






Brief Communication

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

Xiaoming Zheng^{1,†}, Hongbo Pang^{2,†}, Junrui Wang^{1,3}, Xuefeng Yao^{4,5} , Yue Song¹, Fei Li¹ , Danjing Lou¹, Jinyue Ge¹, Zongyao Zhao², Weihua Qiao¹ , Sung Ryul Kim^{6,7}, Guoyou Ye⁷, Kenneth M. Olsen^{8,*} , ChunMing Liu^{1,4,5,*} and Qingwen Yang^{1,*} 

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

Received 18 September 2021;

revised 5 October 2021;

accepted 7 October 2021.

*Correspondence (Tel 010-62189165; fax 010-62189165; email:

yangqingwen@caas.cn (QY); Tel 010-63836663; fax 010-62767138; email:

liuchunming@caas.cn (CL); Tel 1-314-935-7013; fax 1-314-935-4432; email:

kolsen@wustl.edu (KMO)

[†]These authors contribute equally to this work.

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/vya-caas/zheng>). The sequencing depth for these varieties was $>5\times$, 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-J 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–Aguade 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 $-\log_{10}P > 3$) were identified. Among these significant association signals/loci, 25 were located in genic regions (exonic or intronic), including eight for day length (*OsLFL1*, *RNC3*, *OsDof2*, *APG*, *RFT1*, *Hd1*, *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

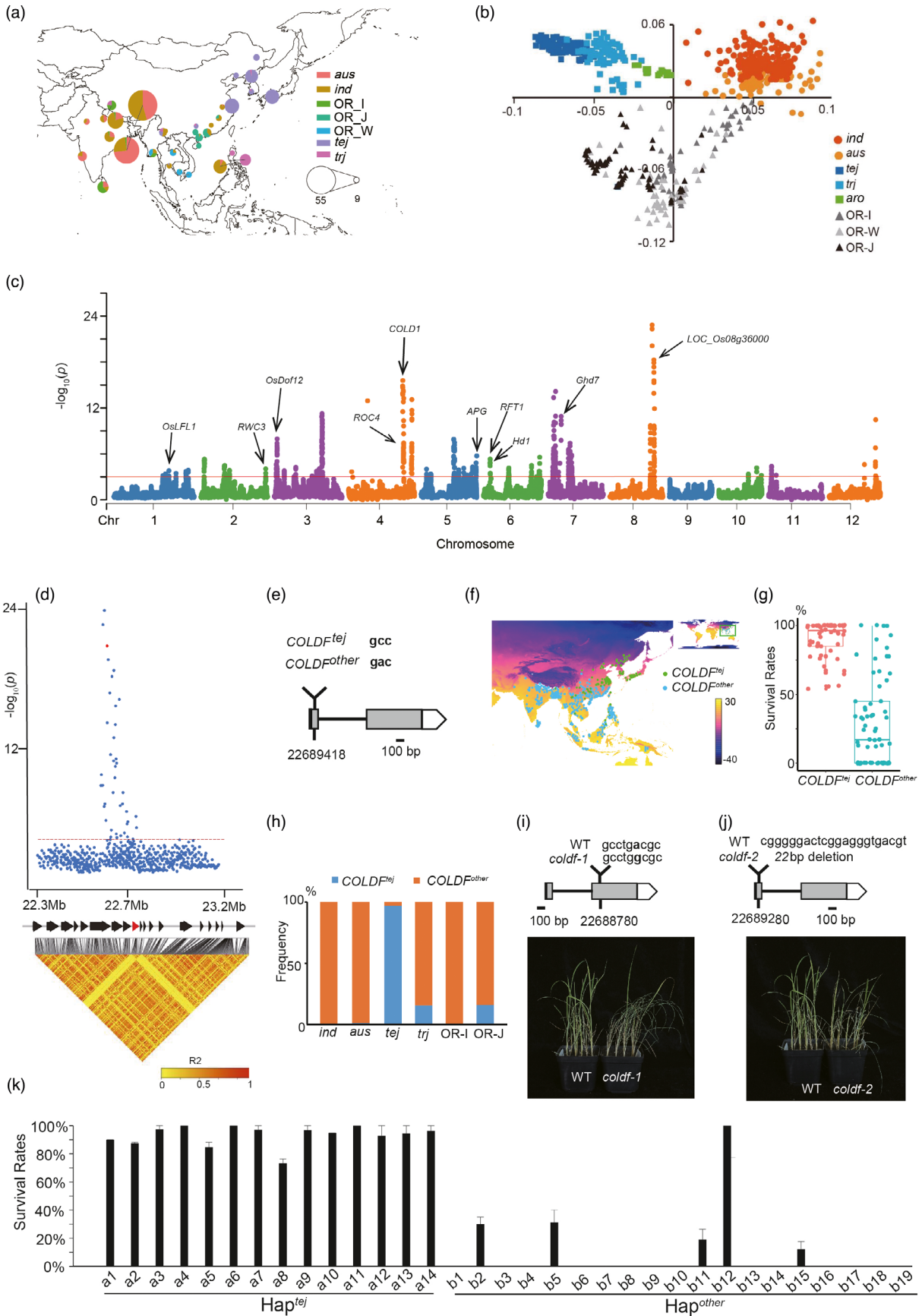


Figure 1 Molecular evolution of genomic regions associated with rice adaptation. (a) Sampling locations and frequencies of wild and cultivated rice varieties. (b) PCA plots of rice varieties. (c) Regional Manhattan plots of geographical locations and adaptation-association for temperature and day length. The red horizontal line indicates the empirical probability (defined by $-\log_{10}P > 3$) of the top 300 for both absolute ρ and Bayes factor across the three chains. (d) Local Manhattan plots, gene positions and LD heatmaps show the regions surrounding the strong peaks of the candidate genes (*COLDF*). (e) Significantly associated SNPs in the genomic location of *COLDF*. (f) Geographical distribution of rice varieties with *COLDF*^{tej} and *COLDF*^{other}. The colour shading scale indicates minimum temperature in May. (g) Chilling tolerance response of rice varieties with the two *COLDF* alleles. (h) The frequency of the two *COLDF* alleles in wild and cultivated rice subgroups. (i–j) Mutations in the *COLDF* genomic region for TILLING (*coldf-1*; i) and gene editing (*coldf-2*; j) mutants and showed increased chilling sensitivity. (k) The cold tolerance response and survival rate of 14 varieties with *COLDF*^{tej} and 19 accessions with *COLDF*^{other}.

association peak ($P = 1.31 \times 10^{-22}$). This SNP was located in the coding region of *LOC_Os08g36000* 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^{tej}) was almost fixed in the *tej* subgroup, with a frequency of 0.985, whereas the 'A' haplotype (Hap^{other}) 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^{tej} allele, we performed cold treatments on 14 cultivated rice varieties with the Hap^{tej} allele and 19 cultivated rice varieties with the Hap^{other} 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^{tej} allele (93.25%) than those with the Hap^{other} allele (10.12%; $P = 1.66 \times 10^{-13}$, 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 as *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,j; $P = 1.07 \times 10^{-10}$, 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^{tej} 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

Z.X.M., Y.Q.W., L.C.M., and K.M.O. designed experiments and wrote manuscript. Z.X.M., P.H.B., W.J.R, Y.X.F., S.Y., L.F., L.D.J., G.J.Y., J.M., and Q.W.H. analysed data.

References

- Coop, G., Witonsky, D., Di Rienzo, A. and Pritchard, J.K. (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.
- Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S. and McCouch, S. (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics*, **169**, 1631.
- Glazmann, J.C. (1987) Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* **74**, 21–30.
- Gutaker, R.M., Groen, S.C., Bellis, E.S., Choi, J.Y., Pires, I.S., Bocinsky, R.K., Slayton, E.R. *et al.* (2020) Genomic history and ecology of the geographic spread of rice. *Nat. Plants*, **6**, 492–502.
- Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y. *et al.* (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497–501.
- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W. *et al.* (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44**, 32–39.
- Oka, H.I. (1988) Emery's elements of medical genetics. In *Origin of Cultivated Rice*, pp. 254. Tokyo: Japan Sci. Soc. Press.
- Wang, H., Xu, X., Vieira, F., Xiao, Y., Li, Z., Wang, J., Nielsen, R. *et al.* (2016) The power of inbreeding: NGS-based GWAS of rice reveals convergent evolution during rice domestication. *Mol. Plant*, **9**, 975–985.
- Yan, Y., Chen, X., Yang, K., Sun, Z., Fu, Y., Zhang, Y. and Fang, R. (2010) Overexpression of an F-box protein gene reduces abiotic stress tolerance and promotes root growth in rice. *Mol. Plant*, **4**, 190–197.