Molecular Plant
Letter to the Editor

CellPress
PARTNER JOURNAL

Genome and Comparative Transcriptomics of African Wild Rice *Oryza longistaminata* Provide Insights into Molecular Mechanism of Rhizomatousness and Self-Incompatibility

Dear Editor,

Oryza longistaminata is an African wild rice species with AA genome type possessing special traits that are highly valued for improving cultivated rice, such as strong resistance to biotic and abiotic stresses (Song et al., 1995) for improving resistance of cultivars, rhizomatousness for perennial breeding (Glover et al., 2010), and self-incompatibility (SI) for new ways to produce hybrid seeds (Ghesquiere, 1986). Deciphering the genome of *O. longistaminata* will be the key to uncovering the mechanism of these hallmark traits and improving cultivated rice. However, deciphering the genome of this species remains a major challenge, since its high heterozygosity (1.3% estimated by k-mer distribution and 2.7 polymorphic sites per kilobase in the final assembly; Supplemental Figure 1, Supplemental Table 1, Supplemental Information S2.8) is the result of its inherent SI.

Here we assembled *de novo* a total length of ~347 Mb of the reference genome of *O. longistaminata* (contig N50 12.5 kb; scaffold N50 363 kb) based on a large quantity of combined sequencing data (~396 × Illumina short reads and ~5.9 × Roche GS FLX + long reads) (Supplemental Tables 2–8 and Supplemental Figures 2–6). Furthermore, we assembled scaffolds of *O. longistaminata* into 12 pseudo-chromosomes (Figure 1A) based on its syntenic relationship with *Oryza glaberrima* (Wang et al., 2014), and the resulting assembled genome was referred to as Oryza_longistaminata_v1.0 (Supplemental Information S8).

Annotations were conducted using the assembled genome. The transposable elements (TEs) annotation results show that 48.73% of the genome assembly were TEs (Supplemental Tables 9 and 10, and Supplemental Figure 7), which is a markedly higher percentage than previously found in *O. glaberrima* (34.25%), *Oryza brachyantha* (29.2%), and *Oryza sativa* (34.8%). We annotated 32 502 protein-coding genes (Supplemental Tables 11 and 12, and Supplemental Figures 8 and 9) and the non-protein-coding RNA genes, including 3954 putative miRNA, 720 tRNA, 34 rRNA, and 669 snRNA genes (Supplemental Table 13).

Divergent times analysis among six Gramineae species, including *O. sativa* ssp. *japonica*, *O. glaberrima*, *O. longistaminata*, *O. brachyantha*, *Phyllostachys heterocycla*, and *Brachypodium distachyon* (Figure 1B) indicates that the genus *Oryza* shared a common ancestor ~15.2 million years ago (MYA), with *O. longistaminata* having potentially diverged from *O. glaberrima* ~1.9 MYA. A total of 120 rapidly evolving genes between *O. longistaminata* and *O. glaberrima* were identified (Supplemental Table 14), eight of which are transcription factor genes with zinc-finger motif, helix-loop-helix DNA-binding domain, or GRAS transcription factor domain, while three are disease-resistance genes with an NB-ARC domain.

Gene family clustering analysis based on pair-wise genes among four rice species (*O. longistaminata*, *O. sativa, O. glaberrima*, and *O. brachyantha*) shows that 13 238 gene families were identified in these four species (Figure 1C). The expansion or contraction of gene families analysis using seven species (four rice species as described above as well as *B. distachyon*, *P. heterocycla*, and *Setaria italica*) showed that 393 gene families contracted and 474 expanded in the genome of *O. longistaminata* (Figure 1D). Our assembly-collapsed genomic region (Supplemental Information S2.6 and Supplemental Table 7) and genomic region recovery analyses (Supplemental Information S4.4) suggest that there are few false-positive family expansion/contraction events in our results. Among the contracted gene families, there is one family annotated as putative receptor protein kinase ZmPK1, which is related to SI in *Brassica* (Walker and Zhang, 1990), significantly contracted in the genome of *O. longistaminata* (the numbers of this gene family: *O. sativa*, 26; *O. brachyantha*, 22; *S. italica*, 12; *B. distachyon*, 12; *O. glaberrima*, 28; *P. heterocycla*, 9; and *O. longistaminata*, 4), indicating that this contracted gene family is probably related to the SI of *O. longistaminata*. Interestingly, among the expanded families, 20 gene families related to resistance, such as disease resistance, LRR receptor-like serine/threonine-protein kinase, or other gene family IDs annotated as stimuli-response (lectin receptor kinase or glutathione *S*-transferases), appeared to have expanded in the *O. longistaminata* genome (Supplemental Table 15). Aside from the resistance (R) gene families, we also carefully identified all R genes in these four rice genomes, and again *O. longistaminata* showed a greater number of R genes than did the other rice species (546 R genes in *O. longistaminata*, 480 in *O. glaberrima*, 466 in *O. sativa*, and 439 in *O. brachyantha*) (Supplemental Figure 11). Although there exist copy-number variations among individuals within a species, the conspicuous difference in R gene/family numbers between *O. longistaminata* and other rice species suggests that *O. longistaminata* as a species has more R genes/families than do other rice species.

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

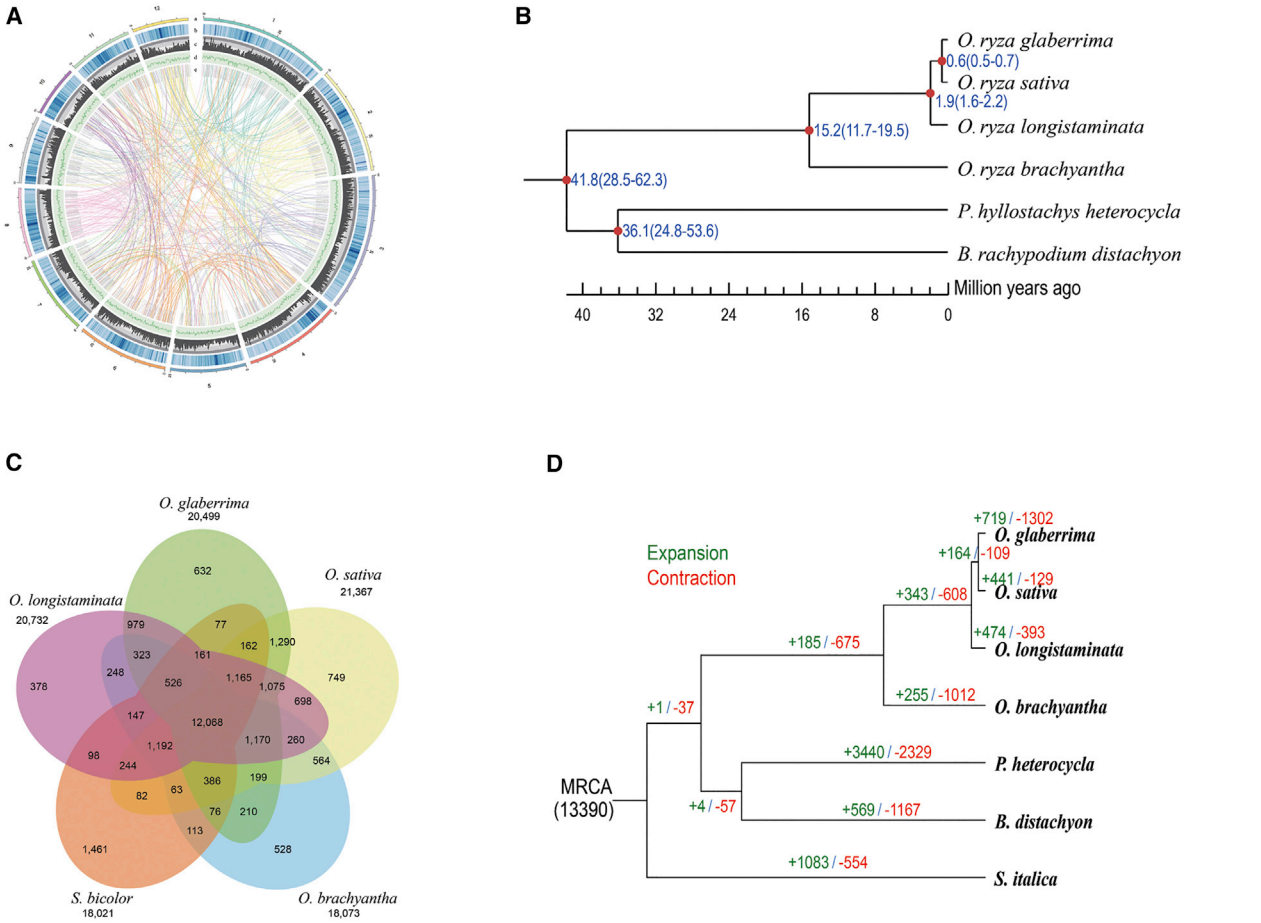

Figure 1. Genome of *O. longistaminata* and Genomic Analysis.

(A) Assembly genome of *O. longistaminata* (Oryza_longistaminata_v1.0). Concentric circles show the structural, functional, and evolutionary aspects of the genome: a, chromosome number; b, heatmap view of repeats; c, density of genes; d, GC content; e, paralogous relationships between chromosomes of *O. longistaminata*.

(B) Divergence time between *O. longistaminata* and other species.

(C) Clustering of gene families.

(D) Expanded and contracted gene families.

To investigate the mechanism of rhizomatousness of *O. longistaminata*, we performed transcriptome sequencing for two groups of tissues (rhizome versus stem and rhizome-tips versus stem-tips) (Supplemental Table 16). In total, 672 and 151 differential expression (DE) genes with four-fold expression differences were respectively identified in the two groups. KEGG enrichment analysis revealed that pathways of photosynthesis, photosynthesis–antenna proteins, carbon fixation in photosynthetic organisms, carbon metabolism, and metabolic pathways were enriched significantly in the DE genes between the rhizome and stem, and the pathway of taurine and hypotaurine metabolism, glyoxylate and dicarboxylate metabolism, carbon fixation in photosynthetic organisms, and plant hormone signal transduction were enriched in the DE genes between rhizome-tips and stem-tips, respectively (Supplemental Table 17). The results related to the plant hormones are consistent with those of a previous study (He et al., 2014), which also found that hormone genes might play key roles in rhizome development. Via QTL mapping, Hu et al. (2003) identified two loci, *Rhz2* and *Rhz3*, that may be dominant-complementary for rhizome formation in *O. longistaminata*. Here, we carefully compared these two QTL genomic regions (*Rhz2*, OSR13-OSR16; *Rhz3*, RM119-RM273) between *O. longistaminata* and *O. sativa* ssp. *japonica* cv. Nipponbare. The correspondent *Rhz2* and *Rhz3* regions in the *O. longistaminata* are about 960 kb and 2.73 Mb, respectively, and 122 and 331 respective annotated genes are in the *Rhz2* and *Rhz3* regions (Supplemental Figures 12 and 13; expression patterns of these genes in the two tissue groups are listed in Supplemental Data 1). In the *Rhz2* region, there are no DE genes and in the *Rhz3* region, only four DE genes were identified (three DE genes in the group of rhizomes versus stems and one gene in the group of rhizome-tips versus stem-tips; Supplemental Table 19). One of these four DE genes, *Olong01m10027813*, encodes a protein phosphatase 2C (PP2C) gene, and expresses greater than four-fold in the rhizomes than the stems. A previous study (Yu et al., 2003) indicated that PP2C is involved in the CLAVATA pathways, controlling stem cell identity in both shoot and flower meristems of *Arabidopsis*. Aligning the putative

protein sequence of *Olong01m10027813* to its homologs in *O. brachyantha*, *Oryza barthii*, *O. glaberrima*, *Oryza rufipogon*, *Oryza nivara*, and *O. sativa*, all of which have no rhizome, revealed that one non-synonymous substitution is consistently different between *O. longistaminata* and other non-rhizome species (Supplemental Figure 14). Based on these two pieces of evidence, it is plausible that *Olong01m10027813* is a likely candidate for the *Rhz3* gene, which offers a starting point for future functional explorations of the mechanism of rhizomatousness. Furthermore, we also identified DE genes in other 10 QTL regions that affect abundance of rhizomes identified by Hu et al. (2003). In total, 41 and 12 DE genes identified in the respective rhizome versus stem and rhizome-tips versus stem-tips groups were mapped within these 10 QTL regions (Supplemental Table 19). All the DE genes within the QTL regions provide us with good candidates to uncover the molecular mechanism of rhizomatousness of *O. longistaminata*.

The mechanism of SI in Gramineae is poorly understood; therefore, exploring the SI in *O. longistaminata* may extend our understanding of the SI in Gramineae, with the potential promise of enhanced ability of producing hybrid seeds. To explore the molecular mechanisms of SI of *O. longistaminata*, we performed transcriptomic analysis for two tissues groups: stamen group (stamens of *O. longistaminata* versus stamens of the self-compatibility hybrid line from cross between *O. longistaminata* and *O. sativa* ssp. *indica* RD23) and pistil group (pistils of *O. longistaminata* versus pistils of the self-compatibility hybrid line) (Supplemental Table 16), of which 571 and 999 DE genes with four-fold expression change were identified, respectively. KEGG analysis revealed that the pathways of pentose and glucuronate interconversions and carbon fixation in photosynthetic organisms were enriched in the DE genes of the stamen group, and no pathways were significantly enriched in the DE genes of the pistil group (Supplemental Table 17). In Gramineae, Yang et al. (2009) mapped the *S* and *Z* loci to perennial ryegrass (*Lolium perenne*, a species with SI of Gramineae) linkage groups one and two, respectively, which are syntenic to regions in the rice chromosome 5 and chromosome 4. Using the syntenic relationship between ryegrass and rice, we identified 15 DE genes within these two corresponding regions in *O. longistaminata* (Supplemental Table 20), of which one gene, *Olong01m10012815*, annotated as an EF-hand calcium-binding protein gene, expressed remarkably higher in pistils of the hybrid line (FPKM value: 112.6) compared with that in *O. longistaminata* (FPKM value: 0.252249), indicating that calcium-dependent signaling, like the SI in Papaveraceae (Franklin-Tong et al., 2002), may be involved in SI of *O. longistaminata*.

In summary, we present a reference genome of *O. longistaminata*, and studies of comparative genomics and transcriptomics identified genes and pathways that may be related to resistance, rhizomatousness, and SI. This work provides a basic evidential foundation for targeted studies of the genes underlying valuable phenotypic traits, future gene mining or breeding efforts, and further study into the evolution of African rice and the *Oryza* genus.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the National Natural Science Foundation of China (U1302264) to F.H., the National Basic Research Program of China (2013CB835200, 2013CB835201) and the Department of Sciences and Technology of Yunnan Province (2013GA004) to W.W. and F.H.

AUTHOR CONTRIBUTIONS

F.H., Wen Wang, and R.W. conceived the project and its components, designed the studies, and contributed to the original concept of the project. Y.Z., S.Z., L.L., and B.F. collected leaves of *O. longistaminata* for genomic sequencing and eight tissue samples for RNA-Seq. W. Wan, and Y.D. constructed short insert libraries and long insert libraries, and sequenced the short insert libraries on the Hiseq 2000 platform. H.L., X.L., J.C., and J.L. performed genome assembly analysis. M.X. and Y.S. conducted genome annotation, divergent time analysis, gene families clustering analysis, and analysis of expansion or contraction of gene families. L.K. and H.H. conducted flow cytometry experiments. Y.Z. analyzed heterozygosity, rapidly evolving genes, resistance (R) genes, rhizomatousness, and self-incompatibility. Wensheng Wang, L.H., J.Z., Q.Y., Q.S., Q.L., W.H., and D.T. took part in the preparation of samples. M.W, M.C., Y.Y., and R.W. modified the manuscript. Y.Z., Wen Wang, and F.H. analyzed the data as a whole and wrote the manuscript.

ACKNOWLEDGMENTS

We would like to thank Xueyan Li of the Kunming Institute of Zoology, Chinese Academy of Science, for helpful discussions. We would also like to thank Andrew Willden for English language editing of the manuscript. No conflict of interest declared.

Received: May 19, 2015
Revised: August 10, 2015
Accepted: August 12, 2015
Published: September 7, 2015

Yesheng Zhang[1,3,10], Shilai Zhang[2,10], Hui Liu[1,3,10], Binying Fu[4,10], Lijuan Li[2,10,11], Min Xie[5], Yue Song[5], Xin Li[1,12], Jing Cai[1,13], Wenting Wan[6], Ling Kui[1,3], Hui Huang[7], Jun Lyu[1], Yang Dong[1,6], Wensheng Wang[4], Liyu Huang[2], Jing Zhang[2], Qinzhong Yang[2], Qinli Shan[2], Qiong Li[2], Wangqi Huang[2], Dayun Tao[2], Muhua Wang[8], Mingsheng Chen[9], Yeisoo Yu[8], Rod A. Wing[8,], Wen Wang[1,*] and Fengyi Hu[2,14,*]*

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[2]Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
[3]University of Chinese Academy of Sciences, Beijing 100039, China
[4]Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing 100081, China
[5]BGI-Shenzhen, Shenzhen 518083, China
[6]Laboratory of Applied Genomics and Synthetic Biology, College of Life Science, Kunming University of Science and Technology, Kunming 650093, P. R. China
[7]Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwest China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China
[8]Arizona Genomics of Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

[9]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
[10]These authors contributed equally to this article.
[11]Present address: Laboratory of Molecular Breeding of Energy Plants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, 88 Xuefu Road, Kunming, Yunnan 650223, China
[12]Present address: Center for Epigenetics, Johns Hopkins University School of Medicine, Rangos 580, 855 North Wolfe Street, Baltimore, MD 21205, USA
[13]Present address: Institute for Systematic Botany, University of Zürich, Zollikerstrasse 107, 8008 Zürich, Switzerland
[14]Present address: School of Agriculture, Yunnan University, Kunming, Yunnan 650091, China
***Correspondence: Fengyi Hu (hfengyiyaas@aliyun.com), Wen Wang (wwang@mail.kiz.ac.cn), Rod A. Wing (rwing@ag.arizona.edu)**
http://dx.doi.org/10.1016/j.molp.2015.08.006

REFERENCES

Franklin-Tong, V.E., Holdaway-Clarke, T.L., Straatman, K.R., Kunkel, J.G., and Hepler, P.K. (2002). Involvement of extracellular calcium influx in the self-incompatibility response of *Papaver rhoeas*. Plant J. **29**:333–345.

Ghesquiere, A. (1986). Evolution of *Oryza longistaminata*. Rice Genet., 15–27.

Glover, J.D., Reganold, J.P., Bell, L.W., Borevitz, J., Brummer, E.C., Buckler, E.S., Cox, C.M., Cox, T.S., Crews, T.E., Culman, S.W., et al. (2010). Increased food and ecosystem security via perennial grains. Science **328**:1638–1639.

He, R., Salvato, F., Park, J.J., Kim, M.J., Nelson, W., Balbuena, T.S., Willer, M., Crow, J.A., May, G.D., Soderlund, C.A., et al. (2014). A systems-wide comparison of red rice (*Oryza longistaminata*) tissues identifies rhizome specific genes and proteins that are targets for cultivated rice improvement. BMC Plant Biol. **14**:46.

Hu, F.Y., Tao, D.Y., Sacks, E., Fu, B.Y., Xu, P., Li, J., Yang, Y., McNally, K., Khush, G.S., and Paterson, A.H. (2003). Convergent evolution of perenniality in rice and sorghum. Proc. Natl. Acad. Sci. USA **100**:4050–4054.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science **270**:1804–1806.

Walker, J.C., and Zhang, R. (1990). Relationship of a putative receptor protein-kinase from maize to the S-locus glycoproteins of *Brassica*. Nature **345**:743–746.

Wang, M., Yu, Y., Haberer, G., Marri, P.R., Fan, C., Goicoechea, J.L., Zuccolo, A., Song, X., Kudrna, D., Ammiraju, J.S., et al. (2014). The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. Nat. Genet. **46**:982–988.

Yang, B., Thorogood, D., Armstead, I.P., Franklin, F.C., and Barth, S. (2009). Identification of genes expressed during the self-incompatibility response in perennial ryegrass (*Lolium perenne* L.). Plant Mol. Biol. **70**:709–723.

Yu, L.P., Miller, A.K., and Clark, S.E. (2003). POLTERGEIST encodes a protein phosphatase 2C that regulates CLAVATA pathways controlling stem cell identity at *Arabidopsis* shoot and flower meristems. Curr. Biol. **13**:179–188.