



Short Communication

Widespread inversions shape the genetic and phenotypic diversity in rice

Wenchuang He^{a,1}, Huiying He^{a,1}, Qiaoling Yuan^{a,1}, Hai Zhang^{b,1}, Xiaoxia Li^a, Tianyi Wang^a, Yingxue Yang^a, Longbo Yang^a, Yuting Yang^a, Xiangpei Liu^a, Hua Wei^a, Hong Zhang^a, Bin Zhang^a, Mingliang Guo^a, Yue Leng^a, Chuanlin Shi^a, Yang Lv^a, Wu Chen^a, Xianmeng Wang^a, Zhipeng Zhang^a, Bohui Yu^a, Binta Zhang^a, Qiang Xu^a, Hongge Qian^a, Yongfeng Zhou^a, Shaokui Wang^{c,*}, Qian Qian^{a,d,e,*}, Lianguang Shang^{a,e,*}

^aShenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China

^bSchool of Life Sciences, Xiamen University, Xiamen 361005, China

^cState Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, Guangdong Provincial Key Laboratory of Plant Molecular Breeding, South China Agricultural University, Guangzhou 510642, China

^dState Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China

^eYazhouwan National Laboratory, Sanya 572024, China

ARTICLE INFO

Article history:

Received 17 September 2023

Received in revised form 4 December 2023

Accepted 26 December 2023

Available online 29 December 2023

© 2023 Science China Press. Published by Elsevier B.V. and Science China Press. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Genomic structural variations affected widely gene function and morphological traits in plants. Chromosomal inversions (INVs), as an important form of structural variation, can be large and extend to megabases in length [1], and form the genetic basis of local adaptation and ecotypic differentiation in sunflowers [2], *Boechera stricta* [3], monkeyflowers [4], mimetic butterflies [5] and ruffs [6], sex determination in nine-spined stickleback [7] and cancer and neurodevelopmental disease in human [8]. At present, structural variants, and in particular, INVs, remain largely uncharacterized in plants despite their importance for local adaptation in a variety of species [1], with only a few studies exploring INVs polymorphisms at the population level. Hence, it remains unclear whether adaptive INVs can directly shape the genetic basis for diverse phenotypes in plants, especially in domesticated crops.

As one of the most important staple food crops worldwide, rice has been extensively studied with abundant genomic sequencing data for wild and cultivated species and subpopulations performing various ecological and agronomic traits. That makes it an ideal subject for assessing the INV landscape and its potential application in population genetics, e.g., gene mining, in a domesticated species, especially following the publication of high-quality geno-

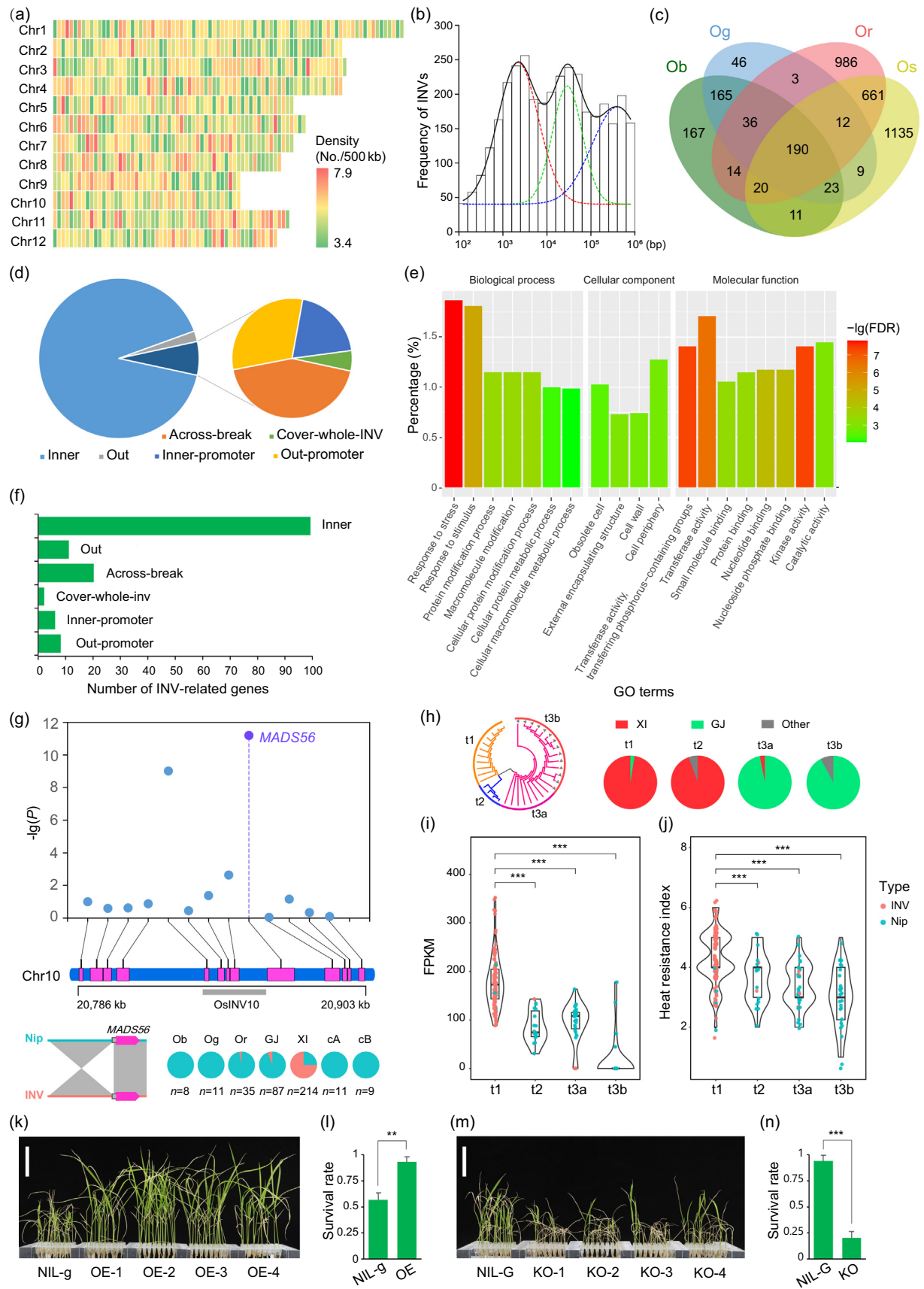
mic, transcriptomic, and phenotypic datasets for a pan-genome rice population in our previous study [9]. Recently, a pan-genome inversion index was built from 1769 non-redundant INVs based on comparative genomic analysis of 73 wild and cultivated genomes [10]. However, to fully release the great potential of INVs in population level, it is still necessary to generate and assess a large scale of INV dataset derived from genomic sequencing dataset that could be an important model for utilizing the large amount of genomic sequencing data accumulated for the past years in rice. In this study, we conducted a large-scale screening of the genomic INVs in 377 cultivated and wild rice accessions to understand the associated genetic and phenotypic variations. Our results provide a comprehensive picture of gene structure and expression changes caused by the widespread INVs in rice, indicating an important role of the genomic inversions in shaping the plant's genetic and phenotypic diversity.

A diverse and representative rice collection with long sequencing reads (Table S1 online) was constructed from two cultivated rice species, Asian cultivated rice (*Oryza sativa*) and African cultivated rice (*Oryza glaberrima*), and their wild relatives, *Oryza rufipogon* and *Oryza barthii*. A total of 3499 INVs were identified among the 12 chromosomes with a non-uniform distribution (Fig. 1a and Table S2 online), e.g., Chr11 contained the largest number and density of INVs (452 and 7.8/500 kb, respectively) while Chr1 have a much smaller number of INVs (286) with the lowest density (3.3/500 kb). Although lengths of the identified INVs varied

* Corresponding authors.

E-mail addresses: shaokuiwang@scau.edu.cn (S. Wang), qianqian188@hotmail.com (Q. Qian), shanglianguang@caas.cn (L. Shang).

¹ These authors contributed equally to this work.



extensively, ranging from 51 to 999,964 bp, we observed three peak modes of 1.8 kb, 24.3 kb, and 304.8 kb in the length distribution based on the gaussian mixture model (Fig. 1b), suggesting potentially different molecular mechanisms for the emergence of INVs with different lengths. Among them, 190 common INVs were shared with all the wild and cultivated rice species and 2334 INVs were species-specific (Fig. 1c and Fig S1 online). To evaluate the quality of the obtained INVs, PCR experiments were conducted to validate 9 selected INVs with specific primers around their inversion breakpoints (Fig. S2, Tables S3 and S4 online), which confirmed the existence of those inversions.

Functional annotation was performed to investigate potential structural changes in genes caused by the INVs. Among all genes located in the proximity of INVs, only 5.9% showed structural variation directly caused by INV breakpoints (Fig. 1d). These were classified according to their structural consequence, namely transcript region across the breakpoint (across-break, 3.0%); cover the whole INV (cover-whole-INV, 0.4%); promoter region across the breakpoint with the transcript region inside the INV (inner-promoter, 1.4%); promoter region across the breakpoint with the transcript region outside the INV (out-promoter, 2.1%). Those INV-related genes were significantly enriched in stress/stimulus response-related functions, cell periphery/wall, and transferase/kinase/catalytic activity (Fig. 1e). Compared to the whole genome dataset, the percentage of stress response-related genes was 1.9% higher than the background dataset. Likewise, the percentage of genes expressed at the cell periphery increased by 1.3% and the transferase activity-related genes increased by 1.7% (Fig. 1e). Those results suggested INVs played an important role in rice stress resistance probably by affecting those resistance genes expressed around the cell surface.

We then identified potential INVs that contribute to gene expression differences and phenotypic variations in rice populations. A total of 863 INVs were detected as expression quantitative trait loci (eQTLs) showing significant associations with the expression of 4722 genes (Fig. S3 online), in which 414 genes were previously associated with plant growth and development, photoperiod and temperature response, and biotic and abiotic stress resistance, etc. Among them, 162 genes were significantly regulated by at least one nearby INV in their different regions (Fig. 1f and Fig. S3 online). As INV-related genes showed a clear bias on stress response functions, we selected a local eQTL (OsINV10, Chr10_20836592_20861493) adjacent to a stress response gene, *LOC_Os10g39130* ($P = 6.2 \times 10^{-12}$), for further analysis (Fig. 1g and Fig. S4 online). The promoter region of this gene covered across the breakpoint of OsINV10 and was rearranged by this inversion event (Fig. 1g). *LOC_Os10g39130* is also known as *MADS56* (or *GL10*) conferring flowering time [11] and grain size [12] in rice. Homologous regions of *MADS56* were further screened in 197 cultivated rice accessions with a total of 37 haplotypes being detected that could be divided into four clusters (Fig. 1h). The INV-related haplotype cluster (t1) showed both significantly higher expression levels and significantly stronger heat-resistance compared to the

non-INV haplotypes as revealed by the Student's *t*-test (Fig. 1i, j), indicating that the *MADS56* was a candidate gene conferring rice heat-resistance. Genetic complementation experiments further confirmed that the transgenic rice plants with overexpressed *MADS56* showed significantly stronger heat-resistance than the wild-type plants (Fig. 1k, l) and knock-out of *MADS56* resulted in a significant decrease in rice heat tolerance (Fig. 1m, n). The *MADS56* has been previously reported to be involved in plant hormone signal transduction, especially in GA biosynthesis and signaling transduction in rice seedlings [12], indicating that the *MADS56* was likely to be involved in GA biosynthesis and signaling transduction related to the regulation of heat resistance.

There have been continuous efforts to understand the genetic and functional mechanism behind INVs for utilizing complex genomic variations in modern crop breeding. In this study, we have generated a large scale of population INV dataset to assess their functional effects on the genic and phenotypic variations in rice, which could serve as an integrated model for employing INVs in gene mining and functional genomics studies. The detection of INV from genomic sequencing data has the advantages of simplicity, efficiency, and a large amount of available data. However, many potential loci that are absent in the reference genome may be lost when compared to a single reference genome. Therefore, it might be a better approach to apply the pan-genome as a reference genome for variation analysis in future studies, after solving the problems in the accuracy and efficiency of mapping tools for pan-genome.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32188102 and 32372148), the National Key R&D Program of China (2022YFE0139400), Guangdong Basic and Applied Basic Research Foundation (2023B1515020053), and the Youth Innovation of Chinese Academy of Agricultural Sciences (Y20230C36).

Author contributions

Liangang Shang, Qian Qian, and Shaokui Wang conceived the study. Wenchuang He, Huiying He, Qiaoling Yuan, Xiaoxia Li, Tianyi Wang, and Longbo Yang performed data statistics analysis and bioinformatic analyses. Hai Zhang, Yingxue Yang, Yuting Yang, Xiangpei Liu, Hua Wei, Hong Zhang, Bin Zhang, Mingliang Guo, Yue Leng, Chuanlin Shi, Yang Lv, Wu Chen, Xianmeng Wang, Zhipeng Zhang, Bohui Yu and Bintao Zhang participated in the collection and assembly of genetic and phenotypic datasets. Wenchuang He, Huiying He, Qiaoling Yuan, and Hai Zhang wrote the manu-

Fig. 1. INV landscape and its effects on functional genes. (a) The chromosome distribution of INVs detected in cultivated and wild rice. (b) Length distribution of all detected INVs. Different colored fit curves represent three potential normal distributions of the INV lengths. (c) Shared and specific INVs among different rice species. Ob: *Oryza barthii*; Og: *Oryza glaberrima*; Or: *Oryza rufipogon*; Os: *Oryza sativa*. (d) Proportion of different INV-related genes. The inner genes denoted those genes located inside the INV regions except for the inner-promoter genes. The out genes denoted those genes located outside an INV region with a distance of < 2 kb except for the out-promoter genes. (e) GO enrichment for INV-related genes, compared to annotated genes in the reference genome database (Nipponbare). (f) Different INV-related genes were significantly regulated by the eQTLs. (g) The *MADS56* was a candidate gene for the eQTL, INV10. Population frequency of reference (Nip) and alternative (INV) alleles was also presented in the pie charts below. XI: Xian/Indica; GJ: Geng/Japonica; cA: circum-Aus; cB: circum-Basmati or Aromatic rice. (h) The haplotypes of *MADS56* genes and their population distribution in cultivated rice. The gray triangle denoted a large deletion in this gene as described by Zhan et al [12]. (i, j) Comparison of expression levels (i) and heat-resistance trait (j) between different haplotype clusters of *MADS56* genes. (k, l) Phenotypic performance (k) and heat-resistance trait (l) of the wild type (NIL-g) and its transgenic over-expression (OE) plants for *MADS56*. (m, n) Phenotypic performance (m) and heat-resistance trait (n) of the wild type (NIL-G) and its transgenic knock-out (KO) plants for *MADS56*. The white scale indicates a length of 5 cm in (k) and (m). The asterisks in (i), (j), (l), and (n) indicate statistical differences according to Student's *t* test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

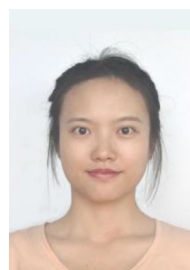
script and revised the manuscript with guidance from Lianguang Shang, Qian Qian, Shaokui Wang, and Yongfeng Zhou. All authors reviewed and commented on the manuscript, and approved its final submission.

Appendix A. Supplementary materials

Supplementary materials to this short communication can be found online at <https://doi.org/10.1016/j.scib.2023.12.048>.

References

- [1] Wellenreuther M, Bernatchez L. Eco-evolutionary genomics of chromosomal inversions. *Trends Ecol Evol* 2018;33:427–40.
- [2] Todesco M, Owens GL, Bercovich N, et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* 2020;584:602–7.
- [3] Lee CR, Wang B, Mojica JP, et al. Young inversion with multiple linked QTLs under selection in a hybrid zone. *Nat Ecol Evol* 2017;1:119.
- [4] Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol* 2010;8.
- [5] Joron M, Frezal L, Jones RT, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 2011;477:203–6.
- [6] Lamichhaney S, Fan G, Widemo F, et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet* 2016;48:84–8.
- [7] Natri HM, Merila J, Shikano T. The evolution of sex determination associated with a chromosomal inversion. *Nat Commun* 2019;10:145.
- [8] Antonacci F, Kidd JM, Marques-Bonet T, et al. Characterization of six human disease-associated inversion polymorphisms. *Hum Mol Genet* 2009;18:2555–66.
- [9] Shang L, Li X, He H, et al. A super pan-genomic landscape of rice. *Cell Res* 2022;32:878–96.
- [10] Zhou Y, Yu Z, Chebotarov D, et al. Pan-genome inversion index reveals evolutionary insights into the subpopulation structure of Asian rice. *Nat Commun* 2023;14:1567.
- [11] Ryu CH, Lee S, Cho LH, et al. *OsMADS50* and *OsMADS56* function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environ* 2009;32:1412–27.
- [12] Zhan P, Ma S, Xiao Z, et al. Natural variations in *grain length 10* (*GL10*) regulate rice grain size. *J Genet Genomics* 2022;49:405–13.



Qiaoling Yuan is a research associate in Prof. Lianguang Shang's lab at Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. Her research interest mainly focuses on superior alleles mining based on pan-genomes.



Hai Zhang is an associate professor at School of Life Sciences, Xiamen University. His research interest mainly focuses on the molecular mechanisms underlying rice heat stress resistance.



Shaokui Wang is a professor at the State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University. Her research interest mainly focuses on the genetics mechanisms of rice quality and yield, as well as molecular breeding by design in rice.



Wenchuang He is an associate professor at the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. His research interest mainly focuses on genetic and evolutionary applications of genomic and pan-genomic variations in rice.



Qian Qian is a professor at the Yazhouwan National Laboratory. His research interest mainly focuses on the molecular mechanisms and breeding application underlying important agronomic traits in rice.



Huiying He is an associate professor at the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. Her research interest mainly focuses on mining the complex structural variations in the rice pan-genome and identifying their effects to gene expressions and phenotypes.



Lianguang Shang is a professor at the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. His research interest mainly focused on elite gene mining and breeding application in rice.