Plant Communications
Correspondence

CellPress
Partner Journal

Landrace introgression contributed to the recent feralization of weedy rice in East China

Dear Editor,

Weedy rice (*Oryza* spp.) is a problematic paddy weed known for traits such as high seed shattering, persistent seed dormancy in the soil seed bank, and robust competitiveness against cultivated varieties, threatening rice production yields worldwide (Nadir et al., 2017). Recent population genomics studies have shown that weedy rice strains around the world have repeatedly and independently evolved from various cultivated ancestors at different times in the history of rice cultivation (Li et al., 2017; Qiu et al., 2017; Sun et al., 2019). This de-domestication (feralization) process involves phenotypic reversions from key domestication traits, such as non-shattering and non-dormant seeds, to more wild-like characteristics. Diverse genetic mechanisms appear to underlie the convergent evolution of weedy rice worldwide (Qi et al., 2015; Qiu et al., 2020). While most global weedy rice directly descends from domesticated rice, introgression from wild rice also contributes to its adaptive evolution, particularly in tropical Asia where wild *Oryzas* occur (Li et al., 2024a).

Recent genome analysis of weedy rice in East Asia has revealed that many current weed strains in this region descended from a handful of Green Revolution (GR) cultivars widely introduced in the late 20th century (Qiu et al., 2020). Despite their close kinship with modern cultivars, these weeds display typical weed-adaptive traits, such as red-pericarp seeds (associated with seed dormancy) and seed shattering, which are absent in modern cultivars. This raises a critical question of how these cultivar-derived weeds could have rapidly acquired their wild-like weedy traits. A detailed examination of potential additional contributors to weedy rice genomes beyond GR cultivars could shed light on the origin and adaptive evolution of these aggressive 21st-century agricultural weeds.

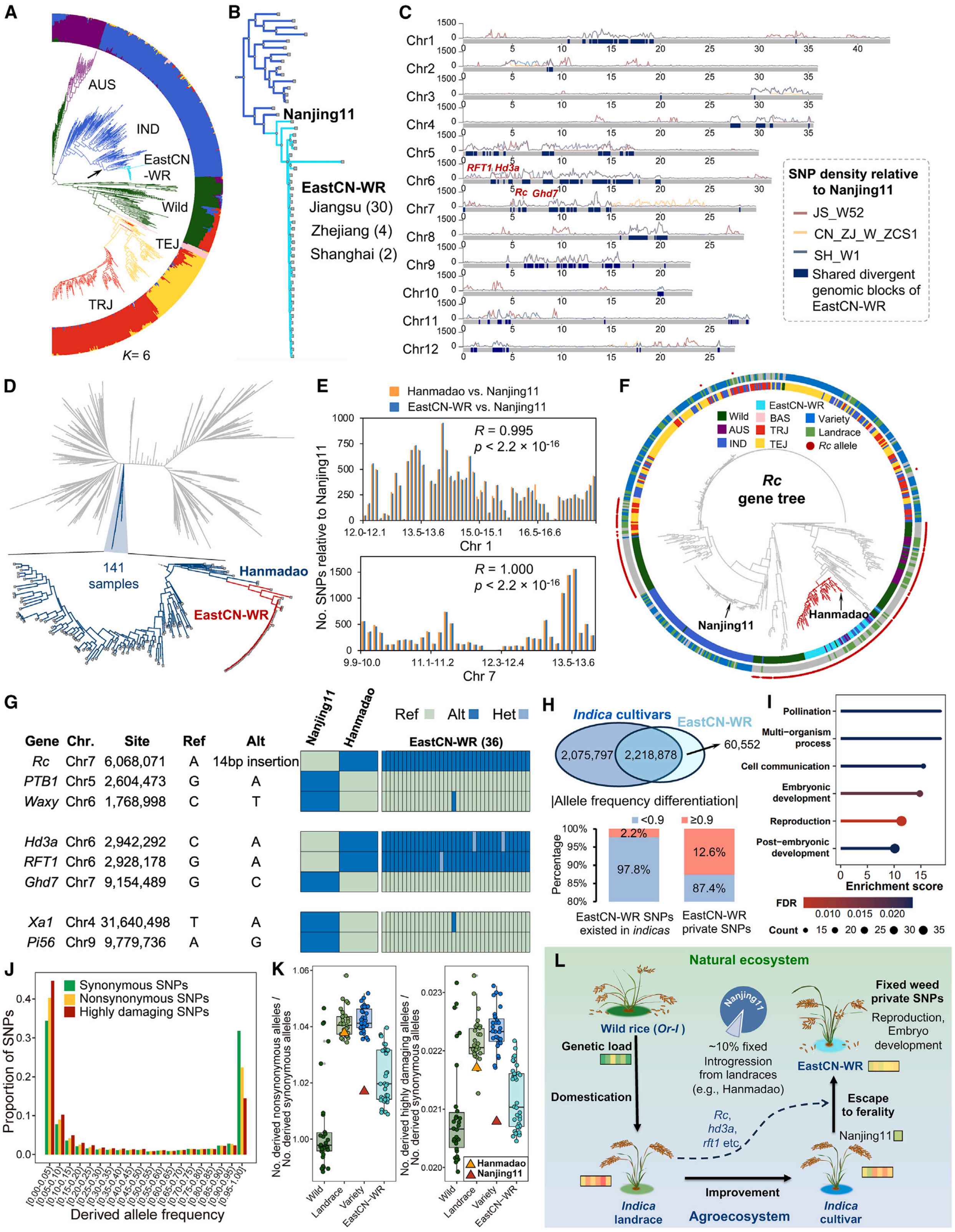
Our previous work revealed that weedy rice originating from Jiangsu and Zhejiang provinces in East China shows the closest genetic relationship with the *indica* GR cultivar "Nanjing11" (Qiu et al., 2020). By analyzing 26 previously sequenced weedy rice samples (22 from Jiangsu and 4 from Zhejiang), and newly sequenced genomes of 10 weedy rice samples (8 from Jiangsu and 2 from Shanghai), we investigated the evolutionary relationship between "Nanjing11" and these weedy rice strains from East China (collectively referred to as EastCN-WR). To create a more comprehensive dataset for evolutionary analysis, publicly available genome sequences of an additional 120 Asian wild rice and 442 Asian cultivated rice accessions were also included (Supplemental Table 1). Ultimately, we obtained a genotypic matrix with 598 samples and 6.87 million SNPs (missing rate < 0.2 and minor allele frequency > 0.01) for further analysis.

Principal-component analysis (Supplemental Figure 1) revealed that the first principal component, accounting for 14.57% of the total variation, mainly distinguishes between the *indica* and *japonica* subspecies, while the second principal component, representing 9.39% of the variation, primarily differentiates a subset of wild rice from all other samples. The 36 Chinese weedy rice accessions (EastCN-WR) clustered tightly within the *indica* group, and compared to other *indica* rice, they appear to be more distantly differentiated from both the *japonica* and wild groups (Supplemental Figure 1). Consistent with this pattern, admixture analysis indicated that the *indica* genetic component dominates in the EastCN-WR genomes (Figure 1A; Supplemental Figure 2). A maximum likelihood phylogenetic tree of 598 rice samples further demonstrated that all samples in the EastCN-WR group cluster on the same independent evolutionary branch (Figure 1B). Furthermore, in line with previous research, all of these weedy rice samples are most closely related to the GR cultivar Nanjing11, suggesting that Nanjing11 is the common progenitor of all sampled EastCN-WR accessions and the primary contributor to their genomic composition.

Identification of Nanjing11 as a common parent of EastCN-WR provides a crucial foundation for studying the genomic mechanisms of feralization in these recently evolved weeds. To examine the weed genomes at a finer scale, we performed 100 kb sliding window analyses to identify regions of genetic differentiation relative to the Nanjing11 genome. Although there were very few SNP differences from Nanjing11 across most of the genome, consistent with direct descent from the GR cultivar, we observed several Mb-level continuous differentiation blocks on multiple chromosomes (Figure 1C; Supplemental Figure 3). By assigning genomic blocks of each EastCN-WR accession as either Nanjing11 derived or divergent, we found that an average of 20.3% of the genome regions exhibit differentiation (Supplemental Figure 4). Additionally, most individuals (30/36) exhibit an overall consistent genomic differentiation pattern (Pearson $R > 0.95$) (Supplemental Table 2; Supplemental Figure 5). Of particular interest, a total of 11.2% of high-differentiation genome regions are shared and fixed in all 36 EastCN-WR accessions, suggesting that genes within these regions could be especially important for weed adaptation (Figure 1C; Supplemental Figure 4).

Due to the Mb-level genomic differentiation observed between each weedy rice accession and Nanjing11 within the EastCN-WR population, we hypothesized that these divergent segments could result from genomic introgression from other rice materials. To identify potential parental donors of these regions, we constructed a maximum likelihood evolutionary tree using only the

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.



Gene	Chr.	Site	Ref	Alt
Rc	Chr7	6,068,071	A	14bp insertion
PTB1	Chr5	2,604,473	G	A
Waxy	Chr6	1,768,998	C	T
Hd3a	Chr6	2,942,292	C	A
RFT1	Chr6	2,928,178	G	A
Ghd7	Chr7	9,154,489	G	C
Xa1	Chr4	31,640,498	T	A
Pi56	Chr9	9,779,736	A	G

Figure 1. Origin and adaptive evolution of weedy rice from East China.
(A) Phylogenetic tree and genetic component analysis of the 598 Asian rice accessions. Maximum likelihood clustering with $K = 6$ is presented.
(B) Fine-scale phylogenetic resolution of the relationship between EastCN-WR and Nanjing11.

(legend continued on next page)

SNPs that were highly differentiated between EastCN-WR and Nanjing11 (allele frequency change > 0.9). Notably, in the resulting tree, the 36 EastCN-WR samples uniformly grouped on a branch closest to the Chinese *indica* landraces "Hanmadao" and "Liushizao" (Supplemental Figure 6); strongly suggesting that genomic regions divergent from Nanjing11 originated from Chinese landraces that predate the widespread introduction of GR cultivars. This pattern was further confirmed by a comprehensive phylogenetic tree combining the 598 samples from the present study with 4591 previously sequenced rice samples (Wang et al., 2018). The resulting tree indicated that the Hanmadao landrace has the closest kinship with EastCN-WR in the Nanjing11-divergent regions (Figure 1D). A closer examination of specific putatively introgressed genomic regions provided additional evidence of their landrace origin; for example, regions spanning 12.0–18.0 Mb on chromosome 1 and 9.9–13.9 Mb on chromosome 7 show striking similarities to Hanmadao in the density distributions of SNP differentiation against Nanjing11 (Figure 1E). Together, these patterns strongly suggest that introgression from Chinese landraces has contributed to the evolution of EastCN-WR.

To further characterize landrace-derived introgression blocks across EastCN-WR genomes, we first employed *Loter* local ancestry scans to identify genomic regions of cultivar versus landrace origin; this confirmed landrace introgression signatures on a genome-wide scale (Supplemental Figure 7). Next, we used the *K*-mer variation approach "IBSpy" to scan complete genome assemblies of two EastCN-WR genomes, "JS_YC_W7" from Jiangsu and "CN_ZJ_W_CX20" from Zhejiang. The genome of JS_YC_W7 was newly sequenced and assembled for this study, yielding a genome size of 398.0 Mb with a contig N50 of 31.9 Mb (Supplemental Table 3). Using Nanjing11 as the reference, we observed a consistent pattern of landrace introgression, which supports our key inferences. Specifically, 36.8 Mb (9.2%) of the JS_YC_W7 genome and 36.9 Mb (9.3%) of the CN_ZJ_W_CX20 genome are closely related to landrace haplotypes (Supplemental Figure 8).

To examine the role of landrace introgression in the evolution of weedy traits in EastCN-WR, we focused on introgression regions containing genes that control weed-associated traits. One of the most characteristic features of weedy rice worldwide is the red pericarp, pigmented by proanthocyanidin and associated with seed dormancy, which is controlled by the functional alleles of the *Rc* gene. Most modern rice cultivars, including Nanjing11, carry a 14 bp deletion in the *Rc gene, which results in* white pericarp and reduced seed dormancy. By contrast, some unimproved landraces still carry functional *Rc* alleles. By generating a phylogenetic tree based on 186 SNPs within the *Rc* gene and its surrounding 2 kb regions, we found that the *Rc* gene in EastCN-WR diverges from the Nanjing11 *Rc* haplotype and clusters with landraces, including Hanmadao, on a distinct branch (Figure 1F). Both Hanmadao and EastCN-WR possess the functional *Rc* allele, whereas Nanjing11 does not (Figure 1G), this pattern suggests that adaptive introgression of the dormancy-conferring *Rc* allele from Hanmadao or related red-pericarp landraces has contributed to the emergence of EastCN-WR. Similarly, the Hanmadao landrace may also explain allelic variations in several other genes associated with weediness that differ between EastCN-WR and Nanjing11 (Figure 1G). These genes include those controlling the heading date (*Hd3a*, *RFT1*, *Ghd7*), biotic stress tolerance (*Xa1*, *Pi56*), seed setting rate (*PTB1*), and amylose content (*Waxy*).

Unlike the *Rc* gene and other genes where alleles differ between rice landraces and modern improved cultivars, major domestication genes, like *sh4,which* control reduced seed shattering, were targets of early selection in rice domestication. Consequently, reduced-shattering alleles are typically present in both landraces and improved cultivars, as well as their weedy descendants. Previous research in weedy rice has demonstrated that the

---

(C) SNP density of three weedy rice samples from East China compared to the Nanjing11 cultivar. The SNP densities for weedy rice from Jiangsu, Zhejiang, and Shanghai are plotted with red, orange, and blue lines, respectively. Common divergent genomic regions in 36 EastCN-WR samples compared to Nanjing11 are highlighted in dark blue rectangles along the chromosomes.

(D) Phylogenetic tree inferring the origin of the divergent genomic blocks in EastCN-WR. Top: Phylogenetic tree including 5189 accessions (598 from this study, 4591 from Wang et al. (2018)) constructed based on SNPs in common divergent genomic regions of EastCN-WR. Bottom: A zoomed-in phylogenetic relationship for a monophyletic clade, including 141 accessions. The EastCN-WR clade is highlighted in red.

(E) Two genomic regions showing similar divergence patterns between Hanmadao and EastCN-WR compared to Nanjing11. The *P* values are obtained using the Pearson correlation test.

(F) Phylogenetic tree of the *Rc* gene constructed based on SNPs in the *Rc* gene and the 2 kb regions upstream and downstream. A monophyletic clade encompassing cultivated and weedy rice is highlighted in red. The evolutionary positions of Nanjing11 and Hanmadao are indicated with arrows. The circles surrounding the tree, from inside to outside, indicate different rice types (landrace and cultivar). Samples with a functional *Rc* allele are indicated with red dots.

(G) Differentiated quantitative trait nucleotides between EastCN-WR and Hanmadao. The divergence of alleles can be explained by Hanmadao, which complements Nanjing11. Light green, blue, light blue, and gray boxes represent the genotypes for the Nipponbare reference (Ref) allele, alternative (Alt) allele, heterozygous (Het.), and missing data, respectively.

(H) Higher allele frequency differentiation of SNPs private to EastCN-WR compared to SNPs found in the cultivated *indica* group.

(I) GO enrichment of genes with high frequency (>0.9) EastCN-WR private SNPs. GO terms with *P* values (hypergeometric) < 0.0001 and false discovery rates (Yekutieli) < 0.05 are shown.

(J) Comparison of site allele frequency spectra of different SNP types in the rice genome.

(K) Comparison of the genetic load of wild rice, *indica* landraces, *indica* varieties, and EastCN-WR. Left: based on the ratio of non-synonymous to synonymous alleles. Right: based on the ratio of highly damaging to synonymous alleles.

(L) Schematic illustration of the origin and evolution of EastCN-WR. EastCN-WR originated from the Green Revolution cultivar Nanjing11, which has a with low genetic load. During its escape to ferality, landrace introgressions occurred, with around 10% being fixed in the EastCN-WR genomes. These introgressions introduced weed-adaptive alleles at genes including *Rc*, *hd3a*, and *rft1*. Private SNPs at high frequency within the EastCN-WR population are predominantly associated with genes potentially controlling reproduction and embryonic development. The gradient-colored rectangles illustrate genetic load levels, with green indicating low loads, yellow indicating intermediate levels, and red indicating high loads.

re-emergence of seed shattering in weedy rice, despite the presence of reduced-shattering alleles, is often polygenic and involves diverse genetic and developmental mechanisms across independently evolved weed strains worldwide (Qi et al., 2015; Li et al., 2024b). Consistent with these findings, Our study revealed that quantitative trait nucleotides for seed-shattering genes, namely *sh4*, *qsh3*, and *sh5*, showed no differences between EastCN-WR and either Nanjing11 or Hanmadao (Supplemental Figure 9).

In addition to known genes for weed-associated traits, we investigated the potential contributions of SNPs with unknown functions that were differentially enriched in the genomes of EastCN-WR compared to domesticated rice. A total of 60 552 SNPs were found to be unique to EastCN-WR in a comparison to all 190 *indica* samples included in this study. Notably, we found that approximately one-eighth (12.6%) of these EastCN-WR-specific SNPs were present at differentially high frequencies (defined as ≥ 0.9) within the weed population. This proportion of high-frequency private SNPs in EastCN-WR is far greater than the proportion (2.2%) of EastCN-WR SNPs also found in cultivated *indicas* (Figure 1H). Further analysis revealed that the high-frequency private SNPs in EastCN-WR are located within 471 genes (Supplemental Table 4). Gene Ontology (GO) enrichment analysis revealed a significant overrepresentation of functions related to "reproduction" (GO: 0000003) and "embryonic development" (GO: 0009790) (Figure 1I). This pattern suggests that adaptations related to reproductive strategies may play a crucial role in the feralization process.

The dynamics of genetic load during crop domestication have been well-documented; however, there are limited reports on genetic load dynamics during crop feralization. Given that EastCN-WR belongs to the *indica* subspecies, our analysis focused on examining the changes in genetic load throughout the domestication and subsequent feralization of *indica* rice, using *O. longistaminata* as the outgroup to infer derived alleles. On a genome-wide scale, non-synonymous and highly damaging SNPs that cause loss of function exhibited lower allele frequencies compared to synonymous SNPs, consistent with the influence of purifying selection (Figure 1J). Genetic load demonstrated a negative correlation with recombination rate across the genome (Supplemental Figure 10), suggesting less effective purifying selection in low-recombination regions. We also analyzed the ratio of non-synonymous and highly damaging alleles to synonymous alleles at the individual level within the wild, cultivated (landrace and variety), and EastCN-WR groups. Consistent with the "cost of domestication" theory (Lu et al., 2006), we found that the genetic load of cultivated rice is higher than that of wild rice (Supplemental Figure 10). Notably, the genetic load of EastCN-WR is intermediate between that of the wild and cultivated *indica* rice groups (Figure 1K; Supplemental Figure 9). Investigating the factors contributing to the observed reduction in genetic load led to an intriguing observation: the progenitor Nanjing11 also displayed a similarly low genetic load (Figure 1K). This finding suggests that the low genetic load of the cultivar progenitor may have played a pivotal role in enhancing the environmental adaptability of weedy rice, thereby contributing to its robust fitness.

Unlike earlier studies that relied on genome-wide genetic backgrounds to investigate the evolution of weedy rice, this study leverages knowledge of the specific known cultivar progenitor of a weedy rice population (EastCN-WR) to dissect the process of rice feralization. We have documented the role of adaptive introgression from local rice landraces in the evolution of these weed strains, particularly highlighting the *Rc* gene, which confers the key weed adaptation of red pericarp/seed dormancy (Figure 1L). Private SNPs in the EastCN-WR population that are present at differentially high frequencies are enriched in genes potentially controlling reproduction and embryological development, suggesting that adaptations related to reproductive strategies may be crucial for weed evolution. Notably, the cultivar progenitor of EastCN-WR, Nanjing11, is characterized by a low genetic load, which may have made it a "naturally selected" material, conferring robust fitness for weedy rice feralization (Figure 1L). Overall, our study offers a more refined genomic perspective on the genesis of weedy rice, highlighting that landrace introgression may be a particularly critical component enabling the feralization of modern cultivars.

DATA AND CODE AVAILABILITY

Sequence data of the weedy rice samples generated in this study have been deposited in the China National Genomics Data Center (https://ngdc.cncb.ac.cn), available under BioProject accession number PRJCA025860. The genome assembly of weedy rice JS_YC_W7 has been submitted to the China National GeneBank with the accession number CNA0142323. Additionally, the genome assembly and gene annotations for JS_YC_W7 can be accessed via the following URL: http://www.xhhuanglab.cn/data/FeralRice.html.

FUNDING

This work was supported by the National Natural Science Foundation of China (32170638 to J.Q.), the Shanghai Science and Technology Committee Rising Star Program (22QA1406800 to J.Q.), and the United States National Science Foundation (Plant Genome Research Program IOS-1947609 to K.M.O.).

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers for their valuable comments, which have significantly improved the quality of this manuscript. We also extend our gratitude to Prof. Hui Xia (South China Agricultural University) and all members of Xuehui Huang's lab for their suggestions and assistance with this project. No conflicts of interest are declared.

AUTHOR CONTRIBUTIONS

J.Q. and Xiaoyi Zhou conceptualized and designed the experiments. J.Q., M.Z., J.C., Xiaofang Zhou, K.L., and X.W. performed bioinformatic analyses. K.Y. and K.X. collected and sequenced the samples. X.H., K.M.O., and L.F. provided insightful suggestions and comments on the manuscript. J.Q. and K.M.O. wrote the manuscript. All authors have read and approved the final manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: May 8, 2024
Revised: July 25, 2024
Accepted: August 18, 2024
Published: August 22, 2024

Min Zhu[1,5], Kaicheng Yong[1,5], Kai Xu[2,5],
Jia Cong[1], Xiaofang Zhou[1], Keyue Liu[1],
Xuechen Wang[1], Longjiang Fan[3],

Kenneth M. Olsen[4], *Xuehui Huang*[1], *Xiaoyi Zhou*[1,*] *and Jie Qiu*[1,*]

[1]Shanghai Key Laboratory of Plant Molecular Sciences, College of Life Sciences, Shanghai Normal University, Shanghai 200234, China
[2]Crop Production Center, Bright Food Group Co., Ltd., Shanghai 224100, China
[3]Institute of Crop Science & Institute of Bioinformatics, College of Agriculture & Biotechnology, Zhejiang University, Hangzhou 310058, China
[4]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130, USA
[5]These authors contributed equally to this article.
***Correspondence: Xiaoyi Zhou (zhouxy@shnu.edu.cn), Jie Qiu (qiujie@shnu.edu.cn)**

https://doi.org/10.1016/j.xplc.2024.101066

REFERENCES

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, L.-F., Pusadee, T., Wedger, M.J., Li, Y.-L., Li, M.-R., Lau, Y.-L., Yap, S.-J., Jamjod, S., Rerkasem, B., Hao, Y., et al. (2024a). Porous borders at the wild-crop interface promote weed adaptation in Southeast Asia. Nat. Commun. **15**:1182.

Li, X., Lowey, D., Lessard, J., and Caicedo, A.L. (2024b). Comparative histology of abscission zones reveals the extent of convergence and divergence in seed shattering in weedy and cultivated rice. J. Exp. Bot. erae221. https://doi.org/10.1093/jxb/erae221.

Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., and Wu, C.-I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet. **22**:126–131.

Nadir, S., Xiong, H.-B., Zhu, Q., Zhang, X.-L., Xu, H.-Y., Li, J., Dongchen, W., Henry, D., Guo, X.-Q., Khan, S., et al. (2017). Weedy rice in sustainable rice production. Agron. Sustain. Dev. **37**:46.

Qi, X., Liu, Y., Vigueira, C.C., Young, N.D., Caicedo, A.L., Jia, Y., Gealy, D.R., and Olsen, K.M. (2015). More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. Mol. Ecol. **24**:3329–3344.

Qiu, J., Jia, L., Wu, D., Weng, X., Chen, L., Sun, J., Chen, M., Mao, L., Jiang, B., Ye, C., et al. (2020). Diverse genetic mechanisms underlie worldwide convergent rice feralization. Genome Biol. **21**:70.

Qiu, J., Zhou, Y., Mao, L., Ye, C., Wang, W., Zhang, J., Yu, Y., Fu, F., Wang, Y., Qian, F., et al. (2017). Genomic variation associated with local adaptation of weedy rice during de-domestication. Nat. Commun. **8**:15323.

Sun, J., Ma, D., Tang, L., Zhao, M., Zhang, G., Wang, W., Song, J., Li, X., Liu, Z., Zhang, W., et al. (2019). Population genomic analysis and de novo assembly reveal the origin of weedy rice as an evolutionary game. Mol. Plant **12**:632–647.

Wang, D.R., Agosto-Pérez, F.J., Chebotarov, D., Shi, Y., Marchini, J., Fitzgerald, M., McNally, K.L., Alexandrov, N., and McCouch, S.R. (2018). An imputation platform to enhance integration of rice genetic resources. Nat. Commun. **9**:3519.