speciation in mountain roses (*Metrosideros*) on an oceanic island. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 375: 20190542.
- Peccoud J, Simon JC. 2010. The pea aphid complex as a model of ecological speciation. *Ecological Entomology* 35: 119–130.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* 8: e1002967.
- Qian C, Yan X, Yin H, Fan H, Yin XY, Sun PP, Li ZJ, Nevo E, Ma XF. 2018. Transcriptomes divergence of *Ricotia lunaria* between the two micro-climatic divergent slopes at “Evolution Canyon” I, Israel. *Frontiers in Genetics* 9: 506.
- Quinlan AR, Hall IM. 2010. BEDTOOLS: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Ren J, Chen L, Sun DK, You FM, Wang JR, Peng YL, Nevo E, Beiles A, Sun DF, Luo MC. 2013. SNP-revealed genetic diversity in wild emmer wheat correlates with ecological factors. *BMC Evolutionary Biology* 13: 169.
- Richards EJ, Servedio MR, Martin CH. 2019. Searching for sympatric speciation in the genomic era. *BioEssays* 41: e1900047.
- Savolainen V, Anstett MC, Lexer C, Hutton I, Clarkson JJ, Norup MV, Powell MP, Springate D, Salamin N, Baker WJ. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441: 210–213.
- Schluter D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16: 372–380.
- Schluter D. 2009. Evidence for ecological speciation and its alternative. *Science* 323: 737–741.
- Serra IA, Innocenti AM, Di Maida G, Calvo S, Migliaccio M, Zambianchi E, Pizzigalli C, Arnaud-Haond S, Duarte CM, Serrão EA. 2010. Genetic structure in the Mediterranean seagrass *Posidonia oceanica*: disentangling past vicariance events from contemporary patterns of gene flow. *Molecular Ecology* 19: 557–568.
- Servedio MR, Noor MAF. 2003. The role of reinforcement in speciation: theory and data. *Annual Review of Ecology, Evolution, and Systematics* 34: 339–364.
- Song X, Li GL, Vernoooy R, Song Y. 2021. Community seed banks in China: achievements, challenges and prospects. *Frontiers in Sustainable Food Systems* 5: 630400.
- Song ZJ, Wang Z, Feng Y, Yao N, Yang J, Lu BR. 2015. Genetic divergence of weedy rice populations associated with their geographic location and coexisting conspecific crop: implications on adaptive evolution of agricultural weeds. *Journal of Systematics and Evolution* 53: 330–338.
- Stuessy TF. 2006. Evolutionary biology: sympatric plant speciation in islands? *Nature* 443: E12.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
- Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution* 16: 381–390.
- Vilella AJ, Blanco-García A, Hutter S, Rozas J. 2005. VARI-SCAN: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics* 21: 2791–2793.
- Wang H, Yin H, Jiao C, Fang X, Wang G, Li G, Ni F, Li P, Su P, Ge W. 2020. Sympatric speciation of wild emmer wheat driven by ecology and chromosomal

- rearrangements. *Proceedings of the National Academy of Sciences, USA* 117: 5955–5963.
- Wang Z, Fu MJ, Zhu HG, Zhu Y, Zhao XX, Lu BR. 2019. Enhanced genetic diversity of weedy rice populations associated with latitude decreases revealed by simple sequence repeat fingerprints. *Journal of Systematics and Evolution* 57: 66–74.
- Watrud LS, Lee EH, Fairbrother A, Burdick C, Reichman JR, Bollman M, Storm M, King G, Water PK. 2004. Evidence for landscape-level, pollen-mediated gene flow from genetically modified creeping bentgrass with *CP4 EPSPS* as a marker. *Proceedings of the National Academy of Sciences, USA* 101: 14533–14538.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256–276.
- Weir BS, Cockerham CC. 1984. Estimating f-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
- Willoughby JR, Harder AM, Tennessen JA, Scribner KT, Christie MR. 2018. Rapid genetic adaptation to a novel environment despite a genome-wide reduction in genetic diversity. *Molecular Ecology* 27: 4041–4051.
- Ye CY, Tang W, Wu D, Jia L, Qiu J, Chen M, Mao L, Lin F, Xu H, Yu X. 2019. Genomic evidence of human selection on Vavilovian mimicry. *Nature Ecology & Evolution* 3: 1474–1482.
- Zhang C, Dong SS, Xu JY, He WM, Yang TL. 2019. POPLD_{DECAY}: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35: 1786–1788.
- Zhang L, Dai W, Wu C, Song X, Qiang S. 2012. Genetic diversity and origin of *Japonica*- and *Indica*-like rice biotypes of weedy rice in the Guangdong and Liaoning provinces of China. *Genetic Resources and Crop Evolution* 59: 399–410.
- Zhu YQ, Fang J, Wang Y, Pang LH, Lu BR. 2021. Key roles of de-domestication and novel mutation in origin and diversification of global weedy rice. *Biology* 10: 828.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Rice fields managed by small farmer households in Leizhou.

Fig. S2 Morphology of weedy rice.

Fig. S3 Analysis of highly diverged genomic regions shared among the three weedy rice population pairs.

Fig. S4 Analysis of selective-sweep genomic regions in the CDE and HJE populations.

Table S1 Information of genomic data used in this study.

Table S2 Functional genes identified in the highly divergent genomic regions of the CDE/L, HJE/L and DCE/L population pairs.

Table S3 Estimation of gene introgression using ABBA-BABA test.

Table S4 Information of 21 previously cloned genes in the selective-sweep genomic regions shared by the CDE and HJE populations.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.