JSE Journal of Systematics and Evolution

doi: 10.1111/jse.12636

Research Article

Genomic landscape of parallel domestication of upland rice and its implications

Mei-Xia Wang[1,2], Chun-Yan Jing[1,2], Xiu-Hua Wang[1,2], Zhe Cai[1], Lian Zhou[1], Mu-Fan Geng[1,2], Jing-Dan Han[1], Jie Guo[1], Fu-Min Zhang[1,2], and Song Ge[1,2]*

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
*Author for correspondence. E-mail: gesong@ibcas.ac.cn
Received 9 April 2020; Accepted 10 May 2020; Article first published online 14 May 2020

Abstract Parallel domestication has been widely acknowledged but its genetic basis remains largely unclear. As an important rice ecotype, upland rice was assumedly domesticated multiple times in two rice subspecies (*Indica* and *Japonica*) and provides a feasible system to explore the genetic basis of parallel domestication. To uncover the genome-wide pattern of genetic differentiation between upland and lowland rice and explore the parallelism of genetic changes during upland rice domestication, we obtained whole-genome sequences of 95 rice landraces and yielded genome-wide expression data for five tissues of representative accessions of upland and lowland rice. Our phylogenetic analyses confirmed multiple domestications of the upland ecotype in two rice subspecies. Genomic scans based on resequencing data identified substantial differentiation between the upland and lowland ecotypes with 11.4% and 14.8% of the genome diverged between the two ecotypes in *Indica* and *Japonica*, respectively. Further genome-wide gene expression analyses found that 30% of effectively expressed genes were significantly differentiated between two ecotypes, indicating the importance of regulation changes in the domestication of upland rice. Importantly, we found that only 1.8% of differentiated genomes and 1.6% of differentially expressed genes were shared by upland *Indica* and upland *Japonica*, suggestive of largely unparallel genetic alterations during upland rice domestication. These findings not only provide new insights into the genetic basis of parallel domestication at the genome scale but could also facilitate genetic improvement and breeding of rice and crops in general.

Key words: domestication, genomics, parallel evolution, upland rice.

1 Introduction

Parallel evolution or parallelism is an important phenomenon and a hot topic in evolutionary biology because it presents strong evidence for adaptation and action of natural selection (Nosil, 2012; Bolnick et al., 2018). To uncover whether and under what conditions phenotypic parallel is associated with parallel at the genetic level could yield additional insights into the predictability of the adaptive process and evolution in general at the genetic level (Conte et al., 2012; Blount et al., 2018). Multiple lines of evidence showed that parallel evolution might occur due to selection for various genetic variations, including those either from the same gene (e.g., Stellari et al., 2010; Jiang et al., 2012; Linnen et al., 2013; Wang et al., 2018; Yang et al., 2018; Zhang et al., 2019) or from different genes in the same pathway (e.g., Bao et al., 2011; van der Knaap et al., 2014; Somssich et al., 2016), and those from the genes in different pathways (e.g., Guan et al., 2012; Blanca et al., 2015; Somssich et al., 2016; Lyu et al., 2020). Parallelism in phenotypes might also evolve from multiple components other than selection, including species history and demography, population size, and gene flow as well as genomic architecture (Conte et al., 2012; Gaut, 2015; Bolnick et al., 2018). In addition, it is increasingly acknowledged that parallel evolution should be treated as a quantitative continuum ranging from parallel to non-parallel, rather than a binary phenomenon because empirical studies have showed that replicate populations and species in similar environments might evolve either similar traits (or genes) or dissimilar traits (or genes) (Bolnick et al., 2018). Therefore, given the complexity of parallelism, the prevalence of parallel evolution has long been debated and disagreements still exist regarding how strong and how variable the parallel evolution can be in the wild (Bolnick et al., 2018).

Parallel domestication is a special type of parallel evolution in which many traits or phenotypes evolve in response to human selection; these modified phenotypic traits are collectively known as the domestication syndrome, such as increased seed size, change in plant habit, loss of seed dormancy, loss of bitterness, and loss of shattering and seed dispersal (Doebley et al., 2006; Pickersgill, 2018; Woodhouse & Hufford, 2019). To date, extensive studies have improved

our understanding of the origin and domestication of many crop species as well as adaptation and diversification in post-domestication (Doebley et al., 2006; Gaut, 2015; Woodhouse & Hufford, 2019). Previous studies have showed that similar domestication syndromes or traits can be shared either by closely related species or by distinctly related taxa (Pickersgill, 2018; Woodhouse & Hufford, 2019). It is also well established that parallelism in domestication traits might arise from genetic alterations in metabolic pathway, gene, DNA sequence, and expression regulation during the process of domestication (Gaut, 2015; Pickersgill, 2018; Woodhouse & Hufford, 2019). Recent studies further indicated that evolution of similar phenotypes during crop adaptation or diversification might not be associate with parallel genetic changes, especially for the traits involving complex developmental pathways or complex networks of genes (Blount et al., 2018; Pickersgill, 2018; Walden et al., 2020). Despite this progress, genome-wide variations at the expression level during domestication and the role of natural selection in crop adaptation and diversification require further investigations (Gaut, 2015; Bolnick et al., 2018; Pickersgill, 2018). It is also unclear whether or to what extent the phenotypic parallel of domestication syndrome is associated with parallelism at the genetic levels (i.e., pathway, gene, and sequence) (Bolnick et al., 2018; Woodhouse & Hufford, 2019).

Rice (*Oryza sativa* L.) is one of the most important crops in the world, with two subspecies (*Indica* and *Japonica*) that are distinct in gross morphology, geographic distribution, habitats/ecosystems, and cultivation culture (Khush, 1997; Sang & Ge, 2007). Rice could also be divided into two major ecotypes (upland and lowland rice) according to its dependence on water (Chang, 1976; Khush, 1997). The lowland or irrigated rice is grown in irrigated fields with water maintenance, whereas the upland or dryland rice is adapted to rainfed, naturally well drained soils in hilly areas (Khush, 1997; Bernier et al., 2008; Lyu et al., 2014). It is well known that upland rice was domesticated to adapt to drought environments and thus widely cultivated in the hilly areas of Southeast Asia, West Africa, and Latin America (Khush, 1997; Bernier et al., 2008). Morphologically, upland rice is taller with low tillering capacity and early maturity, and tends to have longer and thicker roots than lowland rice (Chang, 1976; Lyu et al., 2014). Given its agricultural importance, upland rice has been a major target of various investigations involving morphological and genetic diversity (Lian et al., 2006; Lyu et al., 2014; Xia et al., 2014, 2019), genetic mapping and functional studies of agriculturally important traits (Li et al., 2005; Rabello et al., 2008; Uga et al., 2011; Li et al., 2015; Zhao et al., 2018), as well as its unique characteristics of drought tolerance or resistance (Lian et al., 2006; Zhao et al., 2018; Xia et al., 2019; Lyu et al., 2020), although relatively few studies have been undertaken at the genome scale (Lyu et al., 2014; Xia et al., 2019). Moreover, previous studies on the origin and population genetics of rice diversity suggested that upland rice might be domesticated independently in *Indica* and *Japonica* (Lyu et al., 2014; Wang et al., 2014; Xia et al., 2014) and is presumably a case of parallel domestication. Therefore, upland rice along with its progenitors provides a unique system to explore the patterns and underlying mechanisms of parallel evolution and adaptation of crop species.

In this study, we investigated the evolutionary relationships and population genetics of representative accessions of upland and lowland rice based on resequencing and RNA sequencing (RNA-seq) data. First, we uncovered the population genetic structure based on the genome-wide sequence and expression data and confirmed the hypothesis that upland rice domesticated multiple times in two rice subspecies (*Indica* and *Japonica*). Second, we explored the genome-wide patterns of genetic differentiation between upland and lowland rice in *Indica* and *Japonica* at the sequence and expression levels, and assessed the importance of regulation changes in the domestication of upland rice. Finally, we were interested in how specific and shared were the pathways and genes that associated with the domestication of upland rice in *Indica* and *Japonica* and discussed their implications for the origin of upland rice and crop domestication in general. To answer these questions not only provides additional insights into understanding of the repeatability for evolution, but also facilitates genetic improvement and breeding of rice and crops in general.

2 Material and Methods

2.1 Whole-genome resequencing data and RNA-seq data

To explore comparatively the differentiation between upland and lowland rice, we generated two genome-scale datasets for analyses at the sequence and expression levels. The first dataset includes the resequencing data of 95 accessions of rice landraces, including 13 upland and 43 lowland accessions from *Indica* and 12 upland and 27 lowland accessions from *Japonica* (Table S1). Of them, whole-genome sequences of 82 accessions were downloaded from the 3000 Rice Genome Project (3KRGP) (Li et al., 2014) and the remaining 13 accessions were resequenced in this study by Illumina HiSeq 4000 (Beijing, China) (Tables S1, S2).

The other dataset consists of RNA-seq data of a subset of the resequencing samples. We chose 12 accessions of rice landraces (four upland and eight lowland cultivars) from the above 95 samples for studying genome-wide gene expression (Table S1). Specifically, we grew these 12 accessions in the phytotron in the Institute of Botany, Chinese Academy of Sciences (Beijing, China) at the controlled condition of 30 °C (daytime, 08:00–20:00) and 25 °C (night-time) with humidity of 50%. Each accession was cultivated using at least five individuals for obtaining RNAs at different developmental stages. We collected five types of tissues, that is, leaves at the seedling stage (S), flag leaves (LH), and panicles (PH) at the heading stage, flag leaves (LM) and panicles (PM) at the milk stage, for RNA-seq by Illumina HiSeq 2000. Total RNA was isolated with an SV Total RNA Isolation kit (Promega, Madison, Wisconsin, USA) following the manufacturer's instructions.

2.2 Mapping reads of resequencing data and SNP calling

We obtained an average of 14G clean reads of resequencing data for each sample (Table S2). We mapped the clean reads to the Nipponbare genome (MSU version 7.0; http://rice.plantbiology.msu.edu/pub/data) using BWA (-mem) (Li & Durbin, 2009) and found that, on average, 98.5% of the reads were mapped to the reference genome, with the mean

sequencing depth being 16.6× and the average coverage of the reference genome being 93.1% (Table S2). The "HaplotypeCaller" in GATK 3.5 was used for single nucleotide polymorphism (SNP) calling with parameters "-stand_emit_conf 10, -stand_call_conf 30." To reduce false positives, the raw SNP data were filtered by variant quality score recalibration and a total of 9 282 239 SNPs was obtained for subsequent analyses. To avoid the potential impact of artificial selection on the analyses, we also selected SNPs in the intergenic regions (5 678 050; 61.2% of the total SNPs), as the putatively neutral sites, for various demographic and population genetic analyses.

2.3 Analyses of population genetics and phylogeny

We used the sliding window approach to estimate the genetic diversity (π, 100 kb window sliding in 10 kb steps) of different groups of samples, and the between-group genetic differentiation (F_{ST}, 10 kb window sliding in 10 kb steps) by vcftools (Danecek et al., 2011). Principal component analysis (PCA) was applied to examine the genetic subdivision of all samples using PLINK (version 1.07) (Purcell et al., 2007). We also calculated the pairwise genetic distance of all 95 samples to get a genetic distance matrix using our in-house PERL scripts and constructed a neighbor-joining (NJ) phylogenetic tree by MEGA (version 6.0) (Tamura et al., 2013) based on the distance matrix. The population structure and admixture of all 95 samples were inferred using ADMIXTURE (version 1.3.0) (Alexander et al., 2009) with five replicates and five-fold cross-validation from $K = 2$ to 6 based on the neutral SNPs and all SNPs, respectively.

2.4 Identification of genomic regions that associated with domestication of upland rice

First, we detected the significantly differentiated genomic regions in non-overlapping 10 kb windows along the entire genome based on the divergence (F_{ST}) between upland and lowland rice in each subspecies. We used a bootstrapping method to evaluate the significance of the F_{ST} value while accounting for SNP density per window. For a real window containing *n* SNPs, we simulated 100 pseudo-windows by randomly sampling *n* SNPs with replacement from the genome and calculated a null distribution of the simulated values. The window that fell in the lowest 5% tail of the simulated F_{ST} distribution was defined as a significantly differentiated window (hereafter, outlier window) (Guo et al., 2016). The simulations were carried out using our in-house R scripts. To identify windows with significantly lower diversity in upland rice compared to lowland rice, we calculated the reduction of diversity (*ROD*) of all 10 kb windows along the genome based on the formulation $ROD = \pi\_upland/\pi\_(upland + lowland)$. We considered the outlier windows in the lowest 5% tail of the *ROD* distribution as the putative selected windows, that is, windows with selective sweeps in upland rice.

2.5 Mapping reads of RNA-seq data and identifying differentially expressed genes between upland and lowland rice

After removing reads of low quality and reads containing sequencing adapters, we obtained more than 30M clean read pairs from each RNA-seq sample with the average clean data around 6G for each sample (Table S3). We mapped the clean reads to the reference genome (Nipponbare, MSU version 7.0) by TopHat2 (Kim et al., 2013) and obtained mapping rates ranging from 89.7% to 92.1% depending on the tissues (Table S3). Expression level was measured with the HTSeq.scripts.count feature from HTSeq (version 0.6.1p1) (Anders et al., 2015). Only reads that were uniquely mapped to the reference genome were chosen for computing gene expression values. We calculated cpm (count per million) and RPKM (reads per million reads) to normalize gene expression levels using edgeR (Chen et al., 2014). The trimmed mean of *M*-values method was invoked during normalization procedures. We defined a gene as an effectively expressed gene (EEG) if more than one of the samples had at least one mapped read count per million (rowSums(cpm(d) > 1) ≥ 1). In total, we obtained 29 818 EEGs, that is, the genes that were expressed effectively in at least one tissue. To explore the evolutionary relationship among samples, we calculated pairwise distances of all 12 samples based on the expression quantity of EEGs using our in-house R scripts and acquired a genetic distance matrix for each tissue. Based on the distance matrix, we constructed an NJ phylogenetic tree using PHYLIP (version 3.696) (Felsenstein, 2005).

We detected differentially expressed genes (DEGs) between upland and lowland rice in two subspecies separately by edgeR. Following Guo et al. (2016), we used empirical Bayesian analysis to improve the statistical power of small samples and provided the generalized linear model to take major sources of variation into account by fitting a generalized linear model with a design matrix. We set the false discovery rate (FDR) < 0.05 as the significance threshold for detecting DEGs. These analyses were undertaken independently for each tissue.

2.6 Gene ontology enrichment analysis

To explore candidate pathways and genes that might be related to domestication of the upland rice, we used the R package topGO (version 2.34.0) to conduct gene ontology (GO) enrichment analysis, in which all EEGs were used as the background for analysis of the RNA-seq data and all genes of the reference genome were used as the background for analyzing the resequencing data. We analyzed comparatively the top 20 GO terms (by *P*-value using Fisher's weight01 exact test) of biological process (Alexa et al., 2006) in the enrichment analyses. We also obtained functional information of the selected genes under study from published papers and various databases.

2.7 Detection of distribution patterns of differentiated windows and DEGs

We mapped the physical positions of the outlier and putative selected windows and DEGs between upland and lowland rice across the rice genome to test whether they are randomly distributed in the genome. We further used the χ^2-test to detect regionally enriched clusters for the outlier and putative selected windows with window size of 100 kb with step size of 10 kb, and the DEGs with window size of 200 genes with step size of 10 genes across the genome using our in-house PERL scripts. We set FDR < 0.05 as

the significance threshold for distribution in clusters across the genome (Guo et al., 2016).

3 Results

3.1 Population genetic structure based on resequencing and expression data

We first investigated the genetic diversity and population genetic structure of all samples based on neutral variants from intergenic regions of resequencing data. We detected a higher level of genetic diversity in *Indica* ($\pi = 0.0019$) than in *Japonica* ($\pi = 0.0011$) as expected; the upland ecotype maintained a comparable (slightly higher) amount of genetic diversity to the lowland ecotype in either *Indica* or *Japonica* (Fig. S1). An NJ tree clearly indicated that all 95 samples formed two major groups corresponding to the two subspecies *Indica* and *Japonica*, and the upland ecotype were found in either *Indica* or *Japonica* (Fig. 1A). Similarly, the PCA result with PCA1 explained as high as 44.9% of the variation grouped all samples into two clusters corresponding to *Indica* and *Japonica* without any overlap, and the upland and lowland ecotypes were mixed together within subspecies (Fig. 1B), consistent with the F_{ST} estimates in which genetic differentiation between upland and lowland rice (0.004 in *Indica* and 0.085 in *Japonica*) is substantially lower than that between two subspecies (0.528–0.617) (Fig. S2). Further ADMIXTURE analysis showed the deepest splits ($K = 2$) occurred between two subspecies (*Indica* and *Japonica*) and that upland rice accessions scattered in each of the two subspecies when *K* values were over 3 (Fig. 1C). Using total SNPs, we undertook all the analyses and obtained same results (Fig. S3).

We then analyzed the RNA-seq data and found a pattern that is consistent with the result of the resequencing data, that is, the NJ tree based on the expression quantity of EEGs of each of the five tissues grouped all samples into two major clades corresponding to two subspecies with upland rice found in each subspecies (Figs. 1D, S4). Together, these analyses clearly show that upland rice originated multiple times from *Indica* and *Japonica*.

3.2 Genomic variation patterns of upland rice at the sequence level

Based on the F_{ST} permutation, we detected 4249 and 5500 outlier windows between upland and lowland rice in *Indica* and *Japonica*, respectively, accounting for 11.4% (*Indica*) and 14.8% (*Japonica*) of the total windows (Table 1). Of these outlier windows, 682 accounting for 1.8% of the total windows overlapped (Fig. S5A) and might be candidate regions involving parallel domestication of the upland ecotype. By annotating the genes that fell within the overlapped outlier windows, we identified 1222 genes that accounts for 2.2% of the total genes (Table 1). To evaluate whether the outlier windows are randomly distributed in the genome, we mapped the physical positions of all outlier windows across the rice genome and found that they were not proportionately distributed along the chromosomes (χ^2-test, FDR < 0.05), with 22.4% and 25.0% of the outlier windows occurring in clusters across all 12 chromosomes in *Indica* and *Japonica*, respectively

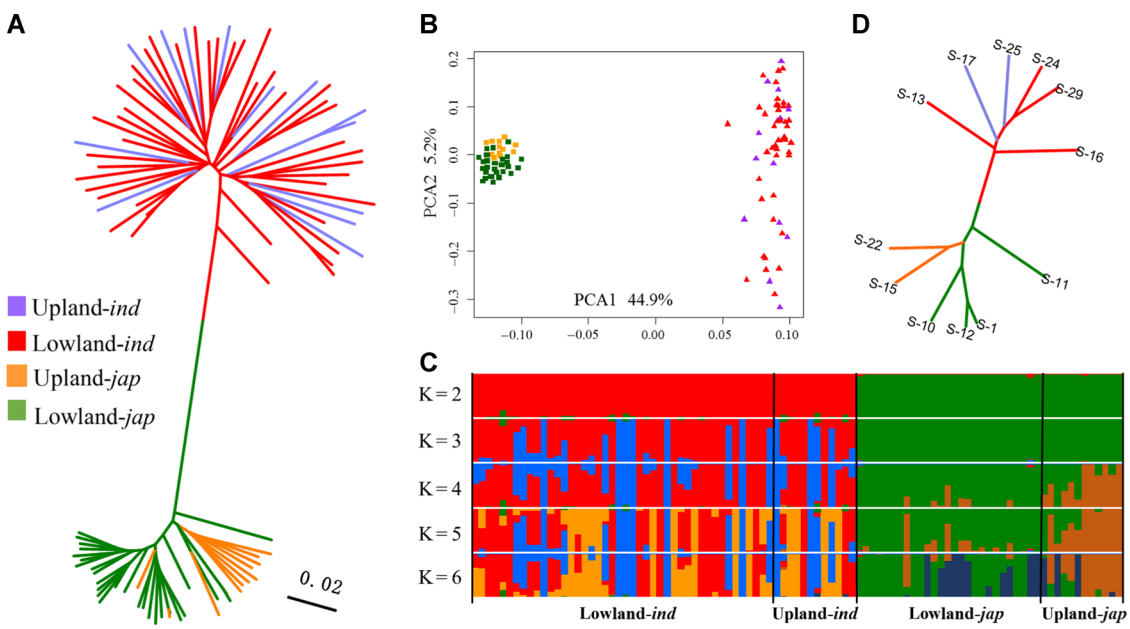

Fig. 1. Analyses of population genetic structure of all 95 rice accessions based on the neutral single nucleotide polymorphisms of resequencing data and the expression quantity of RNA sequencing data. **A**, Neighbor-joining tree based on resequencing data. **B**, Principal component analysis (PCA) based on resequencing data. **C**, Model-based population assignments at *K* from 2 to 6. Each vertical bar represents a sample, with its assignment probability to genetic clusters represented by different colors. **D**, Neighbor-joining tree based on expression quantity of effectively expressed genes.

Table 1 Summary of genomic differentiation between upland and lowland rice in *Indica* and *Japonica* subspecies

	Outlier window		Putative selected window	
Upland–lowland contrast	No. windows (% total windows)	No. genes within windows (% total genes)	No. windows (% outlier windows)	No. genes within windows (% genes in outlier windows)
Within *Indica*	4249 (11.4)	6981 (12.5)	333 (7.8)	586 (8.4)
Within *Japonica*	5500 (14.8)	8737 (15.7)	187 (3.4)	352 (4.0)
Overlap	682 (1.8)	1222 (2.2)	1 (0.1)	3 (0.2)

(Figs. 2, S5), suggesting that approximately one-fourth of the outlier windows involved in genetic differentiation between two ecotypes might be non-randomly distributed across the genome.

To identify the outlier windows that might be under selection and thus be related to domestication of upland rice, we calculated the *ROD* of 10 kb windows along the genome and treated the outlier windows that were in the lowest 5% tail of the *ROD* distribution as the putative selected windows in upland rice. We obtained 333 and 187 putative selected windows in *Indica* and *Japonica*, respectively, accounting for 7.8% and 3.4% of the outlier windows in the two subspecies (Table 1). Such a small proportion of outlier windows with a signature of selection implies that possible factors other than artificial selection might contribute to domestication of upland rice. It is worthwhile noting that only one putative selected window and three genes within the window were shared by *Indica* and *Japonica* (Fig. 3; Table 1). The three shared genes might be involved in parallel domestication of the upland ecotype but information on their functions is not available from published papers and various databases.

3.3 Candidate pathways related to domestication of upland rice

We explored the annotated functions of the genes that fell within the outlier and putative selected windows, respectively, based on resequencing data. As shown in Table 1, 6981 and 8737 genes reside in the outlier windows for *Indica* and *Japonica*, respectively. The GO annotation of these genes indicated that 10 out of the top 20 biological process terms that were enriched in the outlier windows were shared by *Indica* and *Japonica* (Fig. S6), suggestive of a higher level of parallelism at the pathway level than at the gene level. These terms included many basic biological processes such as transport, regulation of gene expression, signal transduction, and various metabolic processes, reflective of similar physiological and metabolic responses during the domestication of the upland *Indica* and upland *Japonica*.

By further analyzing the genes that fell in the putative selected windows and thus might associate with domestication of upland rice, we found 586 and 352 genes within the putative selected windows for *Indica* and *Japonica*, respectively (Table 1). The GO biological process characterization of these genes showed that, out of the top 20 terms, only six were shared by upland *Indica* and upland *Japonica*, while the remaining 14 terms did not overlap between upland *Indica* and upland *Japonica* (Fig. 4). The six shared terms were mainly stress-related, suggesting that the response to abiotic (environmental) stress was a common feature of the upland *Indica* and upland *Japonica*. Notably, the distinct terms in upland *Indica* are related to many vegetative processes such as transport, growth, photosynthesis, anatomical structure morphogenesis, and various metabolic processes; those specific to upland *Japonica* involved mainly the reproduction-related processes such as pollination,

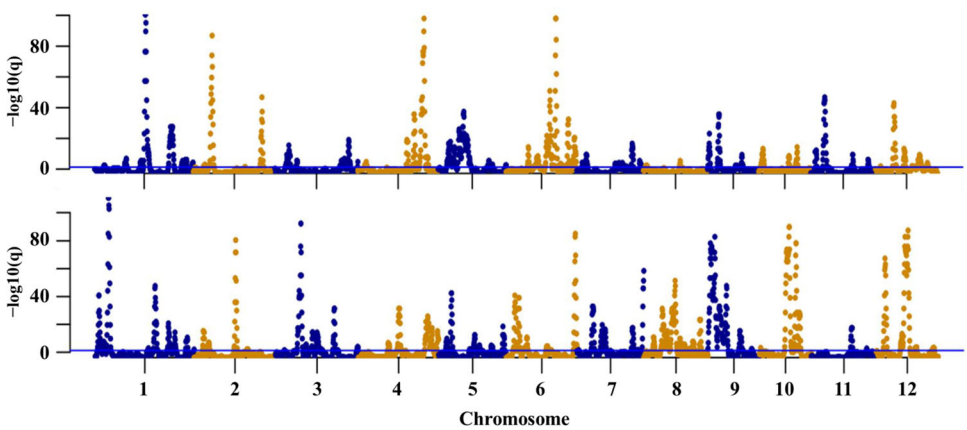

Fig. 2. Distribution patterns of the outlier windows in *Indica* (top) and *Japonica* rice (bottom) on chromosomes. Each dot represents a 10 kb window. Alternating blue and yellow colors indicate different chromosomes.

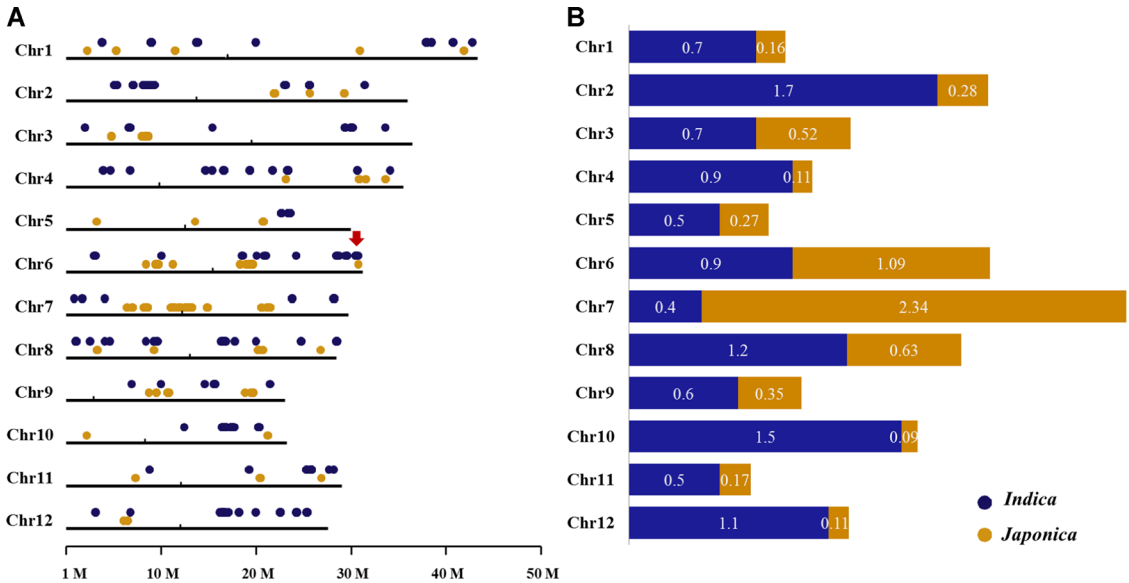

Fig. 3. Distribution patterns of the putative selected windows on chromosomes. **A**, Physical positions of the outlier windows with signature of selection along 12 chromosomes (Chr). Red arrow indicates the only putative selected window shared between *Indica* and *Japonica*. **B**, Proportion of the putative selected windows to the total number of windows in each chromosome.

flower development, and reproduction and reproductive process (Fig. 4). These differences might reflect different strategies of adaptation of upland *Indica* and upland *Japonica* to environments during their domestication.

3.4 Differentially expressed genes between upland and lowland rice

We analyzed the DEGs between upland and lowland rice in *Indica* and *Japonica* separately. A total of 8905 DEGs between upland and lowland rice were identified across five tissues, accounting for 29.9% of the total number of EEGs (29 818), with the number of DEGs being larger in *Japonica* (8131, 27.3% of EEGs) than in *Indica* (1293, 4.3% of EEGs) and more DEGs being in tissue PM than the other tissues for both subspecies (Table 2), implying that the expression differentiation between upland and lowland rice occurs mainly in reproductive stage. Again, a very small proportion of DEGs was shared by two subspecies, ranging from 1.2% (tissue S) to 2.3% (tissue LH), suggesting that expression changes during domestication of upland rice seem not parallel in *Indica* and *Japonica* at the

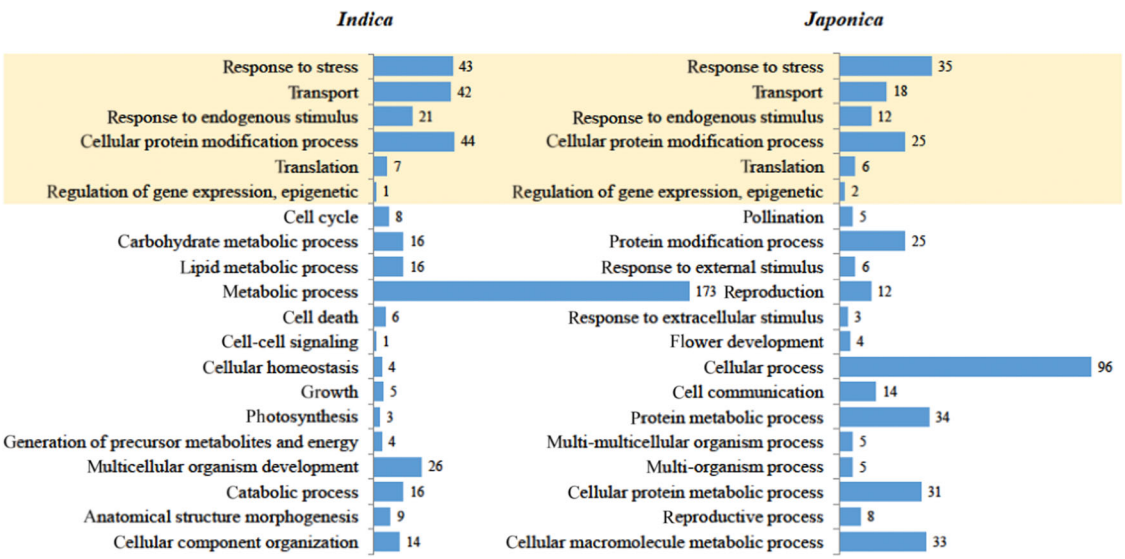

Fig. 4. Gene ontology enrichment of the genes in the putative selected windows in upland rice. Of the top 20 gene ontology biological process terms, six (shaded) were shared by upland *Indica* (left) and upland *Japonica* (right). Numbers of genes for each term are shown to the right of the bars.

Table 2 Number of differentially expressed genes (DEGs) between upland and lowland rice in *Indica* and *Japonica* subspecies

Tissue	No. DEGs in *Indica* (% EEGs)	No. DEGs in *Japonica* (% EEGs)	No. of DEGs combined (% EEGs)	No. of DEGs overlapped (% DEGs combined)	No. DEGs with same direction of expression in two subspecies
S	561 (1.9)	350 (1.2)	900 (3.0)	11 (1.2)	6
LH	399 (1.3)	565 (1.9)	942 (3.2)	22 (2.3)	15
PH	164 (0.6)	556 (1.9)	711 (2.4)	9 (1.3)	4
LM	280 (0.9)	1409 (4.7)	1653 (5.5)	36 (2.2)	29
PM	330 (1.1)	6655 (22.3)	6901 (23.1)	84 (1.2)	19
Total	1293 (4.3)	8131 (27.3)	8905 (29.9)	144 (1.6)	67

EEG, effectively expressed gene; LH, flag leaves at the heading stage; LM, flag leaves at the milk stage; PH, panicles at the heading stage; PM, panicles at the milk stage; S, leaves at the seedling stage.

gene level. We further detected the distribution pattern of DEGs across chromosomes by the χ^2-test (FDR < 0.05) and found that 2.7% and 17.9% of DEGs were distributed in clusters in *Indica* and *Japonica*, respectively (Fig. S7). By applying GO annotation analysis on the 1293 DEGs in *Indica* and 8131 DEGs in *Japonica*, we found that 10 of the top 20 terms overlapped in the two subspecies, with only two terms being the same as those found based on the outlier window analyses (Fig. S8). Similar to the results at the sequence level, genome-wide expression analysis also indicated that parallelism was higher at the pathway level than at the gene level.

Of 144 DEGs that were shared by the two subspecies, 67 (ranging from 4 to 29 in different tissues) (Table 2) showed the same direction of expression changes in upland rice of the two subspecies. To detect the biological functions of the 67 genes, we searched published reports and found nine genes with important known functions (Table S4), including the zinc finger transcription factor *DST* (LOC_Os03g57240) that regulated drought and salt tolerance in rice (Huang et al., 2009), *OsCYP20-2* (LOC_Os05g01270) that is involved in photosynthetic acclimation to help plants cope with environmental stress and enhance multiple abiotic stress tolerance (Kim et al., 2012), as well as several WRKY transcription factors (*OsWRKY62*, *OsWRKY45*, and *OsWRKY50*) that enhanced rice diverse stress responses and disease resistance (Chen et al., 2017). It is evident that these genes involve various functions associated with abiotic resistance and are most likely to evolve in parallel during upland rice domestication; they deserve further in-depth investigations using molecular and functional approaches.

4 Discussion

4.1 Upland rice was domesticated multiple times in *Indica* and *Japonica*

As a well-known ecotype of rice, upland rice has been extensively investigated with the focus on drought resistance (see reviews in Bernier et al., 2008; Xia et al., 2019). However, the domestication history or times of origin of upland rice have never been investigated in detail despite the fact that some of studies implied that upland rice might be domesticated independently in *Indica* and *Japonica* (Lyu et al., 2014; Wang et al., 2014; Xia et al., 2014). Using 47 SSR markers located in drought responding expressed sequence tags, Xia et al. (2014) genotyped 377 rice landraces from China and found that all the samples formed two major groups with each consisting of both upland and lowland rice. Based on 84 upland and 82 lowland accessions of rice from all over the world, Lyu et al. (2014) confirmed a single origin of upland *Japonica* and detected a novel multiple origin pattern in upland *Indica* (i.e., upland *Indica* accessions were bred multiple times). Our phylogenetic study based on genome-wide sequence and expression data showed clearly that the upland ecotype was domesticated in parallel in two rice subspecies (*Indica* and *Japonica*) (Fig. 1). In addition, we found that the upland accessions in either *Indica* or *Japonica* did not form a monophyletic clade, which might result either from multiple domestications or from extensive gene flow between upland accessions of different subspecies. The hypothesis of multiple domestications of upland rice within subspecies seems more plausible given the fact that the upland accessions maintained a comparable level of genetic diversity to the lowland accessions in both *Indica* and *Japonica* (Fig. S1). Nevertheless, the complex genetic structure of rice accessions within subspecies required further investigations using a larger collection of samples and more powerful statistical analysis.

4.2 Substantial differentiation between upland and lowland rice at both sequence and expression levels

It is well understood that the genomic divergence between species or populations was heterogeneous across the genome and varied from low to high at both the sequence and expression levels depending on the organisms and factors underlying the divergence (Nosil, 2012; Shibata et al., 2012; Bolnick et al., 2018). However, relative few studies on crop species have been undertaken to explore the genome-wide variations at the expression level and to reveal the role of natural and artificial selection in regulatory evolution (Lai et al., 2008; Swanson-Wagner et al., 2012; Sauvage et al., 2017). Previous studies on natural populations of many species have indicated the importance of expression changes in adaptation and speciation (Wolf et al., 2010; Guo et al., 2016). For examples, in a study of speciation involving two crow species, Wolf et al. (2010) detected an almost complete lack of sequence divergence at 25 nuclear intronic loci but a clear differentiation at the expression level, supporting the argument that expression changes contributed significantly to the species divergence. Using a genome-wide expression profiling approach, Martinez-Fernandez et al. (2010) found that the differentiation between two snail ecotypes was 7%, 4%, and 3% at the proteome, expression, and sequence levels, respectively, suggesting

that species divergence was largest at the phenotypic level and smallest at the sequence level.

In the present study, based on whole-genome resequencing data of 95 rice accessions, we found that more than 10% of the whole genome was significantly differentiated between upland and lowland rice (12.5% in *Indica* and 15.7% in *Japonica*) (Table 1), consistent with previous reports that detected substantial genetic differentiation between upland and lowland rice (Lyu et al., 2014; Xia et al., 2019). Such a high level of genome-wide divergence between upland and lowland rice at sequence level is of significance considering that upland rice evolved very recently, much less than 10 000 years ago the time when rice was domesticated (Khush, 1997; Sang & Ge, 2007). More strikingly, our genome-wide expression analyses detected approximately 30% of effectively expressed genes was significantly differentiated between upland and lowland rice (Table 2), much higher than the genome-wide expression differentiation between crops and their wild progenitors ranging from 3.3% to 14.1% reported previously (Lai et al., 2008; Swanson-Wagner et al., 2012; Sauvage et al., 2017). In our previous study on two wild rice species (*Oryza rufipogon* and *O. nivara*), we found that only 8% of effectively expression genes was significantly differentiated, although the two species diverged at least 0.4 million years (Zheng & Ge, 2010; Guo et al., 2016). These observations highlight the importance of regulatory evolution in domestication of upland rice and are consistent with previous arguments that the overall expression divergence might evolve faster than overall nucleotide divergence due to correlated effects that the expression change of one gene has on other genes (Wolf et al., 2010). This also supports the idea that gene expression differences were a sensitive indicator of initial species divergence and adaptation, because gene expression allowed for rapid phenotypic change without a long waiting time for new mutations and substitutions in coding regions (Nosil, 2012). It is noted that, although hundreds of genes (586 in *Indica* and 352 in *Japonica*) fell in the putative selected windows based on resequencing data (Table 1), and thousands of DEGs (1293 in *Indica* and 8131 in *Japonica*) were found based on RNA-seq data (Table 2), only 13 genes in *Indica* and 58 genes in *Japonica* were identified simultaneously by both datasets (Table S5), implying that the genetic variations at the sequence and expression levels are largely uncorrelated during domestication of upland rice.

Previous studies based on genomic scans on sequence variation and expression changes have revealed "genomic islands" (or hotspots) of sequence and expression divergence that are involved in adaptation and speciation (Nosil, 2012; Shibata et al., 2012). Similarly, we found that the outlier windows and the DEGs between upland and lowland rice were non-randomly distributed across the genome, suggesting that the phenotypic changes of upland rice are involved in many genomic regions or loci that formed clusters across the genome. These islands or hotspots need particular attentions in research and breeding practice of upland rice.

4.3 Parallel domestication of upland rice is largely non-parallel at the genetic level

Parallel evolution is widespread in animals and plants and has been investigated in great detail (see reviews in Nosil, 2012; Bolnick et al., 2018). Despite great efforts, the genetic basis and causative factors behind parallel phenotypes remain elusive. Crop domestication provides a good system to address this question because the phenotypic parallel of domestication syndrome occurred in diverse plant groups and is widely appreciated (Doebley et al., 2006; Gaut, 2015; Pickersgill, 2018; Woodhouse & Hufford, 2019). Accumulating studies on the parallelism of laboratory and natural systems have shown that parallel or replicated evolution could result in highly similar trajectories or might evolve in distinct directions, reflecting a quantitative continuum of parallelism ranging from parallel to non-parallel (Bolnick et al., 2018). Take stickleback studies as examples. Jones et al. (2012) detected 33% of outlier markers (SNPs) that were shared by two or more of the independent benthic–limnetic ecotype pairs of stickleback from Canada. In contrast, for European lake-stream sticklebacks, only 3% of outlier windows were shared by population pairs (Feulner et al., 2015). Many lines of evidence indicated that, during crop domestication, similar phenotypic changes in different species might be controlled by the same genes or different genes in the same pathway and even genetic variation from the genes in different pathways (Pickersgill, 2018; Woodhouse & Hufford, 2019). Moreover, based on a comparative genomic study on selected genes during domestication of crops, Gaut (2015) did not find evidence of parallel selection events either between distinctly related maize and rice or between two independent domesticates within bean species. Therefore, it was proposed that artificial selection for domestication might have involved largely non-parallel genomic changes (Bolnick et al., 2018).

Here we took advantage of two types of upland rice to explore the genetic basis of parallel domestication. We found a moderate level of parallelism of genetic pathways, as evidenced by our outlier window and DEGs analyses in which 10 of the top 20 enriched GO terms were shared by upland *Indica* and upland *Japonica* (Figs. S6, S8). In striking contrast, based on the analysis of resequencing data, we found that only one putative selected window and three genes within this window are common in upland *Indica* and upland *Japonica*, although many more outlier windows (682) and genes (1222) are shared by two groups of upland rice (Table 1). Similarly, our whole genome expression analysis detected a very small proportion of DEGs that was shared by two groups of upland rice, ranging from 1.2% (tissue S) to 2.3% (tissue LH); particularly, only 67 DEGs show the same direction of expression in both *Indica* and *Japonica* (Table 2). These observations suggest that domestication of upland rice might involve different strategies of adaptation or suffer different artificial selections in *Indica* and *Japonica* and genetic changes during domestication of upland rice do not seem parallel in *Indica* and *Japonica* at the gene level. Accumulating evidence has shown that parallel phenotypic traits can arise either from the same gene (even the same mutations) or from totally unrelated genes in different enzymatic pathways because many evolutionary forces can give rise to parallel evolution (Bolnick et al., 2018; Woodhouse & Hufford, 2019).

As pointed out by Conte et al. (2012), the probability of gene reuse in parallel phenotypic evolution declined with

increasing age of the common ancestor of compared taxa. Therefore, given that two types of upland rice were domesticated within roughly several thousand years and shared a very recent ancestor, the finding that non-parallelism at the gene level during upland rice domestication is of significance. A couple of reasons might be relevant. First, domestication of upland rice involved different progenitors (i.e., *Indica* and *Japonica*) that have distinct genome backgrounds (Sang & Ge, 2007; Lyu et al., 2014; Wang et al., 2014). Although upland *Indica* and upland *Japonica* shared a common feature of adaption to drought environments, differences in morphological characters, physiology, and cultivation cultures exist. In particular, the seemingly common feature of drought tolerance or resistance might have different genetic bases (Bernier et al., 2008; Lyu et al., 2014). This could partially explain why the shared genes that were potentially under selection or differentially expressed by upland and lowland rice were very few and why the pathways shared by the two types of upland rice were exclusively related to the adaptation and responses of plants to biotic and abiotic stress (Figs. S6, S8; Table S5). Second**,** because drought tolerance or resistance is a complex trait or phenotype involving a complex regulatory network with different pathways and genes (Bernier et al., 2008; Joshi et al., 2016; Zhu, 2016), parallelism in such a phenotype is most like to be caused by different genetic components. Research on other domestication traits related to complex developmental networks have proved to be controlled by different genetic routes (Glemin & Bataillon, 2009; Martinez-Ainsworth & Tenaillon, 2016). For example, the increased size in various crops might result from the changes of different genes in the same CLV3-WUS pathway or *CNR* genes (e.g., *FW2.2* in tomato and *ZmCNR1* in maize) (van der Knaap et al., 2014; Somssich et al., 2016) and the reduced branching in many cereals was caused by different routes (Guan et al., 2012; Lyu et al., 2020). Finally, because of different origins and domesticated culture, upland rice in the two subspecies might suffer different artificial selections and thus the same genes related to domestication and adaptation can behave or respond in different ways. Considering all these potential factors, further investigations are required to determine the exact genes associated with parallel alterations and their interaction in metabolic pathways during domestication of upland rice.

Acknowledgements

We would like to thank Yu-Su Du, Xun Xu, Qing-Lin Meng, and other members of Ge's laboratory for technical assistances and valuable discussions. We would also like to thank Xin-Hua Wei (China National Rice Research Institute) for providing seed samples and Yan-Qing Chen (Chinese Crop Germplasm Resources Information System) for sharing information on the drought-tolerant type of the samples. This work was supported by the National Natural Science Foundation of China (91731301) and grants from the Chinese Academy of Sciences (XDB31000000 and XDA08020103) to S.G. and the National Natural Science Foundation of China (31670224) to J.G.

References

Alexa A, Rahnenfuehrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22: 1600–1607.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655–1664.

Anders S, Pyl PT, Huber W. 2015. HTSeq — A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169.

Bao Y, Hu G, Flagel LE, Salmon A, Bezanilla M, Paterson AH, Wang Z, Wendel JF. 2011. Parallel up-regulation of the profilin gene family following independent domestication of diploid and allopolyploid cotton (*Gossypium*). *Proceedings of the National Academy of Sciences USA* 108: 21152–21157.

Bernier J, Atlin GN, Serraj R, Kumar A, Spaner D. 2008. Breeding upland rice for drought resistance. *Journal of the Science of Food and Agriculture* 88: 927–939.

Blanca J, Montero-Pau J, Sauvage C, Bauchet G, Illa E, Díez MJ, Francis D, Causse M, van der Knaap E, Cañizares J. 2015. Genomic variation in tomato, from wild ancestors to contemporary breeding accessions. *BMC Genomics* 16: 257.

Blount ZD, Lenski RE, Losos JB. 2018. Contingency and determinism in evolution: Replaying life's tape. *Science* 362: eaam5979.

Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. 2018. (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics* 49: 303–330.

Chang TT. 1976. The origin, evolution, cultivation, dissemination and diversification of Asian and African rices. *Euphytica* 5: 425–441.

Chen F, Hu Y, Vannozzi A, Wu K, Cai H, Qin Y, Mullis A, Lin Z, Zhang L. 2017. The WRKY transcription factor family in model plants and crops. *Critical Reviews in Plant Sciences* 36: 311–335.

Chen Y, Lun ATL, Smyth GK. 2014. Differential expression analysis of complex RNA-seq experiments using edgeR. In: Datta S, Nettleton D eds. *Statistical analysis of next generation sequencing data. Frontiers in probability and the statistical sciences*. Cham: Springer. 51–74.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The probabilityof genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* 279: 5039–5047.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127: 1309–1321.

Felsenstein J. 2005. *PHYLIP (Phylogeny Inference Package) version 3.6*. Seattle: Department of Genome Sciences, University of Washington.

Feulner PG, Chain FJJ, Panchal M, Huang Y, Eizaguirre C, Kalbe M, Lenz TL, Samonte IE, Stoll M, Bornberg-Bauer E, Reusch TBH, Milinski M. 2015. Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics* 11: e1004966.

Gaut BS. 2015. Evolution is an experiment: Assessing parallelism in crop domestication and experimental evolution. *Molecular Biology and Evolution* 32: 1661–1671.

Glemin S, Bataillon T. 2009. A comparative view of the evolution of grasses under domestication. *New Phytologist* 183: 273–290.

Guan JC, Koch KE, Suzuki M, Wu S, Latshaw S, Petruff T, Goulet C, Klee HJ, McCarty DR. 2012. Diverse roles of strigolactone

signaling in maize architecture and the uncoupling of a branching-specific subnetwork. *Plant Physiology* 160: 1303–1317.

Guo J, Liu R, Huang L, Zheng X-M, Liu P-L, Du Y-S, Cai Z, Zhou L, Wei X-H, Zhang F-M, Ge S. 2016. Widespread and adaptive alterations in genome-wide gene expression associated with ecological divergence of two *Oryza* species. *Molecular Biology and Evolution* 33: 62–78.

Huang XY, Chao DY, Gao JP, Zhu MZ, Shi M, Lin HX. 2009. A previously unknown zinc finger protein, *DST*, regulates drought and salt tolerance in rice via stomatal aperture control. *Genes & Development* 23: 1805–1817.

Jiang N, Visa S, Wu S, van der Knaap E. 2012. Rider transposon insertion and phenotypic change in tomato. In: Grandbastien MA, Casacuberta JM eds. *Plant transposable elements*. Heidelberg: Springer. 297–312.

Jones FC, Chan YF, Schmutz J, Grimwood J, Brady SD, Southwick AM, Absher DM, Myers RM, Reimchen TE, Deagle BE, Schluter D, Kingsley DM. 2012. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* 22: 83–90.

Joshi R, Wani SH, Singh B, Bohra A, Dar ZA, Lone AA, Pareek A, Singla-Pareek SL. 2016. Transcription factors and plants response to drought stress: Current understanding and future directions. *Frontiers in Plant Science* 7: 1029.

Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology* 35: 25–34.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology* 14: R36.

Kim S-K, You YN, Park JC, Joung Y, Kim B-G, Ahn JC, Cho HS. 2012. The rice thylakoid lumenal cyclophilin *OsCYP20-2* confers enhanced environmental stress tolerance in tobacco and *Arabidopsis*. *Plant Cell Reports* 31: 417–426.

Lai Z, Kane NC, Zou Y, Rieseberg LH. 2008. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. *Genetics* 179: 1881–1890.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li J, Han Y, Liu L, Chen Y, Du Y, Zhang J, Sun H, Zhao Q. 2015. *qRT9*, a quantitative trait locus controlling root thickness and root length in upland rice. *Journal of Experimental Botany* 66: 2723–2732.

Li J-Y, Wang J, Zeigler RS. 2014. The 3,000 rice genomes project: New opportunities and challenges for future rice research. *GigaScience* 3: 8.

Li ZC, Mu P, Li CP, Zhang HL, Li ZK, Gao YM, Wang XK. 2005. QTL mapping of root traits in a doubled haploid population from a cross between upland and lowland japonica rice in three environments. *Theoretical and Applied Genetics* 110: 1244–1252.

Lian H-L, Yu X, Lane D, Sun W-N, Tang Z-C, Su W-A. 2006. Upland rice and lowland rice exhibited different PIP expression under water deficit and ABA treatment. *Cell Research* 16: 651–660.

Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* 339: 1312–1316.

Lyu J, Huang L, Zhang S, Zhang Y, He W, Zeng P, Zeng Y, Huang G, Zhang J, Ning M, Bao Y, Zhao S, Fu Q, Wade LJ, Chen H, Wang W, Hu F. 2020. Neo-functionalization of a *Teosinte branched 1* homologue mediates adaptations of upland rice. *Nature Communications* 11: 725.

Lyu J, Li B, He W, Zhang S, Gou Z, Zhang J, Meng L, Li X, Tao D, Huang W, Hu F, Wang W. 2014. A genomic perspective on the important genetic mechanisms of upland adaptation of rice. *BMC Plant Biology* 14: 160.

Martinez-Ainsworth NE, Tenaillon MI. 2016. Superheroes and masterminds of plant domestication. *Comptes Rendus Biologies* 339: 268–273.

Martinez-Fernandez M, Bernatchez L, Rolan-Alvarez E, Quesada H. 2010. Insights into the role of differential gene expression on the ecological adaptation of the snail *Littorina saxatilis*. *BMC Evolutionary Biology* 10: 356.

Nosil P. 2012. *Ecological speciation*. Oxford: Oxford University Press.

Pickersgill B. 2018. Parallel vs. convergent evolution in domestication and diversification of crops in the Americas. *Frontiers in Ecology and Evolution* 6: 56.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.

Rabello AR, Guimaraes CM, Rangel PHN, da Silva FR, Seixas D, de Souza E, Brasileiro ACM, Spehar CR, Ferreira ME, Mehta A. 2008. Identification of drought-responsive genes in roots of upland rice (*Oryza sativa* L). *BMC Genomics* 9: 485.

Sang T, Ge S. 2007. Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics & Development* 17: 533–538.

Sauvage C, Rau A, Aichholz C, Chadoeuf J, Sarah G, Ruiz M, Santoni S, Causse M, David J, Glemin S. 2017. Domestication rewired gene expression and nucleotide diversity patterns in tomato. *Plant Journal* 91: 631–645.

Shibata Y, Sheffield NC, Fedrigo O, Babbitt CC, Wortham M, Tewari AK, London D, Song L, Lee B-K, Iyer VR, Parker SCJ, Margulies EH, Wray GA, Furey TS, Crawford GE. 2012. Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genetics* 8: e1002789.

Somssich M, Je BII, Simon R, Jackson D. 2016. *CLAVATA-WUSCHEL* signaling in the shoot meristem. *Development* 143: 3238–3248.

Stellari GM, Mazourek M, Jahn MM. 2010. Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity* 104: 460–471.

Swanson-Wagner R, Briskine R, Schaefer R, Hufford MB, Ross-Ibarra J, Myers CL, Tiffin P, Springer NM. 2012. Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences USA* 109: 11878–11883.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30: 2725–2729.

Uga Y, Okuno K, Yano M. 2011. *Dro1*, a major QTL involved in deep rooting of rice under upland field conditions. *Journal of Experimental Botany* 62: 2485–2494.

van der Knaap E, Chakrabarti M, Chu YH, Clevenger JP, Illa-Berenguer E, Huang Z, Keyhaninejad N, Mu Q, Sun L, Wang Y, Wu S. 2014. What lies beyond the eye: The molecular mechanisms regulating tomato fruit weight and shape. *Frontiers in Plant Science* 5: 227.

Walden N, Lucek K, Willi Y. 2020. Lineage-specific adaptation to climate involves flowering time in North American *Arabidopsis lyrata*. *Molecular Ecology* 29: 1436–1451.

Wang CH, Zheng XM, Xu Q, Yuan XP, Huang L, Zhou HF, Wei XH, Ge S. 2014. Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity* 112: 489–496.

Wang M, Li W, Fang C, Xu F, Liu Y, Wang Z, Yang R, Zhang M, Liu S, Lu S, Lin T, Tang J, Wang Y, Wang H, Lin H, Zhu B, Chen M,

Kong F, Liu B, Zeng D, Jackson SA, Chu C, Tian Z. 2018. Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics* 50: 1435–1441.

Wolf JBW, Bayer T, Haubold B, Schilhabel M, Rosenstiel P, Tautz D. 2010. Nucleotide divergence vs. gene expression differentiation: Comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. *Molecular Ecology* 19: 162–175.

Woodhouse MR, Hufford MB. 2019. Parallelism and convergence in post-domestication adaptation in cereal grasses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374: 20180245.

Xia H, Luo Z, Xiong J, Ma X, Lou Q, Wei H, Qiu J, Yang H, Liu G, Fan L, Chen L, Luo L. 2019. Bi-directional selection in upland rice leads to its adaptive differentiation from lowland rice in drought resistance and productivity. *Molecular Plant* 12: 170–184.

Xia H, Zheng X, Chen L, Gao H, Yang H, Long P, Rong J, Lu B, Li J, Luo L. 2014. Genetic differentiation revealed by selective loci of drought-responding EST-SSRs between upland and lowland rice in China. *PLoS One* 9: e106352.

Yang L, Wang H-N, Hou X-H, Zou Y-P, Han T-S, Niu X-M, Zhang J, Zhao Z, Todesco M, Balasubramanian S, Guo Y-L. 2018. Parallel evolution of common allelic variants confers flowering diversity in *Capsella rubella*. *The Plant Cell* 30: 1322–1336.

Zhang X, Lin Z, Wang J, Liu H, Zhou L, Zhong S, Li Y, Zhu C, Liu J, Lin Z. 2019. The *tin1* gene retains the function of promoting tillering in maize. *Nature Communications* 10: 5608.

Zhao Y, Zhang H, Xu J, Jiang C, Yin Z, Xiong H, Xie J, Wang X, Zhu X, Li Y, Zhao W, Rashid MAR, Li J, Wang W, Fu B, Ye G, Guo Y, Hu Z, Li Z, Li Z. 2018. Loci and natural alleles underlying robust roots and adaptive domestication of upland ecotype rice in aerobic conditions. *PLoS Genetics* 14: e1007521.

Zheng XM, Ge S. 2010. Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Molecular Ecology* 19: 2439–2454.

Zhu JK. 2016. Abiotic stress signaling and responses in plants. *Cell* 167: 313–324.

Supplementary Material

The following supplementary material is available online for this article at http://onlinelibrary.wiley.com/doi/10.1111/jse.12636/suppinfo:

Table S1. Information on 95 rice accessions used in this study.
Table S2. Statistics of resequencing data.
Table S3. Statistics of RNA sequencing data.
Table S4. Information of the nine known function genes identified in previous studies.
Table S5. Functions of the intersections between putative selected genes and DEGs.
Fig. S1. Genetic diversity of upland and lowland rice in two rice subspecies *(Indica and Japonica)*. Sample sizes are in parentheses.
Fig. S2. Matrix of pairwise F_{ST} values of upland and lowland rice from two rice subspecies. The lower left and upper right are the values calculated based on the neutral and total SNPs, respectively.
Fig. S3. Analyses of population genetic structure of all 95 rice accessions based on total SNPs of resequencing data. (**A**) Neighbor-joining (NJ) tree. (**B**) Principal components analysis (PCA). (**C**) Model-based population assignments at K from 2 to 6. Each vertical bar represents a sample, with its assignment probability to genetic clusters represented by different colors.
Fig. S4. NJ trees based on expression quantity of the effectively expressed genes in five tissues, i.e., leaves at the seedling stage (S), flag leaves (LH) and panicles (PH) at the heading stage, flag leaves (LM) and panicles (PM) at the milk stage.
Fig. S5. Distribution patterns of the outlier windows on chromosomes. (**A**) Physical positions of the outlier windows on 12 chromosomes. Each dot represents a 10 kb window. Alternating colors represent different chromosomes. (**B**) Proportion of the outlier windows to the total number of windows in each chromosome.
Fig. S6. Gene ontology (GO) enrichment of the genes in the outlier windows. Of the 20 top functional terms, ten (shaded) were shared between upland *Indica* (left) and upland *Japonica* (right). Numbers of genes for each term were to the right of the bars.
Fig. S7. The distribution patterns of DEGs between upland and lowland rice in *Indica* rice (top) and *Japonica* rice (bottom) across the genome. Each dot represents a 200-gene window. Alternating blue and yellow colors indicate different chromosomes.
Fig. S8. Gene ontology (GO) enrichment of the DEGs in two subspecies. Of the 20 top terms, ten (shaded) were shared by upland *Indica* (left) and upland *Japonica* (right). Numbers of genes for each term were to the right of the bars. Red stars indicate the terms that are common to those identified by the outlier window approach (Fig. S6).