Brief Communication

Genomic signatures of domestication and adaptation during geographical expansions of rice cultivation

Xiaoming Zheng1,*, Hongbo Pang1,†, Junrui Wang1,3, Xuefeng Yao4,5,†, Yue Song1, Fei Li1,6, Danjing Lou1, Jinyue Ge1, Zongyao Zhao2, Weihua Qiao1,†, Sung Ryul Kim6,7, Guoyou Ye7, Kenneth M. Olsen8,*,†, ChunMing Liu1,4,5,† and Qingwen Yang1,*,†

1National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China
2College of Life Science, Shenyang Normal University, Shenyang, China
3Guangxi Key Laboratory for Polysaccharide Materials and Modifications, School of Marine Sciences and Biotechnology, Guangxi University for Nationalities, Nanning, China
4Key Laboratory of Plant Molecular Physiology, Institute of Botany, The Chinese Academy of Sciences, Beijing, China
5The University of Chinese Academy of Sciences, Beijing, China
6Crop Biotech Institute & Department of Genetic Engineering, Kyung Hee University, Yongin, Korea
7International Rice Research Institute, Metro Manila, The Philippines
8Biology Department, Washington University, St. Louis, MO, USA

Keywords: rice, adaptation, association, cold tolerance, breeding.

Rice (Oryza sativa L.), a global staple food now grown on all inhabited continents, was domesticated from its wild progenitor, O. rufipogon Griff., in tropical and subtropical regions of Asia (Oka, 1988). After domestication, the expansions of rice landraces into the present-day range required a diverse array of adaptations to local environments, which included changes in daylight sensitivity, expanded thermal tolerance (for excess cold and heat), adaptations to water availability (drought and waterlogging), and resistance to biotic stresses (Garris et al., 2005; Glaszmann, 1987). Although a large amount of genomic data has been available for wild and cultivated rice varieties, and genetic characterizations of important agronomic traits were obtained in the past two decades (Gutaker et al., 2020; Huang et al., 2011, 2012; Wang et al., 2016), a complete landscape of genomic variations underlying regional adaptations remains elusive.

We selected 185 wild rice (O. rufipogon) and 743 cultivated rice varieties, which represent 33 major rice-growing regions worldwide (Figure 1a). Of them, all 185 wild rice varieties and 371 cultivated rice varieties (203 japonica and 168 indica varieties) were newly sequenced (see methods and material on https://github.com/nya-caas/zheng). The sequencing depth for these varieties was ≥5, which is for the first time to sequence such a large amount of wild rice varieties at this sequencing depth. Through a combination of principal component analysis (PCA), we confirmed five cultivated rice subgroups, including aus, indica (ind), tropical japonica (trj), temperate japonica (tej), and aromatic (aro) (Figure 1b). In particular, the aus and ind varieties, cultivated primarily in tropical and subtropical regions of Asia, were derived from the wild rice clade designated as OR-I. The trj and tej varieties, both originated from the OR-I clade of wild rice, were cultivated primarily in tropical, high-elevation regions of Southeast Asia (trj) and colder regions of Northeast Asia (tej).

To understand the selection on critical traits and the recent rapid speciation, three robust indicators of selective sweeps were examined to identify top outliers: (i) maximum differentiation from wild rice; (ii) maximum negative residuals of diversity/differentiation metrics based on the Hudson–Kreitman–Agadze test; and (iii) top scoring windows from a site frequency spectrum composite likelihood ratio test based on the length of haplotype. Using high-stringency cut-offs, considerable overlaps in outliers were observed among these methods. We identified 131 genomic regions with strong selection signals. Within individual rice subgroups, the numbers of targeted regions specific to tej, trj, ind, and aus were 20, 11, 20, and 12, respectively, including 791, 1477, 2647, and 2786 genes, respectively. Some of them were strong candidates for selection imposed by human preferences and local environmental conditions.

To identify loci associated with local adaptation, we collected temperature and day-length information in May and August of the last 68 years across the sampling sites in the Figure 1a and performed a genotype-environment association analysis based on a Bayesian analysis based on allele frequency data (Coop et al., 2010). As shown in Manhattan plots (Figure 1c,d), 128 significant association signals/loci (defined by −log10p > 3) were identified. Among these significant association signals/loci, 25 were located in genic regions (exonic or intronic), including eight for day length ($OsFL11, RNC3, OsDoF2, APG, RFT1, Hdt1, Ghd7$, and $ROC4$; Figure 1c) and one for temperature sensitivity ($COLD1$; Figure 1c). The remaining 103 loci were located in non-genic regions, more than half of which (59, or 57.28%) coincided with top selective sweep candidate regions detected in the four cultivated rice subgroups, including 32 specific to tej, 16 to trj, 8 to aus, and 3 to ind. These were strong candidates as targets of selection for local adaptation.

Based on the adaptation-associated results, the top five significant sites were found on chromosome (Chr.) 8 (Figure 1d). We only identified a non-synonymous SNP (C/A) with a significant
Genomic signatures of rice adaptation

(a) Map showing genomic signatures of rice adaptation.
(b) Scatter plot illustrating the distribution of various genomic signatures.
(c) Chromosome map with highlighted regions.
(d) Graph depicting log2(p) values.
(e) Schematic of gene expression analysis.
(f) Survival rates comparison.
(g) Percentage survival rates.
(h) Frequency distribution of gene expression.
(i) Phenotypic analysis of rice plants.
(j) Phenotypic analysis of rice plants with specific deletions.
(k) Survival rates across different haplotypes.

association peak ($P = 1.31 \times 10^{-22}$). This SNP was located in the coding region of LOC_Os08g360000 on Chr. 8 (+22 689 418 bp, C > A; +40 aa, Val > Phe) (Figure 1e). The encoded protein contained an F-box and two LRR domains. Many genes in this family respond to biotic and abiotic stresses, including temperature and hormones (Yan et al., 2010). The ‘C’ haplotype (Hap$^{69\text{C}}$) was almost fixed in the tej subgroup, with a frequency of 0.985, whereas the ‘A’ haplotype (Hap$^{69\text{A}}$) predominated in other subgroups included aus, ind, trj, and wild rice varieties, with a frequency of 0.961 (Figure 1f,h). To address the function of the Hap$^{69\text{A}}$ allele, we performed cold treatments on 14 cultivated rice varieties with the Hap$^{69\text{A}}$ allele and 19 cultivated rice varieties with the Hap$^{69\text{C}}$ allele in a growth chamber, and the results showed that under cold stress (4°C for 48 h), the survival rate was higher in rice varieties with the Hap$^{69\text{A}}$ allele (93.25%) than those with the Hap$^{69\text{C}}$ allele (10.12%; $P = 1.66 \times 10^{-15}$, two-tailed t-test) (Figure 1g,h,k). We therefore named the gene CHILLING-TOLERANCE DIVERGENCE F-box (COLDF).

To further characterize the function of COLDF, we obtained TILLING and CRISPR mutants of the gene in the ZH11 background (O. sativa ssp. japonica var. Zhonghua 11) (Figure 1i,j). The TILLING mutant carried a G to A mutation in the coding region (+22 688 780 bp, G > A; +256 aa, Ala > Thr), and the knockout mutant carried a 22-bp insertion in the coding region, designated coldf-1 and coldf-2, respectively. Fourteen-day-old seedlings of the homozygous coldf-1 and coldf-2 mutants were subjected to cold stress at 4°C for 48 h, and then returned to normal growth conditions at 30°C for recovery. Results showed that these coldf seedlings were susceptible to cold stress, with a survival rate of 37% and 12% in coldf-1 and coldf-2, respectively, in comparison with a survival rate of 82% in ZH11 (Figure 1i); $P = 1.07 \times 10^{-10}$, two-tailed t-test for coldf-1 and $P = 1.94 \times 10^{-12}$, two-tailed t-test for coldf-2). These results suggest that the Hap$^{69\text{C}}$ of COLDF is responsible for enhanced cold tolerance in the tej cultivars. Our findings provide insight into the mechanistic basis of rice domestication and improvement by identification of adaptation-associated loci, which would help in mitigating these effects by facilitating rapid development of improved varieties that are better able to tolerate the stresses accompanying the changing climate.

### Acknowledgements
This work was supported by the National Natural Science Foundation of China (31970237), and the Hundred Talent Program of Shenyang Normal University (SSDBRJH2002012).

### Conflict of interest
The authors declare no conflict of interest.

### Author contributions

### References


