Population genetic analysis and scans for adaptation and contemporary selection footprints provide genomic insight into *aus*, *indica* and *japonica* rice cultivars diversification

Seyed Milad Vahedi
Dalhouse University Department Animal Science and Aquaculture https://orcid.org/0000-0002-5846-2107

Moslem Momen
University of Hormozgan

Seyedeh Fatemeh Mousavi
University of Kurdistan https://orcid.org/0000-0001-7198-4252

Mohammad Hossein Banabazi
ASRI: Animal Science Research Institute

Mohammad Saeed Hasanvandi
Agricultural Research, Education & Extension Organization (AREEO)

Madhav Bhatta
University of Wisconsin-Madison https://orcid.org/0000-0003-3986-8301

Mahmoud Amiri Roudbar (mahmood.amiri225@gmail.com)
Safiabad-Dezful Agricultural and Natural Recourses Research and Education Center
https://orcid.org/0000-0002-2023-9292

Siavash Salek Ardestani
Department of Animal Science, University of Zanjan, Zanjan, Iran

Research Article

Keywords: rice, ecotype, population structure, signatures of selection, adaptation, diversification

Posted Date: July 21st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-121958/v2

Abstract

Following domestication, rice cultivars have been spread worldwide to different climates and have experienced selection pressures to improve desirable traits. This has resulted in diverse cultivars that display variation in phenotypic traits, such as stress tolerance, grain size, and yield. To better understand the genomic composition arising from cultivars development and local adaptation, high-density genotypes [containing 286,183 single nucleotide polymorphisms (SNPs) after the quality control] of 1,284 rice cultivars of *aus*, *indica*, and *temperate* and *tropical japonica* were scanned for diversifying signatures applying a pairwise comparison of fixation index (F_{st}) test. Each cultivar's population was investigated for contemporary selection using the integrated haplotype score (iHS) test. Signatures of diversifying selection among the pairwise comparisons were found in genomic regions mainly involved in response to stress (pathogens, drought, heat, cold) and development and morphology of various structures, such as root, pollen, spikelet, and grain. The most significant diversification signal between *indica* and *japonica* cultivars was detected at the location of *ROX2* gene. A*us* with *indica* comparison detected the most divergent signal at important candidate genes of *OsEXPA8* and *OsEXPA9*, whereas *temperate* with *tropical japonica* comparison resulted in two well-known candidate genes *OsHCT4* and *OsGpx4.* Recent selection analysis detected different patterns of contemporary selection in genomic regions related to rice breeding standard criteria such as stress tolerance, seed germination, starch content, and flowering time. Our findings highlight the underlying molecular basis of adaptive divergence and propose that modern rice breeding may provide additional diversification among rice cultivars.

Introduction

Crop domestications took long-term selection processes and could significantly advance human civilization. Cultivated rice *(Oryza sativa L.),* which is cultivated worldwide and is one of the most important food crop, is considered to have been domesticated from wild rice *(Oryza rufipogon)* thousands of years ago in China (Zheng *et al.* 2016). Since then, natural and artificial selection have resulted in genetically, morphologically, and physiologically various commercial rice cultivars. Despite rice being a significant cereal and a model system for plant biology, there is still widespread controversy about cultivated rice's evolution and differentiation (Huang *et al.* 2012).

Two major eco-geographical subspecies of cultivated rice are *indica* and *japonica*. *Indica*, called lowland rice, grows throughout tropical Asia, whereas *japonica* is typically cultivated in highlands of southern China, Southeast Asia, and Indonesia, as well as outside of Asia, in Africa and North and South America (Londo *et al.* 2006; Xiong *et al.* 2011). Conventionally, *japonica* rice is classified into two ecotypes of *temperate japonica* and *tropical japonica* according to the ecosystems they belong to (Yoshida and Benta 1983). *Temperate japonica* is cultivated in temperate regions due to its cold tolerance, while *tropical japonica* is mainly cultivated in tropical zones (Lee *et al.* 2018). In tropical conditions, *indica* rice has higher yields than *japonica* rice, whereas, under cold conditions, *japonica* rice

performs better (Jing *et al.* 2010). Photoperiod sensitivity and spikelet sterility induced by high temperatures were proposed as the main reasons for the poor growth of *temperate japonica* rice in tropical regions (Lee *et al.* 2018). *Aus* is a close ecotype to *indica*, and this type has evolved and cultivated in the tropical condition of India and Bangladesh (Londo *et al.* 2006).

The diversification of *indica* and *japonica* cultivated rice has been an interesting topic in evolutionary biology. The adaptation process and natural and artificial selection have left detectable footprints within the rice genome. Advances in genotyping technology have provided an opportunity to reveal the genomic selection signatures in rice (Li *et al.* 2017). Some of these footprints of selection (also known as selection signatures) reflect the 'historical or old selection' during domestication. In contrast, some represent selection within the past few generations for economically important traits, such as higher yield or environmental adaptation, referred to as 'contemporary or recent selection' (Decker *et al.* 2014). Multiple analytical methods have been introduced to identify recent or old positive selection signals on a genome-wide scale using genotypic data. Single-site population differentiation methods such as the fixation index (F_{st}; Weir and Cockerham, 1984) are proper for identifying the signature of selections associated with adaptation processes. In contrast, the linkage disequilibrium (LD)-based method of integrated haplotype score (iHS) developed by Voight *et al.* (2006) is helpful for the detection of recent selection signatures. Using selection signatures methods and comparing cultivated and wild rice genome, genes involved in rice domestication, e.g., *Sh4*, *qSW5*, *qSH1*, *prog1*, *sd1*, *Wx*, *Badh2*, and *Rc*, were identified (Huang *et al.* 2012; Zhao *et al.* 2018).

Although the genetic basis underlying economically important traits in rice have been well studied, our knowledge of the genetic mechanism responsible for the adaptation to local environments and rice cultivars diversification process is limited. Exploring the genetic information of cultivated rice can clear its adaptation process and provide essential insights into breeding elite varieties for sustainable agriculture. Studying locally adapted rice ecotypes is promising to underpin the genes involved in differentiation, as well as ecologically and economically important traits. This study aimed to assess population structure and detect selection signatures in the genomes of the four cultivated rice ecotypes of *aus*, *indica*, and *temperate* and *tropical japonica* adapted to different environments using high-density single nucleotide polymorphism (SNP) genotyping data. Our results provided more understanding of the evolution and biology underlying rice cultivars diversification and may provide tools to increase the efficiency of selection programs in rice agronomy.

Materials and Methods

Data collection and quality control

Genotypic data of 1,284 of four cultivated rice ecotypes, including *aus* (14.56%), *indica* (38.71%), and *temperate* (18.69%) and *tropical japonica* (28.04%), was obtained from Rice Diversity Project database (http://www.ricediversity.org). Three germplasm collections of RDP1 (343 accessions), RDP2 (880 accessions), and NIAS (61 accessions) were included in this study. Samples were already collected from 94 countries worldwide, and genotypes were generated by the high-density rice array (HDRA) comprising 700 K SNPs. This SNP array This SNP array was developed using the genome of different rice species and cultivars, including wild (*O. rufipogon*/*O. nivara*, *O. meridionalis*, *O. officinalis*, *O. punctata*) and *O. sativa* (*temperate japonica*, *tropical japonica*, aromatic/Group V, *indica*, *aus*, admixed) (McCouch et al. 2016). Table S1 presents the number and origin of samples included in this study.

Quality control was performed using the PLINK 2.0 software (Purcell *et al.* 2007). SNPs with a call rate < 80%, minor allelic frequency < 0.01, and Hardy-Weinberg equilibrium p-value < 10^{-7} were removed. Moreover, samples with less than 80% genotype call rate were discarded from downstream analyses. Finally, genotypes were phased, and missing genotypes were imputed using Beagle software 5.2 (Browning *et al.* 2018).

Population structure analyses

Principal component analysis (PCA) was conducted using PLINK 2.0 software (Purcell *et al.* 2007). Then, the plot of principal components was constructed using ggplot2 package (Villanueva and Chen 2019) in the R 4.0.5 software (R Core Team 2020). The phylogenetic analysis was performed using the neighbor-joining approach in VCF-kit 0.2.9 (Cook and Andersen 2017). Visualization of the phylogenetic analysis was based on midpoint rooting in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree). Maximum likelihood analysis of population structure was conducted using Admixture 1.3 (Alexander *et al.* 2009) for K values ranging from 2 to 4 with 10 iterations per each K. The Admixture software uses a cross-validation procedure to estimate the preferable ancestral populations (K). Linkage disequilibrium (LD) was estimated based on the allele frequency correlations ( ) (Hill and Robertson 1968) using the PopLDdecay 3.4 software (Zhang *et al.* 2019), with a maximum distance between two SNPs of 300 Kb.

Fixation index (F_{st})

Population genetic analysis results showed high genetic differences between *indica* and *japonica* (*temperate* and *tropical*) groups. Moreover, our analyses indicated that *aus* cultivar was derived from *indica*, and the two subpopulations of *tropical* and *temperate japonica* also derived from *japonica* over time. Based on these findings, we designed three main comparisons, including comparing the genome of *indica* with two *japonica* populations of *temperate* (Ind_Tem) and *tropical* (Ind_Tro), *aus* with *indica*

(Aus_Ind), and *tropical japonica* with *temperate japonica* (Tem_Tro), searching for regions involved in adaptation and diversification processes using Weir & Cockerham's F_{st} (Weir & Cockerham 1984). Table 1 presents the study design and the number of samples in each pairwise test. F_{st} values were estimated and averaged along 100 kb genomic windows with a step size of 25 Kb using VCFtools 0.1.15 software (Danecek *et al.* 2011). F_{st} values were then Z-transformed ($Z(F_{st})$) using *scale* function in R software (R Core Team 2020), and genomic windows ranked in the highest 99^{th} percentile of $Z(F_{st})$ values considered as regions under significant selection. Finally, these genomic regions were applied to gene annotation analysis to find candidate genes involved in the adaptation process of rice ecotypes.

<< Table 1 around here >>

iHS test

We employed the iHS test to evaluate recent positive selection evidence based on haplotype frequencies (Voight *et al.* 2006). The iHS statistic measures the extent of local LD, then partitions them into two classes: haplotypes centered upon an SNP that carry the ancestral versus the derived allele. This statistic is applied to samples SNPs and begins by calculating the integrated extended haplotype homozygosity (EHH) (Sabeti *et al.* 2002), defined as the integral of the observed decay of EHH away from a particular core allele until EHH reaches the value of 0.05. The integrated EHH (iHH), summed over both directions away from the core SNP, is referred to iHH_A or iHH_D, based on whether it is computed for the ancestral or derived core allele. The unstandardized iHS value is then calculated as follows:

$$Unstandardized\ iHS = In\left(\frac{\mathrm{iHH}_A}{\mathrm{iHH}_D}\right)$$

This quantity is standardized with a mean of 0 and variance of 1 irrespective of allele frequency at the core SNP (for more details, see Voight *et al.* (2006)) as follows:

$$iHS = \frac{In\left(\frac{\mathrm{iHH}_A}{\mathrm{iHH}_D}\right) - E\left[In\left(\frac{\mathrm{iHH}_A}{\mathrm{iHH}_D}\right)\right]}{SD\left[In\left(\frac{\mathrm{iHH}_A}{\mathrm{iHH}_D}\right)\right]}$$

Large positive or negative values of iHS refer to unusually extended haplotypes carrying the ancestral or derived allele, respectively. Then, iHS values were fitted to a normal distribution using the robust linear model (*rlm* function) of the MASS R package (Venables and Ripley 2013) and model = rlm (iHS ~ 1), where the iHS object is a vector containing the iHS values. The outputs of the fitted model, including mean and standard deviation, were used by the *pnorm* R function to calculate the two-sided p-values of the iHS statistics (lower.tail = TRUE, log.p = FALSE). Finally, to control multiple testing false discovery rate (FDR) among rejected null hypotheses, the iHS p-values were transformed to the corresponding q-values using the *qvalue* R function and the Benjamini and Hochberg method (Benjamini and Hochberg 1995). The significant threshold of q-value < 0.01 was considered to identify SNP under intensive recent selection pressure. This analysis was separately conducted within each of the four ecotypes populations.

Gene annotation and functional enrichment analysis

The most recent rice genome assembly of *Oryza sativa* Japonica Group genes (IRGSP-1.0; https://plants.ensembl.org/) and the biomaRt 2.46.3 R package (Durinck *et al.* 2009) were used to identify candidate genes. F_{st} values were Z-transformed ($Z(F_{st})$), and genomic regions ranked in the highest 99^{th} percentile of $Z(F_{st})$ were identified. Genes within these genomic regions were considered candidate genes involved in the rice ecotypes diversification process. Since in all ecotypes, a sufficient level of LD (LD > 0.2) was detected at the distance of 10 Kb, we considered a distance of ±5 Kb on both the upstream and downstream sides from significant markers in the iHS test to find the candidate genes under recent selection pressure. Manhattan plots of the results were created using the R package CMplot 3.6.2 (Yin *et al.* 2021).

Gene ontology was conducted based on the reference list of *Oryza sativa* Japonica Group. Terms of Gene Ontology biological process (GO:BP), molecular function (GO:MF), and cellular component (GO:CC), and biological pathway terms of Kyoto Encyclopedia of Genes and Genomes (KEGG) were assessed for all genes using functional profiling (g:GOSt) of gProfiler online software (Raudvere *et al.* 2019). The FDR adjusted p-value < 0.05 was considered the threshold for identifying the overrepresented terms in all functional enrichment analyses.

Results

A total number of 1,284 rice genotypes from 94 countries, including four cultivated rice ecotypes of *aus*, *indica*, and *temperate* and *tropical japonica*, were included in this study. After quality control, 286,183 SNPs from 1,284 samples remained for further analyses.

Population genetic structure

Our phylogenetic analysis illustrated *aus*, *indica*, and *japonica* ecotypes in three separate main branches (Fig. 1a). As expected, *temperate* and *tropical japonica* were located in one clade, and *aus* showed a closer phylogenetic relationship with *indica* than *japonica* populations. Similar population affinities were obtained based on the PCA results, in which a clear genetic structure with samples from each ecotype clustering together was observed, while *temperate* and *tropical japonica* were placed close together (Fig. 1b). A total of 37.83% of genetic variation was explained by the top three principal components (PC1-PC3). The first component (PC1 = 24.53%) was driven by the difference between two large clusters of *indica* and *japonica* ancestry. However, *aus* cluster separated from the *indica* population in the second component (PC2 = 8.13%). Finally, the third component broke the *japonica* ecotype into two *temperate* and *tropical* subpopulations (PC3 = 5.17%).

The ancestral lineage compositions of four ecotypes are shown in Figure 1c—the K value, representing the number of ancestral populations. Considering the history of the different populations, we chose to plot the admixture results from two assumed ancestries (K = 2) to K of 4 (Fig. 1c). In K = 2, the samples were split into two groups: (1) pure *indica* ancestry; and (2) pure *japonica* ancestry. However, limited admixture between two ancestries was observed. In K = 3, *temperate* and *tropical japonica* were located in one population, while *aus* and *indica* separated. In this K, the proportion of *indica* alleles in *aus* was estimated at 0.001±0.037, while the ancestral *japonica* allele contributed to 0.000±0.002 of *aus* genome. In K = 4, two *japonica* subpopulations of *temperate* and *tropical* separated. Admixture analysis showed that only 0.002±0.008 of alleles are ancestrally identical between *temperate* and *tropical japonica*. In this K, 0.000±0.001 and 0.001±0.001 of introgressed alleles from *japonica* into *indica* populations were originated from *temperate* and *tropical* cultivars, respectively. Clusters identified by increasing K beyond four did not contain a single individual with a majority of ancestry in the new cluster, indicating that for the number of markers evaluated here, K = 5 does not improve population admixture resolution.

<< Figure 1 about here>>

Fixation index

A total of 14,855 100 Kb-genomic windows with step sizes of 25 Kb was scanned along the rice genome to calculate F_{st} values. Figure 2 presents the $Z(F_{st})$ distribution in four pairwise comparisons of Ind_Tem, Ind_Tro, Aus_Ind, Tem_Tro, showing the potential regions involved in adaptation and diversification.

<< Figure 2 around here >>

Ind_Jap pairwise comparison

Genomic windows ranked in highest 99th percentile of $Z(F_{st})$ values (N = 148) were identified ($Z(F_{st})_{Ind\_Tro} \geq 2.19$; $Z(F_{st})_{Ind\_Tem} \geq 2.11$). The most significant genomic region in both Ind_Tem and Ind_Tro comparisons was located on chromosome (chr) 2 (chr2:12.70-12.80 Mb) with $Z(F_{st})$ values of 3.79 and 3.57, respectively. A total of 92 genomic regions were overlapped between Ind_Tro and Ind_Tem top 1% genomic regions, out of which the largest number of regions were located on chr2 (N = 18, chr2:12.3-15.52 Mb) followed by chr10 (N = 15, chr10:14.1-23.2 Mb).

Gene annotation analysis of identified regions in the pairwise comparison of Ind_Tem and Ind_Tro detected 248 and 245 candidate genes, respectively (Table S2). We identified 168 candidate genes overlapping the two Ind_Jap analyses (Fig. S1). These genes were associated with a diverse range of traits, such as response to pathogens and disease resistance (*OsWRKY52*, *OsWRKY46*, *OsWRKY40*, *OSRAC7*, *ROX2*, *ONAC109*), salt stress tolerance (*OsWRKY50*, *OsCAX1c*, *OsTPS6*, *OsSOD4*, *RD22*, *ONAC110*), drought stress response (*OsCYP18-2*, *OsWRKY104*, *OsDSR6*, *OsDi19-3*, *OsITPK5*, *ONAC110*), cold stress response (*OsCHB701*, *TSV3*), seed germination and growth (*OsWRKY50*, *OsIPMS2*, *OsWRKY50*), flavone accumulation and UV-tolerance (*OsUGT706D2*, *OsUGT706C4*, *OsUGT706C3*), tapetum degeneration (*API5*), root development (*YUCCA2*), lignin and cellulose synthesis (*OsCesA8*), leaf angle (*OsbHLH098*), chloroplast development (*YSS1*), arsenite root uptake (*OsNIP3*), and spikelet (*OPR7*) and pollen development (*UGP1*).

Gene enrichment analysis of Ind_Jap candidate genes found 57 significantly overrepresented terms. The most significant terms in which Ind_Jap candidate genes were overrepresented were regulation of biological process (GO term = GO:0050789; FDR-adjusted p-value = 0.001) and regulation of cellular process (GO term = GO:0050794; FDR-adjusted p-value = 0.001). Table S3 presents the results of gene enrichment analysis in Ind_Jap comparison.

Aus_Ind pairwise comparison

Genomic regions with top 1% $Z(F_{st})$ values (N = 148) were identified ($Z(F_{st})_{Aus\_Ind} \geq 2.98$). The most significant signal was detected on chr1 (chr1:8.15-8.25 Mb) with $Z(F_{st})_{Aus\_Ind}$ value of 5.51, and the largest number of windows were located on chr4 (N = 33, chr4:16.75-16.95 Mb & chr4:30.52-35.45 Mb) followed by chr1 (N = 31, chr1:3.05-3.37 Mb & chr1:7.75-8.77 Mb).

Gene annotation analysis of identified regions resulted in 199 candidate genes (Table S4). These genes were associated with a diverse range of traits, such as disease tolerance and response to pathogens (*OsRBX1a*, *OsTVLP1*, *OsABCF3*, *LML1*), salt stress response (*SRWD5*, *OsADF2*, *OsRMC*, *OsIPK1*), drought stress response (*OsLEA20*, *DREB2A*, *OsSKgamma*, *OsADF2*, *OsIPK1*), heat stress response (OsBiP2, OsFKBP62b), arsenic stress (*OsABCB11*), anther development (*OsbHLH035*), root growth and architecture (*EXPA8*, *OsEXPA9*, *OsRAA1*, *RR6*), leaf senescence and morphology (*OsPME1*, *ADL1*), iron homeostasis (*OsOPT8*), seed germination (*OsDOG1L-2*, *OsGLN2*), pollen tube germination and growth (*OsUCL8*), pollen wall development (*OsUAM2*), grain filling (*GF14F*), and flowering time (*OsCOL10*).

Gene enrichment analysis of Aus_Ind candidate genes found four significantly overrepresented terms (FDR adjusted p-value < 0.05) associated with cinnamyl-alcohol dehydrogenase and sinapyl alcohol dehydrogenase activity, as well as two pathways of aromatic compound and organic cyclic compound biosynthetic processes (Table S3).

Tem_Tro pairwise comparison

Genomic regions possessing top 1% $Z(F_{st})$ values (N = 148) were extracted ($Z(F_{st})_{Tem\_Tro} \geq 3.21$), mostly located within three segments on chr5 (N = 59, chr5:5.55-6.25 Mb & chr5:16.3-20.07 Mb & chr5:27.12-28.50 Mb). The most significant signal was detected on chr9 (chr9:4.87-4.97 Mb) with $Z(F_{st})_{Tem\_Tro}$ value of 4.75.

A total of 165 candidate genes were detected in gene annotation analysis (Table S5). The identified genes were associated with different traits, e.g., response to pathogens (*OsSLRL1*, *APIP10*, *JIOsPR10*, *OsVQ10*, *OsVQ11*, *OsBIHD1*), salinity stress response (*OsRFP*), drought tolerance (*CK1*, *OsWRKY55*, *OsCTR1*), heat tolerance (*OsCam1-1*), cold tolerance (*OsSIP1*), seed and root development (*OsSultr5*), chloroplast development (*YL1*), anther development (*RR24*, *OsHFP*), pollen development (*OsPME10*), grain size and weight (*LRK1*, *LRK7*, *LRK8*), phosphate hemostasis (*OsRLCK64*), and rice dwarfism (*SSD1*).

Gene enrichment analysis of Aus_Ind candidate genes found two significantly overrepresented terms (FDR-adjusted p-value < 0.05). Two cellular component terms of integral and intrinsic components of peroxisomal membrane were significantly overrepresented by Aus_Ind candidate genes (Table S3).

Recent selection signals

Following obtaining iHS scores, p-values were calculated and corrected for multiple testing. Figure 3 presents the distribution of -$\log_{10}$(p-value) of iHS scores along the rice genome of different ecotypes showing the potential recent selection signals. A total of 458, 443, 222, and 332 markers were found to be under recent selective pressure for *aus*, *indica*, *temperate japonica,* and *tropical japonica*, respectively (q-value < 0.01) (Table S6). The most significant signal in *aus* and *indica* was located on chr1 (chr1:13,549,676:13,559,676) and chr1 (chr1:18,532,173:18,542,173 bp), whereas in *temperate japonica* and *tropical japonica* were on chr12 (chr12:10,794,098:10,804,098 bp) and chr9 (chr9:1,722,050:1,732,050 bp).

Gene annotation analysis of detected genomic regions found 110, 99, 43, and 67 candidate genes under recent selection pressure in *aus*, *indicia*, *temperate japonica*, and *tropical japonica* populations, respectively (Table S6). Figure 4 presents the list and number of potentially overlapping genes under recent selection among different ecotypes. The identified genes were associated with different favorable rice breeding traits; such as drought tolerance (*OsRLCK17*, *OsGATA8*, *OsNPC2*, *OsRLK5*, *SAPK6*, *OsCCR10*), disease tolerance (*OsRLCK18*, *RBBI2-3*, *APIP12*, *OsPR2*, *Gns6*, *OsPAL1*), heat tolerance (*tms5*, *SAPK6*), cold tolerance (*OsMYB30*, *CTB4a*), plant height (*OsGA2ox7*, *DRUS2*), salinity tolerance (*OsGATA8*, *OsNPC2*, *OsNUC1-S*, *ZRP4*), arsenic tolerance (*R2R3-MYB*), seed germination (*OsBgal3*), starch components (*DRUS2*), lignin content (*GH2*, *OsC4H*), grain yield (*OsPLL3*, *RPBF*, *OsATG8b*), flowering time (*OsPRR1*, *OSCRY2*), pigment formation (*COW1*), pest resistance (*Osr9-LOX1*), flavonoid accumulation and UV-B tolerance (*OsRLCK160*).

<< Figure 3 around here >>

<< Figure 4 around here >>

Identified candidate genes in each ecotype were separately applied to gene enrichment analysis. In *aus* ecotype, the discovered candidate genes were significantly overrepresented in molecular function of some enzymes, such as terpene synthase (term = GO:0010333; FDR adjusted p-value = 0.00) and lyase (term = GO:0016829; FDR adjusted p-value = 0.02). Concerning *indicus* ecotype, we found that candidate genes were significantly overrepresented in carbohydrate-binding molecular function (term = GO:0030246; FDR adjusted p-value = 0.04). Gene enrichment analysis of *temperate japonica* candidate genes showed significant overrepresentation in 38 terms, among which the most significant molecular function and biological pathway terms were cellobiose glucosidase activity (term = GO:0080079; FDR-adjusted p-value = 0.00) and phenylpropanoid biosynthesis pathway (term = KEGG:00940; FDR adjusted p-value = 0.00), respectively. Regarding *tropical japonica*, the most significant terms were related to the molecular function of polysaccharide binding (term= GO:0030247; FDR-adjusted p-value = 0.00) and symplast cellular component (term = GO:0055044; FDR-adjusted p-value = 0.00). Table S7 presents the

functional enrichment analyses results of candidate genes under recent selection in four studied ecotypes.

Discussion

Cultivated rice was mainly domesticated from wild rice into two main subspecies of *indica* and *japonica*. From the beginning to the present, this process involved genetic variation, natural and artificial selection, local adaptation, and various evolutionary events, such as bottleneck, expansion, and introgression (Sweeney *et al.* 2007; Huang *et al.* 2012). Subsequent to the domestication process, rice has been spread to various areas with different climates worldwide to meet the need for human civilization expansion. In general, *indica* has been grown in the low latitude areas with high growth temperature, while *japonica* has been cultivated in temperate regions. Additionally, selective regimes owing to cultural preference of rice farmers or breeding programs might affect favorable traits in each ecotype population and could contribute to the shaping of phenotype and the underlying genome of multiple rice ecotypes (Kiple and Ornelas 2000; Tang and Shi 2007; Fuller *et al.* 2010).

The high-throughput genomic data of different rice ecotypes has allowed studying the genetic basis of rice cultivars' differentiation during rice domestication and subsequent improvement. As rice cultivars adapted to different ecogeographical environments and were to some extent reproductively isolated, the genes that have contributed to cultivars' differences are likely to be highly divergent between their populations (Tang and Shi 2007). Here, we first investigated the population genetic structure of four cultivated rice ecotypes using high-density genome-wide polymorphism data. Then, based on the results of population genetic analyses, we designed four pairwise comparisons using F_{st} test to detect genomic regions underlying the diversification process. Moreover, we performed the test of iHS to compare the patterns of recent selection signatures among different cultivars.

A genomic region showing high signals for Fst and iHS could be the result of selection sweeps and/or presence-absence variations (PAVs) (Jensen *et al.* 2005; Sabeti *et al.* 2007). Selection sweeps occur when a beneficial mutation arises and rapidly increases in frequency in a population due to natural selection. This can result in a region of the genome with reduced genetic diversity and high levels of linkage disequilibrium (Jensen *et al.* 2005; Sabeti *et al.* 2007). F_{st} can be used to detect population differentiation resulting from selection, as it measures the extent to which genetic variation is partitioned among populations relative to the total variation (Nei and Chesser 1983). PAVs refer to the presence or absence of entire genomic segments, rather than single nucleotide variants. PAVs can also be subject to natural selection, as they can alter gene dosage or disrupt gene function (Feuk *et al.* 2006). iHS is a statistic that is commonly used to detect signatures of positive selection on single nucleotide variants that have

rapidly increased in frequency in a population. However, it has also been shown to be effective at detecting selection on PAVs (Grossman *et al.* 2010).

We used a genotypic dataset that was already generated by HDRA. This SNP array was developed using the genome of different wild and domesticated rice species (McCouch *et al.* 2016), and it includes SNPs that are informative for all the cultivars being compared in our study. Therefore, it represents the genetic diversity among different rice species populations. This could be beneficial for pairwise signatures of selection tests, such as F_{st}. However, in single-population tests, like iHS, it could potentially result in false positives or false negatives in the analysis, as the underlying genetic variation in each cultivar could be overrepresented/underrepresented. Therefore, we suggest further confirmation of our results of the iHS test using whole-genome sequence data.

Population genetic structure

In neighbor-joining tree and PCA, *indica* and *japonica* populations were placed in separate clusters, consistent with admixture results. The admixture results in K=3 indicated very limited introgression between *indica* and *japonica*, suggesting high genetic diversity between these two populations. Moreover, *aus* was located close to *indica* in both neighbor-joining tree and PCA and separated from *indica* population in K = 3. Our results indicated that *aus* is ancestrally derived from *indica* over time. In contrast, two *temperate* and *tropical japonica* subpopulations were located in the same clade in the neighbor-joining tree and PCA (PC1 and PC2). These two populations were originated from *japonica* rice in K = 4. Our results were consistent with historical information of *aus*, *indica*, and *temperate* and *tropical japonica* cultivars uncovered by previous studies (Zhao *et al.* 2010; Seo *et al.* 2020). Based on our findings in population structure analyses, we designed four comparisons aimed to investigate genomic regions underlying the diversification process of rice cultivars (Table 1). Subsequently, we applied iHS test to discover genomic regions under contemporary selection pressure.

Fixation index (F_{st})

Regarding spatially separated populations inhabiting different environments or sympatric populations that exploit different ecological niches, it is possible to identify chromosomal regions involved in adaptive divergence by comparing the levels of differentiation among multiple loci. By comparing two populations, F_{st} efficiently captures significant allele frequency differences between ecotypes and thus identifies outlier markers that are fixed or close to fixation for opposite alleles. This means that the identified signals are mainly the markers that have been differentially selected for a relatively large number of generations; therefore, they can be referred to as historical selection signals (Bomba *et al.* 2015).

Indica-japonica diversification

We detected the most significant diversification signal on chr2 (chr2:12.70-12.80 Mb) in both Ind_Tem and Ind_Tro comparisons with $Z(F_{st})_{Ind\_Tem} = 3.79$ and $Z(F_{st})_{Ind\_Tro} = 3.57$. However, no regions within or flanking this region had significant influence at subpopulation level. The study performed by Yuan et al. (Yuan *et al.* 2017) detected a close region on chr2 (chr2:13,139,975–13,160,003 bp) under significant positive selection. Gene annotation analysis of this region detected the candidate genes of *ROX2* and *OsWD40-43 ROX2* is a member of the NOL1/NOL2/sun gene family and a positive regulator of biotic stress response in rice, particularly *XA21*-mediated immunity (Lee *et al.* 2011). *XA21* gene plays a significant role in broad-spectrum immunity against the rice bacterial blight disease (Peng *et al.* 2015). This disease is one of the most devastating afflictions of inbred and hybrid rice throughout the world, specifically in southeast Asia and west Africa, induced by a Gram-negative bacterial pathogen, *Xanthomonas oryzae* pv. *Oryzae* (Niño-liu *et al.* 2006). Our study indicated that *ROX2* alleles involved in *XA21* gene regulation are highly divergent between *indica* and *japonica* cultivars, which might be associated with the geographical distribution and prevalence of this disease and the subsequent adaptation process in the local cultivars. Bacterial blight disease is distributed among rice plants in both tropical and temperate environments. However, rice in irrigated and rainfed lowland areas, with optimum temperature 25–34°C and relative humidity of more than 70%, are more prone to the disease (Sahu *et al.* 2020; Joshi *et al.* 2021). *OsWD40* performs diversified biological functions in rice, including seed growth and development (Ouyang *et al.* 2012), and recently it has been revealed that this gene is a vital regulator of anthocyanin biosynthesis (Yang *et al.* 2021).

Another significant signal detected in both Ind_Tem and Ind_Tro comparisons was located on chr10 (chr10:14.1-23.2 Mb). The genes of *OsCAD3*, *OsMB42*, *OsMB30*, *OsMB27*, and *OsMB22* were located within this genomic region and ranked in top 10 Z(Fst) values (Figure 2). Park *et al.* (2018) showed that *OsCAD3* expression is induced by wounding and cold stress, suggesting that this gene is participate in defense responses to diverse biotic and abiotic stresses. *OsMB42*, *OsMB30*, *OsMB27*, and *OsMB22* are the members of Meprin And TRAF Homology (MATH) domain containing protein (MDCP) family that has been known to be involved in biotic stress response (Kushwaha *et al.* 2016). Kushwaha *et al.* (2016) revealed that *OsMB22* might be involved in heat stress response, while all the MATH domain encoding genes showed down-regulation under cold stress.

Gene enrichment analysis was performed for 168 overlapped genes between Ind_Tem and Ind_Tro comparisons. We found 57 significantly overrepresented terms using gene enrichment of identified candidate genes (FDR adjusted p-value < 0.05). We revealed that many genes involved in regulating biological and cellular processes contribute to the diversification of *indica* and *japonica* cultivars. The

most important molecular function term was associated with UDP-glycosyltransferase enzymes activity (GO term = GO:0016757; FDR adjusted p-value = 0.014), which was overrepresented by 15 genes. UDP-glycosyltransferase is an important biotransformation superfamily of enzymes that catalyze glucosidation and help to transfer glycosyl from UDP-glycosyl donator to a variety of lipophilic compounds. This family consists of more than 200 genes in rice, which are believed to be involved in different phenotypes, particularly coping with abiotic stress and grain size (Dong *et al.* 2020; Liu *et al.* 2021a). Among these 15 candidate genes detected in this study, *OSAPT1* is essential for pollen tube germination and elongation (Liu *et al.* 2021b), *OSDGL1* for root development (Qin *et al.* 2013), *OSIRX9L* for Xylan biosynthesis (Chiniquy *et al.* 2013), and *STT3A* for response to salt stress (Koiwa *et al.* 2003).

Aus-indica diversification

We detected the most significant signal on chr1 (chr1:8.15-8.25 Mb) with $Z(F_{st})_{Aus\_Ind}$ = 5.51. Gene annotation analysis found three candidate genes of *OsEXPA8*, *OsEXPA9*, *OsGLP1-1* in this region. *OsEXPA8 and OsEXPA9* are members of the large expansins family. Expansins are key wall-loosening factors that have been implicated in controlling plant growth processes via their role as modulators of cell wall extensibility (Cosgrove 2005). *OsEXPA8* in rice plants is involved in increased plant height, enhanced leaf number, enlarged leaf size, as well as improved root system architecture, including longer primary roots and more lateral roots and root hairs (Ma *et al.* 2013). *OsEXPA9* is specifically expressed in the root of the rice (Shin *et al.* 2005), but its function is not comprehensively understood. The aforementioned genes might contribute to the root system diversification between *aus* and *indica* cultivars. A study performed by Kondo *et al.* (2000) revealed that *aus* varieties tend to have a larger xylem vessel diameter than the lowland *indica* varieties.

Gene enrichment analysis was performed using all 199 candidate genes identified by Aus_Ind comparison. The most significant terms were related to the activity of cinnamyl-alcohol dehydrogenase and sinapyl alcohol dehydrogenase enzymes, in both three genes of *OSCAD8A*, *OSCAD8B*, and *OSCAD8D* were overrepresented (Table S3). Cinnamyl alcohol dehydrogenase and sinapyl alcohol dehydrogenase are involved in the biosynthesis of lignin and culm structure in rice (Li *et al.* 2009; Ponniah *et al.* 2017). Therefore, these genes may contribute to the differences in lignin levels and culm mechanical strength between *aus* and *indica* cultivars. Further evaluation of these genes and their function are necessary to confirm our results.

Temperate japonica-tropical japonica diversification

In the Tem_Tro comparison, the most significant signal was identified on chr9 (chr9:4.87-4.97 Mb). However, gene annotation analysis did not find any genes in this region. Among the identified genes, the highest $Z(F_{st})_{Tem\_Trop}$ value was obtained by four candidate genes of *OsHCT4*, *OsGpx4*, *OsSTA170*, and

OsSCP32 located within a window on chr6 (chr6: 42.5-43.5 Mb). *OsHCT4* is a member of hydroxycinnamoyltransferases (HCTs), which catalyzes the cinnamoyl moiety transfer from hydroxycinnamoyl-CoA to various acceptors such as shikimic acid, quinic acid, hydroxylated acid, and glycerol (Kim *et al.* 2012). Ponniah *et al.* (2017) revealed that down-regulation of *OsHCT4* results in a significant reduction of lignin content. The protein encoded by *OsGpx4* belongs to the glutathione peroxidase family, which catalyzes the reduction of hydrogen peroxide, organic hydroperoxides, and lipid hydroperoxides, thereby protecting cells against oxidative damage. *OsGPX4* is critical for redox homeostasis, which is important for rice's normal growth and development (Passaia *et al.* 2014). *OsGPX4* is upregulated by drought and oxidative stress but downregulated by salinity, heat, and cold in the shoot (Islam *et al.* 2015). We showed divergent allele frequencies of genes that contributed to the lignin synthesis and stress responses between the *tropical* and *temperate japonica* cultivars. Belanger *et al.* (2015) showed that lignin signatures in tropical plants are distinguishable from temperate ones because of their high ratios of Acid/aldehyde of vanillyls and $\frac{p-hydroxyls}{\text{vanillyls+syringyls}}$. However, more assessment of lignin characteristics in *temperate* and *topical japonica* rice cultivars is necessary.

Gene enrichment analysis of Tem_Tro candidate genes identified two significant terms associated with integral and intrinsic components of peroxisomal membrane, both significantly overrepresented by three genes of *PEX11-2*, *PEX11-3*, and *OSRFP*. PEX11 gene family has been indicated to be involved in peroxisome biogenesis. In agreement with our results, Nayidu et al. (2008) indicated that rice PEX11 genes have diversification in sequences. They also showed different expression patterns of PEX11 family members under normal and various stress conditions. *PEX11-3* is responsive to abscisic acid and H_2O_2 treatments; therefore, it is potentially involved in stress responses, while *PEX11-2* shows no response to stress (Nayidu *et al.* 2008).

Recent selection signals

The iHS test identifies extended haplotypes that segregate at high frequency in the population and thus are, by definition, recent selection signals (Bomba *et al.* 2015). No overlapped candidate gene was detected among cultivars included in this study (Fig. 4). Our study showed different patterns of recent selection pressure among the four cultivars, which can further contribute to the diversification of these varieties.

The most significant iHS score in *the aus* population was located on chr1 (chr1:13,549,676:13,559,676), but no candidate was detected in this region in gene annotation. The most significant candidate gene under recent selection pressure was *COW1* which was located within a region on chr3 (chr3:3,358,056-3,368,056 bp). This gene is a member of YUCCA gene family and, through mediating indole-3-acetic acid biosynthesis, plays an essential role in maintaining root to shoot rice ratios (Woo *et al.* 2007). High root

penetration ability could maximize soil moisture capture and maintain a high plant water status under drought conditions (Chu *et al.* 2014). Moreover, deeper root distribution, longer root length, greater root length density contribute to higher grain yield, higher Nitrogen uptake efficiency, and higher Nitrogen utilization efficiency in rice (Ju *et al.* 2015); therefore, they are favorable traits in rice breeding.

Regarding *indica*, the most significant iHS score was located on chr1 (chr1:18,532,173:18,542,173 bp). However, gene annotation analysis of found intensive recent selection pressure on *OsFKBP57* and *OsSPX5* candidate genes which are located on regions on chr1 (chr1:21,375,176:21,385,176 bp) and chr3 (chr3:16,622,635:16,632,635 bp). Proteins having the SPX domain are key players in controlling processes involved in phosphate homeostasis. *OsSPX5* genes redundantly modulate recovering phosphate homeostasis after phosphate starvation (Shi *et al.* 2014). Phosphorus deficiency has been identified as one of the main nutrients limiting crop production on acidic soil of lowland tropical regions where *indica* cultivar is typically grown (Fageria *et al.* 2003; Fukuda *et al.* 2021).

With regards to *temperate japonica* the most significant region under selection pressure was detected on chr12 (chr12:10,794,098:10,804,098 bp). Gene annotation analysis did not find any candidate gene in this region. Instead, the most significant candidate genes were *OsWAK108* and *OspPGM* located within two segments on chr10 (chr10:5,194,390-5,204,390 & chr10:6,160,827:6,170,827). *OsWAK108* is a member of the wall-associated kinase (WAK) gene family, contributing to cell expansion, pathogen resistance, and heavy-metal stress tolerance in plants (de Oliveira *et al.* 2014). However, the function of *OsWAK108* is not still well understood in rice. *OspPGM* encodes a phosphoglucomutase enzyme that converts Glc-6-P to Glc1-P in the plastids of rice. Recent studies revealed the significant impact of this gene in multiple important favorable traits in rice breeding. This gene is involved in starch synthesis in rice pollen grains (Lee *et al.* 2016) and shoot gravity-sensing cells (Huang *et al.* 2021), rice yield (Pan *et al.* 2021), pollen maturation (He *et al.* 2021), and male fertility (Lee *et al.* 2016).

Concerning *tropical japonica,* the most significant signal was detected on chr9 (chr9:1,722,050:1,732,050 bp), and the most significant candidate gene of *OsPLT3* was found on chr2 (chr2: 24,257,149-24,267,149 bp). *Plethora* (PLT) genes encode transcription factors containing the AP2-domain and have been shown to be involved in regulating hormone-mediated development of main, crown, and lateral roots of rice (Li and Xue 2011). A study performed by Lavarenne *et al.* (2020) showed that *OsPLT3* might contribute to crown root primordia formation in rice. Crown roots are a type of adventitious roots making up most of the root system in rice (Rebouillat *et al.* 2009). This type of roots develop post-embryonically from the stem and play an important role in adapting plants to the soils' different hydro-mineral statuses (Ahmadi *et al.* 2014).

Gene enrichment analysis of candidate genes under significant recent selection pressure in the *aus* population found significantly overrepresented terms in connection with terpene synthase activity (Table S7). Terpene synthases are localized to the cytoplasm or chloroplast, essential mediators of ecological interactions. They play an important role in rice defense against herbivores, disease resistance, the attraction of mutualists such as wasps, as well as potentially plant-plant communication (Lee *et al.* 2015; Kiryu *et al.* 2018; Li *et al.* 2018). Regarding *tropical japonica*, we found that a term related to protein kinase molecular function was significantly overrepresented. By chemically incorporating substrate proteins with phosphate groups, protein kinases regulate the activity, localization, and protein-protein interactions. It is known that protein kinases are central components in plant responses to environmental stresses such as drought, high salinity, cold, and pathogen attack (Wang *et al.* 2020). Moreover, their role in phosphorus deficiency tolerance and seed germination and growth in rice has been characterized (Lu *et al.* 2007; Gamuyao *et al.* 2012). Regarding *temperate japonica*, we detected the phenylpropanoid biosynthesis biological pathway significantly overrepresented. Phenylpropanoids contribute to a wide range of plant responses towards biotic and abiotic stimuli. They are indicators of plant stress responses upon light or mineral treatment variation and are also key mediators of the plants' resistance towards pests (La Camera *et al.* 2004). No significant overrepresented term was found in gene enrichment analysis of *indica* candidate genes.

We detected a few candidate genes under recent selection pressure that were common between *indica* and *japonica* subpopulations (Figure 4). Among these genes, *Pi63* is a member of the R gene family cloned to *M. oryzae* for rice blast resistance (Chen *et al.* 2006). *PDIL2* is a protein disulfide isomerase that is involved in endoplasmic reticulum stress responses (Onda *et al.* 2011; Takahashi *et al.* 2014). Another candidate gene was *Osr9-LOX1*, a chloroplast-located C9 position-specific LOX, which mediates root development and resistance to chewing and piercing-sucking herbivores (Zhou *et al.* 2014). More studies with larger sample sizes are required to explore genomic regions under recent selection pressure, which are common among different rice cultivars.

The process of adaptation and natural and artificial selection have left footprints within the rice genome that can be detected by selection signatures analyses. Rice cultivars adapted to different agro-ecosystems can provide the genes variation essential for elite rice breeding. Our study involved pairwise comparisons of rice cultivars' genomes in combination with an investigation of contemporary selection signals. This combination can represent a suitable strategy for clarifying the evolutionary diversity among rice varieties. In this study, several outlier loci were detected in pairwise F_{st} comparisons, as well as recent selection signature analysis of iHS. Results from this study revealed that the genetic differentiation among the studied rice ecotypes is most likely driven by divergent selection on a wide range of traits, most importantly stress response (disease, drought, heat, cold, and salinity), roost system structure, and yield. Moreover, we found different patterns of contemporary selection among rice cultivars

proposing that modern rice breeding provides additional diversification among them. The findings of this study not only help to understand the underlying molecular basis of adaptive divergence but also provide valuable implications for rice breeding.

Declarations

Acknowledgment

The authors are grateful to all the researchers involved in Rice Diversity Project for sharing the data publicly for research purposes.

Authors' contribution

SV, MA, SS conceived the study. SV, MM, SM carried out the analyses. SV, MA, and SS interpreted the results. SV drafted the manuscript. MB, MH, and MB edited and reviewed the manuscript. All authors read and approved the final manuscript.

References

Ahmadi N., Audebert A., Bennett M. J., Bishopp A., de Oliveira A. C., Courtois B., Diedhiou A., Diévart A., Gantet P. and Ghesquière A. 2014. The roots of future rice harvests. Rice **7**, 1–9.

Alexander D. H., Novembre J. and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**, 1655–1664.

Belanger E., Lucotte M., Gregoire B., Moingt M., Paquet S., Davidson R., Mertens F., Passos C. and Romana, C. 2015. Lignin signatures of vegetation and soils in tropical environments. Adv. Environ. Res. **4**, 247–262.

Benjamini Y. and Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc., B: Stat. Methodol. **57**, 289–300.

Bomba L., Nicolazzi E.L., Milanesi M., Negrini R., Mancini G., Biscarini F., Stella A., Valentini A. and Ajmone-Marsan P. 2015. Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. Genet. Sel. **47**, 1–14.

Browning B. L., Zhou Y. and Browning S. R. 2018. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics 103, 338–348.

Chen X., Shang J., Chen D., Lei C., Zou Y., Zhai W., Liu G., Xu J., Ling Z. and Cao, G. 2006. AB-lectin receptor kinase gene conferring rice blast resistance. The plant journal. **46**, 794–804.

Chiniquy D., Varanasi P., Oh T., Harholt J., Katnelson J., Singh S., Auer M., Simmons B., Adams P. D. and Scheller, H.V. 2013. Three Novel Rice Genes Closely Related to the Arabidopsis *IRX9*, *IRX9L*, and *IRX14* Genes and Their Roles in Xylan Biosynthesis. Front. Plant Sci. **4**, 83.

Chu G., Chen T., Wang Z., Yang J. and Zhang, J. 2014. Reprint of "Morphological and physiological traits of roots and their relationships with water productivity in water-saving and drought-resistant rice." Field Crops Res. **165**, 36–48.

Cook D. E. and Andersen E. C. 2017. VCF-kit: assorted utilities for the variant call format. Bioinformatics **33**, 1581–1582.

Cosgrove D. J. 2005. Growth of the plant cell wall. Nat. Rev. Mol. Cell Biol. **6**, 850–861.

Danecek P., Auton A., Abecasis G., Albers C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T. and Sherry, S. T. 2011. The variant call format and VCFtools. Bioinformatics **27**, 2156–2158.

Decker J. E., McKay S. D., Rolf M. M., Kim J., Molina Alcalá A., Sonstegard T. S., Hanotte O., Götherström A., Seabury C. M. and Praharani L. 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genet. **10**, e1004254.

Dong N.-Q., Sun Y., Guo T., Shi C.-L., Zhang Y.-M., Kan Y., Xiang Y.-H., Zhang H., Yang Y.-B. and Li Y.-C. 2020. UDP-glucosyltransferase regulates grain size and abiotic stress tolerance associated with metabolic flux redirection in rice. Nat. Commun. **11**, 1–16.

Durinck S., Spellman P. T., Birney E. and Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. **4**, 1184–1191.

Fageria N., Slaton N. and Baligar V. 2003. Nutrient management for improving lowland rice productivity and sustainability. Advances in agronomy 80, 63–152.

Feuk L., Carson A. R. and Scherer S. W. 2006. Structural variation in the human genome. Nat. Rev. Genet. **7**, 85–97.

Fukuda M., Soma D. M., Iwasaki S., Nakamura S., Kanda T., Ouattara K. and Nagumo F. 2021. Site-specific responses of lowland rice to acidulated and calcined phosphate rock fertilizers in the Center-West region of Burkina Faso. Plos one. **16**, e0250240.

Fuller D. Q., Sato Y.-I., Castillo C., Qin L., Weisskopf A. R., Kingwell-Banham E. J., Song J., Ahn S.-M. and Van Etten J. 2010. Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol. Anthropol. Sci. **2**, 115–131.

Gamuyao R., Chin J. H., Pariasca-Tanaka J., Pesaresi P., Catausan S., Dalid C., Slamet-Loedin I., Tecson-Mendoza E. M., Wissuwa M. and Heuer S. 2012. The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature. **488**, 535–539.

Grossman S. R., Shylakhter I., Karlsson E. K., Byrne E. H., Morales S., Frieden G., Hostetter E., Angelino E., Garber M. and Zuk O. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. Science. **327**, 883–886.

He Z., Zou T., Xiao Q., Yuan G., Liu M., Tao Y., Zhou D., Zhang X., Deng Q. and Wang S. 2021. An L-type lectin receptor-like kinase promotes starch accumulation during rice pollen maturation. Development. **148**, dev196378.

Hill W. and Robertson A. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. **38**, 226–231.

Huang L., Wang W., Zhang N., Cai Y., Liang Y., Meng X., Yuan Y., Li J., Wu D. and Wang Y. 2021. LAZY2 controls rice tiller angle through regulating starch biosynthesis in gravity-sensing cells. New Phytol. **231**, 1073–1087.

Huang X., Kurata N., Wang Z.-X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W. and Guo Y. 2012. A map of rice genome variation reveals the origin of cultivated rice. Nature. **490**, 497–501.

Islam T., Manna M., Kaul T., Pandey S., Reddy C.S. and Reddy M. 2015. Genome-wide dissection of Arabidopsis and rice for the identification and expression analysis of glutathione peroxidases reveals their stress-specific and overlapping response patterns. Plant Mol. Biol. Rep. **33**, 1413–1427.

Jensen J. D., Kim Y., DuMont V. B., Aquadro C. F. and Bustamante C. D. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics. **170**, 1401–1410.

Jing Q., Spiertz J., Hengsdijk H., Van Keulen H., Cao W. and Dai T. 2010. Adaptation and performance of rice genotypes in tropical and subtropical environments. NJAS-Wagen. J. Life. Sc. **57**, 149–157.

Joshi T., Pandey S. C., Maiti P., Tripathi M., Paliwal A., Nand M., Sharma P., Samant M., Pande V. and Chandra S. 2021. Antimicrobial activity of methanolic extracts of Vernonia cinerea against *Xanthomonas oryzae* and identification of their compounds using in silico techniques. Plos one. **16**, e0252759.

Ju C., Buresh R. J., Wang Z., Zhang H., Liu L., Yang J. and Zhang J. 2015. Root and shoot traits for rice varieties with higher grain yield and higher nitrogen use efficiency at lower nitrogen rates application. Field Crops Res. **175**, 47–55.

Kim I. A., Kim B.-G., Kim M. and Ahn J.-H. 2012. Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. Phytochemistry. **76**, 25–31.

Kiple K. F. and Ornelas K. 2000. The Cambridge world history of food. Cambridge University Press.

Kiryu M., Hamanaka M., Yoshitomi K., Mochizuki S., Akimitsu K. and Gomi K. 2018. Rice terpene synthase 18 (*OsTPS18*) encodes a sesquiterpene synthase that produces an antibacterial (E)-nerolidol against a

bacterial pathogen of rice. J. Gen. Plant Pathol. **84**, 221–229.

Koiwa H., Li F., McCully M. G., Mendoza I., Koizumi N., Manabe Y., Nakagawa Y., Zhu J., Rus A. and Pardo J. M. 2003. The STT3a subunit isoform of the Arabidopsis oligosaccharyltransferase controls adaptive responses to salt/osmotic stress. The Plant Cell. **15**, 2273–2284.

Kondo M., Aguilar A., Abe J. and Morita S. 2000. Anatomy of nodal roots in tropical upland and lowland rice varieties. Plant Production Science 3(4): 437–445. Taylor & Francis.

Kushwaha H. R., Joshi R., Pareek A. and Singla-Pareek S. L. 2016. MATH-domain family shows response toward abiotic stress in Arabidopsis and rice. Front. Plant Sci. **7**, 923.

La Camera S., Gouzerh G., Dhondt S., Hoffmann L., Fritig B., Legrand M. and Heitz T. 2004. Metabolic reprogramming in plant innate immunity: the contributions of phenylpropanoid and oxylipin pathways. Immunol. Rev. **198**, 267–284.

Lavarenne J., Gonin M., Champion A., Javelle M., Adam H., Rouster J., Conejéro G., Lartaud M., Verdeil J.-L. and Laplaze L. 2020. Transcriptome profiling of laser-captured crown root primordia reveals new pathways activated during early stages of crown root formation in rice. Plos one. **15**, e0238736.

Lee G. W., Lee S., Chung M.-S., Jeong Y. S. and Chung B. Y. 2015. Rice terpene synthase 20 (*OsTPS20*) plays an important role in producing terpene volatiles in response to abiotic stresses. Protoplasma. **252**, 997–1007.

Lee I., Seo Y.-S., Coltrane D., Hwang S., Oh T., Marcotte E. M. and Ronald P. C. 2011. Genetic dissection of the biotic stress response using a genome-scale gene network for rice. PNAS. **108**, 18548–18553.

Lee J., Torollo G., Ndayiragije A., Berchmans Bizimana J., Choi I., Gulles A., Yeo U., Jeong O., Venkatanagappa S. and Kim B. 2018. Genetic relationship of tropical region-bred temperate japonica rice (*Oryza sativa*) plants and their grain yield variations in three different tropical environments. Plant Breed. **137**, 857–864.

Lee S.-K., Eom J.-S., Hwang S.-K., Shin D., An G., Okita T. W. and Jeon J.-S. 2016. Plastidic phosphoglucomutase and ADP-glucose pyrophosphorylase mutants impair starch synthesis in rice pollen grains and cause male sterility. J. Exp. Bot. **67**, 5557–5569.

Li F., Li W., Lin Y., Pickett J. A., Birkett M. A., Wu K., Wang G. and Zhou J. 2018. Expression of lima bean terpene synthases in rice enhances recruitment of a beneficial enemy of a major rice pest. Plant Cell Environ. **41**, 111–120.

Li L.-F., Li Y.-L., Jia Y., Caicedo A. L. and Olsen K. M. 2017. Signatures of adaptation in the weedy rice genome. Nat. Genet. **49**, 811–814.

Li P. and Xue H. 2011. Structural characterization and expression pattern analysis of the rice PLT gene family. Acta Biochim. Biophys. Sin. **43**, 688–697.

Li X., Yang Y., Yao J., Chen G., Li X., Zhang Q. and Wu C. 2009. FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. Plant Mol. Biol. **69**, 685–697.

Liu Q., Dong G.-R., Ma Y.-Q., Zhao S.-M., Liu X., Li X.-K., Li Y.-J. and Hou B.-K. 2021a. Rice Glycosyltransferase Gene *UGT85E1* Is Involved in Drought Stress Tolerance Through Enhancing Abscisic Acid Response. Front. Plant Sci. **12**, 790195–790195.

Liu S., Zhong J., Ling S., Liu Y., Xu Y. and Yao J. 2021b. *OsAPT1* a pollen preferentially expressed gene is essential for pollen tube germination and elongation in rice. Plant Mol. Biol. Rep. **39**, 87–97.

Londo J. P., Chiang Y.-C., Hung K.-H., Chiang T.-Y. and Schaal B. A. 2006. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, Oryza sativa. PNAS. **103**, 9578–9583.

Lu C.-A., Lin C.-C., Lee K.-W., Chen J.-L., Huang L.-F., Ho S.-L., Liu H.-J., Hsing Y.-I. and Yu S.-M. 2007. The SnRK1A protein kinase plays a key role in sugar signaling during germination and seedling growth of rice. The Plant Cell. **19**, 2484–2499.

Ma N., Wang Y., Qiu S., Kang Z., Che S., Wang G. and Huang J. 2013. Overexpression of *OsEXPA8*, a root-specific gene, improves rice growth and root system architecture by facilitating cell extension. PLoS One. **8**, e75997.

McCouch S. R., Wright M. H., Tung C.-W., Maron L. G., McNally K. L., Fitzgerald M., Singh N., DeClerck G., Agosto-Perez F. and Korniliev P. 2016. Open access resources for genome-wide association mapping in rice. Nat. Commun. **7**, 1–14.

Nayidu N. K., Wang L., Xie W., Zhang C., Fan C., Lian X., Zhang Q. and Xiong L. 2008. Comprehensive sequence and expression profile analysis of PEX11 gene family in rice. Gene. **412**, 59–70.

Nei M. and Chesser R. K. 1983. Estimation of fixation indices and gene diversities. Ann. Hum. Genet. **47**, 253–259.

Niño-liu D. O., Ronald P. C. and Bogdanove A. J. 2006. *Xanthomonas oryzae* pathovars: model pathogens of a model crop. Mol. Plant Pathol. **7**, 303–324.

de Oliveira L. F. V., Christoff A. P., de Lima J. C., de Ross B. C. F., Sachetto-Martins G., Margis-Pinheiro M. and Margis R. 2014. The Wall-associated Kinase gene family in rice genomes. Plant Sci. **229**, 181–192.

Onda Y., Nagamine A., Sakurai M., Kumamaru T., Ogawa M. and Kawagoe Y. 2011. Distinct roles of protein disulfide isomerase and P5 sulfhydryl oxidoreductases in multiple pathways for oxidation of structurally diverse storage proteins in rice. The Plant Cell. **23**, 210–223.

Ouyang Y., Huang X., Lu Z. and Yao J. 2012. Genomic survey, expression profile and co-expression network analysis of *OsWD40* family in rice. BMC genom. **13**, 1–15.

Pan X., Li Y., Zhang H., Liu W., Dong Z., Liu L., Liu S., Sheng X., Min J. and Huang R. 2021. The chloroplast-localized protein LTA1 regulates tiller angle and yield of rice. The Crop Journal. **10**, 952–961.

Park H. L., Kim T. L., Bhoo S. H., Lee T. H., Lee S.-W. and Cho M.-H. 2018. Biochemical characterization of the rice cinnamyl alcohol dehydrogenase gene family. Molecules. **23**, 2659.

Passaia G., Caverzan A., Fonini L., Carvalho F., Silveira J. and Margis-Pinheiro M. 2014. Chloroplastic and mitochondrial *GPX* genes play a critical role in rice development. Biol. Plant. **58**, 375–378.

Peng H., Chen Z., Fang Z., Zhou J., Xia Z., Gao L., Chen L., Li L., Li T. and Zhai W. 2015. Rice Xa21 primed genes and pathways that are critical for combating bacterial blight infection. Sci. Rep. **5**, 1–12.

Ponniah S. K., Shang Z., Akbudak M. A., Srivastava V. and Manoharan M. 2017. Down-regulation of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase, cinnamoyl CoA reductase, and cinnamyl alcohol dehydrogenase leads to lignin reduction in rice (*Oryza sativa* L. ssp. *japonica* cv. *Nipponbare*). Plant Biotechnol. Rep. **11**, 17–27.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D., Maller J., Sklar P., De Bakker P. I. and Daly M. J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. AJHG. **81**, 559–575.

Qin C., Li Y., Gan J., Wang W., Zhang H., Liu Y. and Wu P. 2013. *OsDGL1*, a homolog of an oligosaccharyltransferase complex subunit, is involved in N-glycosylation and root development in rice. Plant and Cell Physiol. **54**, 129–137.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Raudvere U., Kolberg L., Kuzmin I., Arak T., Adler P., Peterson H. and Vilo J. 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. **47**, W191–W198.

Rebouillat J., Dievart A., Verdeil J.-L., Escoute J., Giese G., Breitler J.-C., Gantet P., Espeout S., Guiderdoni E. and Perin C. 2009. Molecular genetics of rice root development. Rice **2**, 15–34.

Sabeti P. C., Reich D. E., Higgins J. M., Levine H. Z., Richter D. J., Schaffner S. F., Gabriel S. B., Platko J. V., Patterson N. J. and McDonald G. J. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature. **419**, 832–837.

Sabeti P. C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., Xie X., Byrne E. H., McCarroll S. A. and Gaudet R. 2007. Genome-wide detection and characterization of positive selection in human

populations. Nature. **449**, 913–918.

Sahu A., Das A., Saikia K. and Barah P. 2020. Temperature differentially modulates the transcriptome response in *Oryza sativa* to *Xanthomonas oryzae* pv. *oryzae* infection. Genomics. **112**, 4842–4852.

Seo J., Lee G., Jin Z., Kim B., Chin J. H. and Koh H.-J. 2020. Development and application of indica–japonica SNP assays using the Fluidigm platform for rice genetic analysis and molecular breeding. Mol. Breed. **40**, 1–16.

Shi J., Hu H., Zhang K., Zhang W., Yu Y., Wu Z. and Wu P. 2014. The paralogous *SPX3* and *SPX5* genes redundantly modulate Pi homeostasis in rice. J. Exp. Bot. **65**, 859–870.

Shin J.-H., Jeong D.-H., Park M. C. and An G. 2005. Characterization and transcriptional expression of the *α-Expansin* gene family in rice. Mol. Cells. 20, 210–218.

Sweeney M. T., Thomson M. J., Cho Y. G., Park Y. J. Williamson S. H., Bustamante C. D. and McCouch S. R. 2007. Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet. **3**, e133.

Takahashi H., Wang S., Hayashi S., Wakasa Y. and Takaiwa F. 2014. Cis-element of the rice PDIL2-3 promoter is responsible for inducing the endoplasmic reticulum stress response. J. Biosci. Bioeng. **117**, 620–623.

Tang T. and Shi S. 2007. Molecular population genetics of rice domestication. J. Integr. Plant Biol. **49**, 769–775.

Venables W. N. and Ripley B. D. 2013. Modern applied statistics with S-PLUS. Springer Science & Business Media.

Villanueva R. A. M. and Chen Z. J. 2019. ggplot2: Elegant graphics for data analysis (2nd ed.). Measurement: Interdisciplinary Research and Perspectives. **17**, 160-167

Voight B. F., Kudaravalli S., Wen X. and Pritchard J. K. 2006. A map of recent positive selection in the human genome. PLoS Biol. **4**, e72.

Wang P., Hsu C.-C., Du Y., Zhu P., Zhao C., Fu X., Zhang C., Paez J. S., Macho A. P. and Tao W. A. 2020. Mapping proteome-wide targets of protein kinases in plant stress responses. PNAS. **117**, 3270–3280.

Weir B. S. and Cockerham C. C. 1984. Estimating F-statistics for the analysis of population structure. Evolution. **38**, 1358–1370.

Woo Y.-M., Park H.-J., Su'udi M., Yang J.-I., Park J.-J., Back K., Park Y.-M. and An G. 2007. Constitutively wilted 1, a member of the rice *YUCCA* gene family, is required for maintaining water homeostasis and an appropriate root to shoot ratio. Plant Mol. Biol. **65**, 125–136.

Xiong Z., Zhang S., Ford-Lloyd B., Jin X., Wu Y., Yan H., Liu P., Yang X. and Lu B. 2011. Latitudinal distribution and differentiation of rice germplasm: its implications in breeding. Crop Sci. **51**, 1050–1058.

Yang X., Wang J., Xia X., Zhang Z., He J., Nong B., Luo T., Feng R., Wu Y. and Pan Y. 2021. *OsTTG1*, a WD40 repeat gene, regulates anthocyanin biosynthesis in rice. The Plant Journal. **107**, 198–214.

Yin L., Zhang H., Tang Z., Xu J., Yin D., Zhang Z., Yuan X., Zhu M., Zhao S. and Li X. 2021. rmvp: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics Bioinformatics. **19**, 619–628.

Yoshida S. and Benta W. 1983. Potential productivity of field crops under different environments. IRRI, Los Banos, Philippines.

Yuan Y., Zhang Q., Zeng S., Gu L., Si W., Zhang X., Tian D., Yang S. and Wang L. 2017. Selective sweep with significant positive selection serves as the driving force for the differentiation of japonica and indica rice cultivars. BMC genom. **18**, 1–13.

Zhang C., Dong S.-S., Xu J.-Y., He W.-M. and Yang T.-L. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. **35**, 1786–1788.

Zhao K., Wright M., Kimball J., Eizenga G., McClung A., Kovach M., Tyagi W., Ali M. L., Tung C.-W. and Reynolds A. 2010. Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PloS one. **5**, e10780.

Zhao Q., Feng Q., Lu H., Li Y., Wang A., Tian Q., Zhan Q., Lu Y., Zhang L. and Huang T. 2018. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. **50**, 278–284.

Zheng Y., Crawford G. W., Jiang L. and Chen X. 2016. Rice domestication revealed by reduced shattering of archaeological rice from the Lower Yangtze valley. Sci. Rep. **6**, 1–9.

Zhou G., Ren N., Qi J., Lu J., Xiang C., Ju H., Cheng J. and Lou Y. 2014. The 9-lipoxygenase Osr9-LOX1 interacts with the 13-lipoxygenase-mediated pathway to regulate resistance to chewing and piercing-sucking herbivores in rice. Physiol. Plant. **152**, 59–69.

Tables

Table 1. Two pairwise signature of selection studies applying fixation index (F_{st}) to find genomic regions and candidate genes involved in adaptation process of rice ecotypes.

Pairwise	Population 1		Population 2	
	Ecotype	Number of samples	Ecotype	Number of samples
Ind_Tem	*Indica*	497	*Temperate japonica*	240
Ind_Tro	*Indica*	497	*Tropical japonica*	360
Aus_Ind	*Aus*	187	*Ind*	497
Tem_Tro	*Temperate japonica*	240	*Tropical japonica*	360

Figures

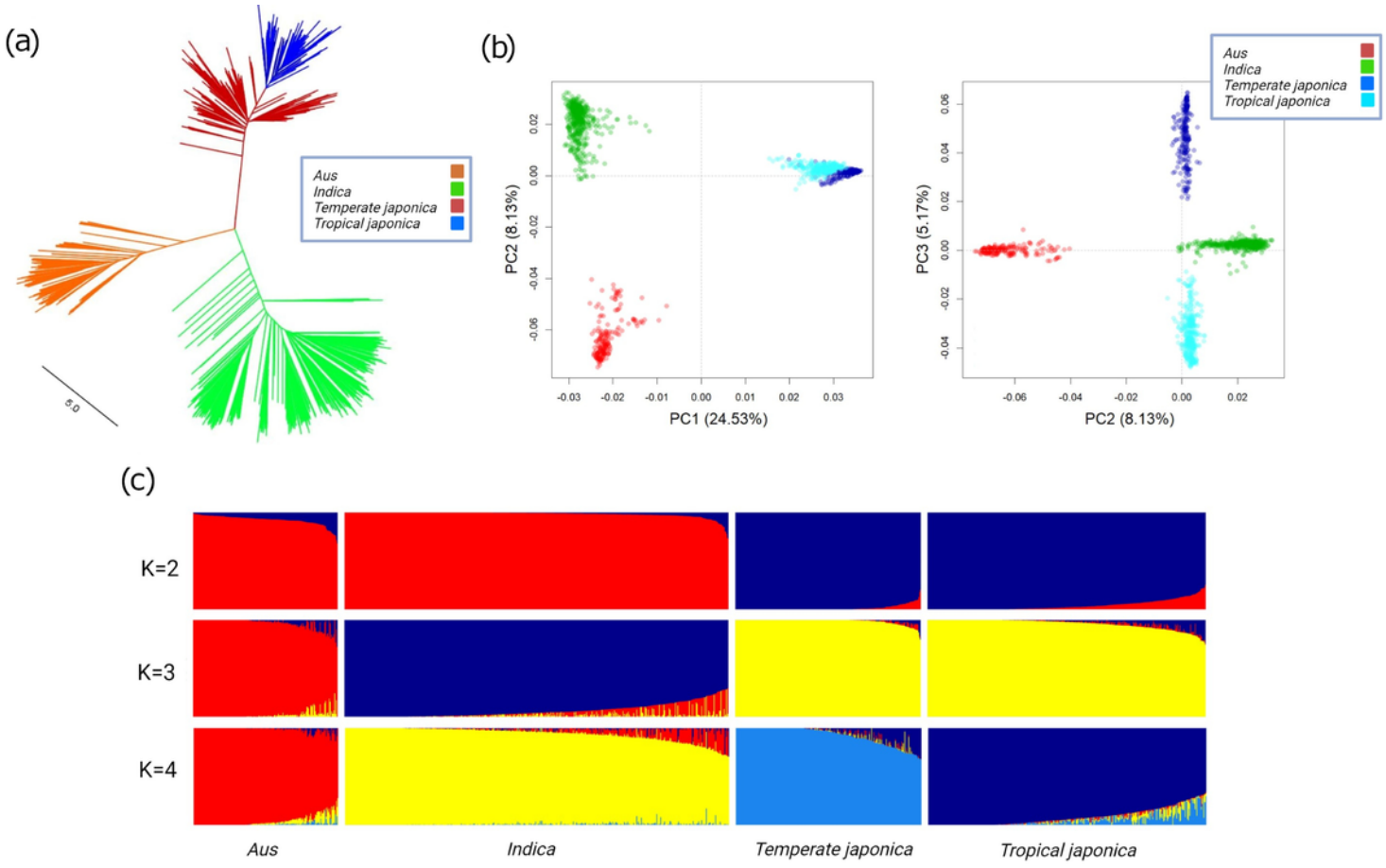

Figure 1

Population structure and relationship of *aus*, *indica*, *temperate japonica*, and *tropical japonica* ecotypes tested in this study. (a) A neighbor-joining phylogenetic tree was constructed using whole-genome SNP data. The scale bar represents pairwise distances between different samples. (b) Principal component analysis showing PC1 against PC2 and PC2 against PC3. (c) Model-based clustering of four cultivated rice ecotypes using admixture analysis with the assumed number of ancestries of 2, 3, and 4.

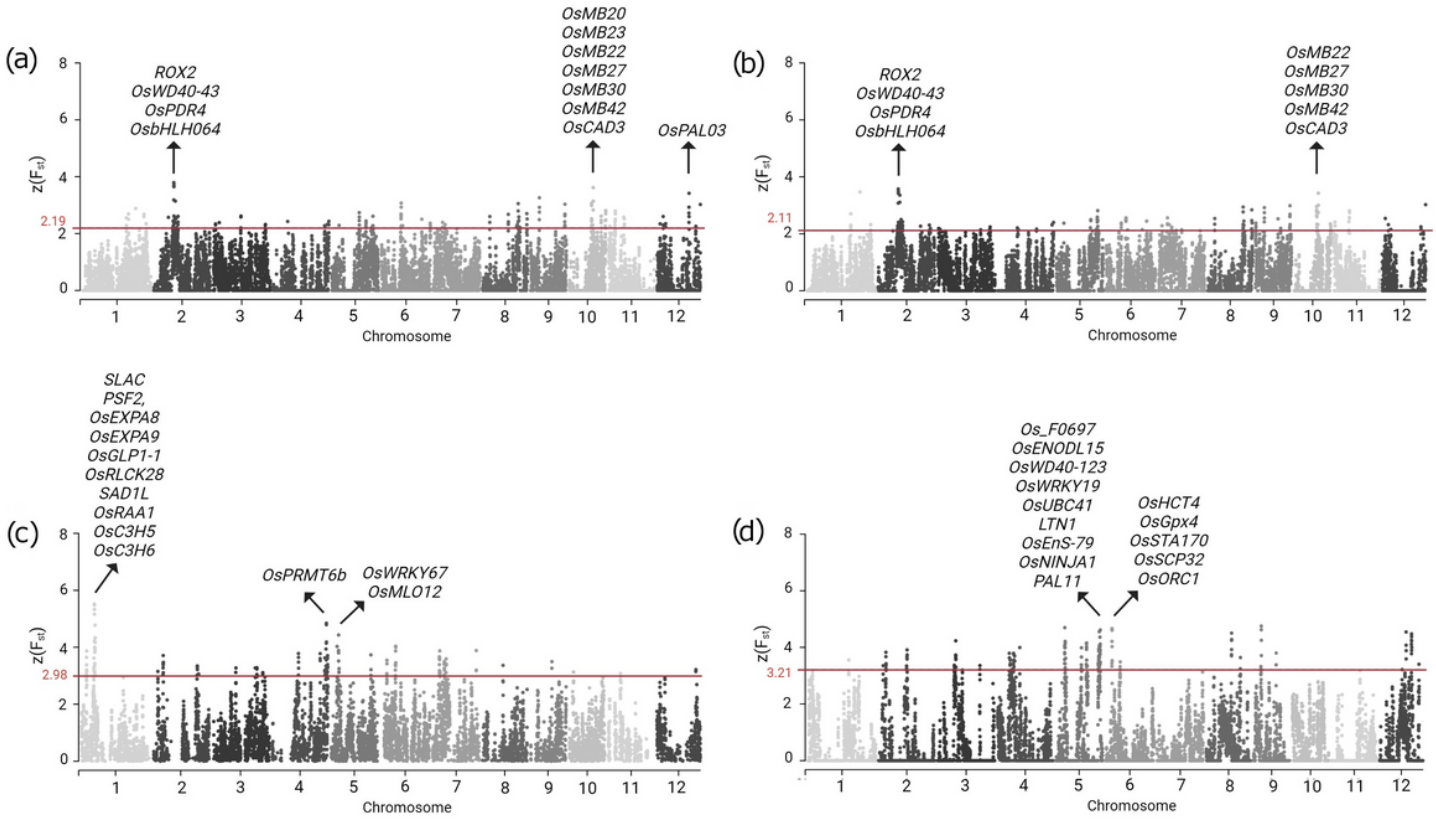

Figure 2

Manhattan plots (a), (b), (c), and (d) depict the distribution of Z-transformed F_{st} ($Z(F_{st})$) values in four pairwise analysis of *indica* versus *temperate japonica*, *indica* versus *tropical japonica*, *aus* versus *indica*, and *temperate japonica* versus *tropical japonica*, respectively. The $Z(F_{st})$ values calculated using sliding window approach. The red lines in plots (a), (b), (c), and (d) represent the significant threshold level set at the top 1% of the $Z(F_{st})$ values distribution. Genes located within the genomic regions ranked in top 10 $Z(F_{st})$ values are pointed with arrows.

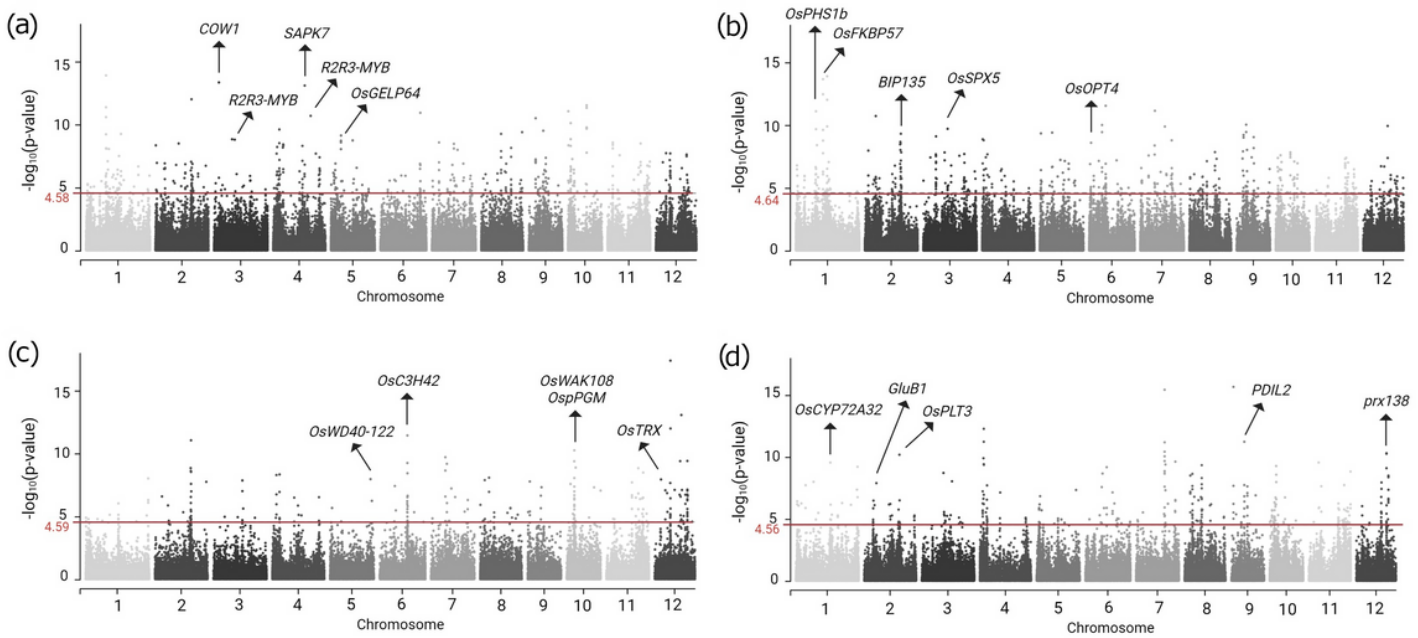

Figure 3

Manhattan plots (a), (b), (c), and (d) depict the distribution of standardized iHS values in four populations of *aus*, *indica*, *temperate japonica*, and *tropical japonica*, respectively. The red lines in plots (a), (b), (c), and (d) represent the significant threshold of q-value < 0.01. Top five significant candidate genes are pointed with arrows.

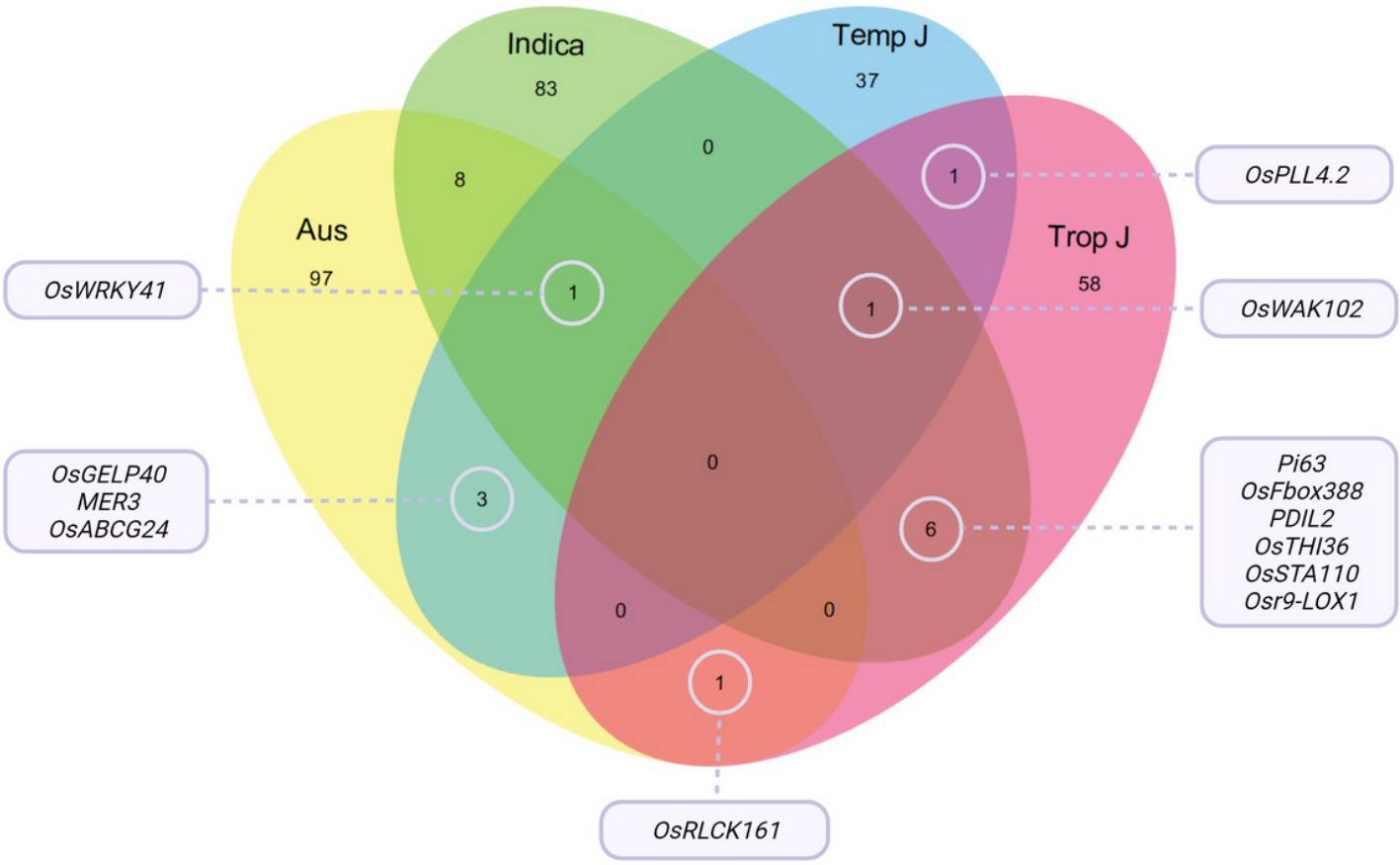

Figure 4

Overlapped candidate genes under recent selection among *aus* (Aus), *indica* (Indica), *temperate japonica* (Temp J) and *tropical japonica* (Trop J).

Supplementary Files

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

- SupplementaryTables1.pdf