Draft genomes of two outcrossing wild rice, *Oryza rufipogon* and *O. longistaminata*, reveal genomic features associated with mating-system evolution

Wei Li¹,² | Qun-Jie Zhang¹,² | Ting Zhu¹,³ | Yan Tong² | Kui Li¹,² | Cong Shi²,⁴ | Yun Zhang² | Yun-Long Liu² | Jian-Jun Jiang² | Yuan Liu² | En-Hua Xia² | Hui Huang² | Li-Ping Zhang² | Dan Zhang¹ | Chao Shi² | Wen-Kai Jiang² | You-Jie Zhao² | Shu-Yan Mao² | Jun-ying Jiao² | Ping-Zhen Xu² | Li-Li Yang² | Li-Zhi Gao¹,²

¹Institution of Genomics and Bioinformatics, South China Agricultural University, Guangzhou, China
²Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species in Southwestern China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China
³College of Life Science, Liaoning Normal University, Dalian, China
⁴University of the Chinese Academy of Sciences, Beijing, China

Correspondence Li-Zhi Gao, Institution of Genomics and Bioinformatics, South China Agricultural University, Guangzhou 510642, China. Email: Lgaogenomics@163.com

Abstract

*Oryza rufipogon* and *O. longistaminata* are important wild relatives of cultivated rice, harboring a promising source of novel genes for rice breeding programs. Here, we present de novo assembled draft genomes and annotation of *O. rufipogon* and *O. longistaminata*. Our analysis reveals a considerable number of lineage-specific gene families associated with the self-incompatibility (SI) and formation of reproductive separation. We show how lineage-specific expansion or contraction of gene families with functional enrichment of the recognition of pollen, thus enlightening their reproductive diversification. We documented a large number of lineage-specific gene families enriched in salt stress, antifungal response, and disease resistance. Our comparative analysis further shows a genome-wide expansion of genes encoding NBS-LRR proteins in these two outcrossing wild species in contrast to six other selfing rice species. Conserved noncoding sequences (CNSs) in the two wild rice genomes rapidly evolve relative to selfing rice species, resulting in the reduction of genomic variation owing to shifts of mating systems. We find that numerous genes related to these rapidly evolving CNSs are enriched in reproductive structure development, flower development, and postembryonic development, which may associate with SI in *O. rufipogon* and *O. longistaminata*.

Keywords

adaptation, genome and transcriptome assemblies, mating system evolution, *O. longistaminata* A. Chev. & Roehr., *Oryza rufipogon* Griff.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Plant Direct published by American Society of Plant Biologists, Society for Experimental Biology and John Wiley & Sons Ltd
1 | INTRODUCTION

The genus *Oryza* belongs to the grass family and consists of more than 20 wild species and two cultivated species (Khush, 1997; Vaughan, 1994). These rice species have been assigned ten genome types (AA, BB, CC, EE, FF, GG, BBCC, CCDD, HHJJ, and HHKK) (Aggarwal, Brar, & Khush, 1997; Ge, Sang, Lu, & Hong, 1999; Nayar, 1973), representing an enormous gene pool for genetic improvement of modern rice cultivars. The majority of alien genes involved in rice improvement are derived from wild AA-genome species to broaden gene pool of cultivated rice through introgression lines from the other wild relatives of *Oryza*. This bunch of AA-genome *Oryza* species comprises two cultivated rice (*O. sativa* and *O. glaberrima*) and six wild relatives (*O. barthii*, *O. longistaminata*, *O. nivara*, *O. rufipogon*, *O. meridionalis*, and *O. glumaepatula*), respectively (Vaughan, Morishima, & Kadowaki, 2003). They span a wide range of global pantropical regions and are disjunctively distributed in Asia, Africa, Australia, and South America (Vaughan, 1994). Phylogenomic studies showed that they have generated extensive AA-genome diversity from a common AA-genome ancestor and diverged within 4.8 million years (Myr) (Gao et al., 2019; Zhang et al., 2014; Zhu et al., 2014). There were frequent switches of mating systems between selfing (*O. sativa*, *O. glaberrima*, *O. barthii*, *O. nivara*, *O. meridionalis* and *O. glumaepatula*) and outcrossing species (*O. rufipogon* and *O. longistaminata*). This closely spaced series of recent speciation events has occurred with different life-history and breeding traits (Oka, 1988; Morishima, Sano, & Oka, 1992), providing an unparalleled system for understanding the gene and genome divergence that determines the wealth of phenotypic diversity and adaptive differences among multiple plant lineages.

*O. rufipogon* Griff., belonging to the genus Oryza, is thought to be the wild progenitor of Asian cultivated rice, *O. sativa* (Oka, 1988; Cheng et al., 2003; Fuller et al., 2010; Huang, Chen, et al., 2012; Huang, Kurata, et al., 2012; Khush, 1997). This perennial outcrossing plant species widely grows in diverse habitats of tropical and subtropical regions of Asia and Australia (Morishima et al., 1992). As a result of long historical domestication and improvement, *O. sativa* has experienced a considerable loss of genetic diversity through genome-wide bottlenecks and artificial selection for domestication traits compared to *O. rufipogon* (Kovach, Sweeney, & McCouch, 2007; Xu et al., 2011). Thus, *O. rufipogon* definitely harbors abundant source of novel alleles that are critical and necessary for rice breeding programs in the future. A large number of alien genes from *O. rufipogon* have successfully been introduced into cultivated rice, generating many environmentally resistant, and high-yielding varieties (Brar & Ramos, 2007), for example, the application of the “wild-abortive” allele for the fruitful raising of hybrid rice varieties (Lin & Yuan, 1980). In China, *O. rufipogon* is extensively found in Guangdong, Guangxi, Hainan, Yunnan, Hunan, Jiangxi, Fujian, and Taiwan provinces (Gao, Zhang, Zhou, Ge, & Hong, 1996). Unfortunately, human disturbance has driven this species at the edge of extinction (Gao et al., 1996). Recent decades have witnessed the progress in population and conservation genetic studies using multiple molecular markers to obtain novel insights into the extent, distribution, and dynamics of genetic diversity within and among natural populations of Chinese *O. rufipogon* (Gao, 2004). Despite the release of nuclear genomes of *O. rufipogon* and the seven other AA-genome *Oryza* species (Du et al., 2017; Goff et al., 2002; IRGSP, 2005; Stein, Yu, Copetti, Zwickl, & Zhang, 2018) the lack of a typical *O. rufipogon* genome may seriously impede its evolutionary and functional genomic studies. In-depth knowledge of genomic variation and population structure within the species is urgently needed to provide timely information useful for developing appropriate and efficient conservation management of wild rice germplasms.

*O. longistaminata* A. Chev. & Roehr., which is native to Africa, is among the eight AA-genome species in the genus *Oryza*. This wild rice species possesses highly valued traits to improve cultivated rice, including rhizomatousness for perennial rice breeding program (Glover et al., 2010), strong resistance to diseases and abiotic stresses (Song et al., 1995), and self-incompatibility (SI) for new procedures to generate seeds of hybrid rice (Ghesquiere, 1986). Recently, Zhang et al. (2015) reported a draft genome assembly of *O. longistaminata* with a relatively short (12.5 kb) contig N50 length. However, deciphering the *O. longistaminata* genome is fundamental to uncover molecular mechanisms that determine these remarkable agronomic traits to enhance rice genetic improvement.

The past decades have witnessed the completion of nuclear genomes of the two cultivated rice subspecies and a number of their relatives (Du et al., 2017; Goff et al., 2002; IRGSP, 2005; Shi et al., 2020; Stein et al., 2018; Wang et al., 2014; Yu et al., 2002; Zhang et al., 2014, 2015, 2016). Here, we sequenced and de novo assembled the two outcrossing wild rice genomes, *O. rufipogon* and *O. longistaminata*. We report the draft genome assembly and annotation of a typical *O. rufipogon* as well as its large transcriptome datasets using Illumina and 454 sequencing platforms. We also present an improved genome assembly and the annotation of *O. longistaminata* as well as large transcriptome datasets using Illumina sequencing platforms. Comparisons of all eight *Oryza* AA-genomes with well-defined phylogenetic framework and splitting times may undoubtedly serve as an important model that deserves endeavors for obtaining the full-genome patterns and multispecies perspective of evolutionary dynamics and genome dissimilarity. Access to the unprecedented data of these *O. rufipogon* and *O. longistaminata* genome sequences will be necessary for mining valuable alleles from this wild rice to boost the future rice breeding programs.

2 | METHODS

2.1 | Plant materials

An individual plant of *O. rufipogon* (RUF) was collected from a typical natural population grown in Yuanjiang County, Yunnan Province, China (Gao, Wei, Yang, Hong, & Ge, 2001). Another single plant of *O. longistaminata* (LON), which was originally from Botswana, was introduced from IRRI, Los Banos, the Philippines,
and grown in the greenhouse of Kunming Institute of Botany, the Chinese Academy of Sciences. High-quality genomic DNA was extracted from leaves using a modified CTAB method (Porebski, Bailey, & Baum, 1997). Total RNA was individually isolated from the four tissues of *O. rufipogon* and *O. longistaminata*, including 30-d-roots, 30-d-shoots, panicles at booting stage and flag leaves at booting stage, using a Water Saturated Phenol method (Zhang et al., 2014).

### 2.2 Genome sequencing on Illumina and 454 platforms

Three short-insert (180, 300, and 500 bp) paired-end and six long-insert (2, 4, 6, 8, 20, and 40 Kb) mate-pair genomic DNA libraries of *O. rufipogon* were constructed and sequenced using Hiseq2000 platform. A 454 sequencing library was also prepared and sequenced using 454 FLX platform. Three short-insert (300, 360 and 500 bp) paired-end and five long-insert (2, 4, 5, 6 and 8 Kb) mate-pair genomic DNA libraries of *O. longistaminata* were constructed and sequenced using Hiseq2000 platform.

### 2.3 RNA sequencing on Illumina and 454 platforms and transcriptome de novo assembly

RNA sequencing (RNA-Seq) libraries were constructed and paired-end sequenced using the Illumina Hiseq2000 platform to generate RNA-Seq data for *O. rufipogon*. We constructed the four libraries for 30-d-roots, 30-d-shoots, panicles at booting stage and flag leaves at booting stage, which were sequenced using Roche 454 platform to yield sequencing reads. To obtain the transcriptome of *O. rufipogon*, we assembled Illumina sequencing reads using Trinity (version r20140717) with default parameters (Grabherr et al., 2011). The transcriptome of *O. rufipogon* was also assembled based on 454 sequencing reads using MIRA (Version 4.0.2) with default parameters (Chevreux et al., 2004). RNA-Seq libraries were also constructed and paired-end sequenced using the Illumina Hiseq2000 platform to produce RNA-Seq data for *O. longistaminata*. To obtain the transcriptome of *O. longistaminata*, we assembled Illumina sequencing reads using Trinity (version r20140717) with default parameters (Grabherr et al., 2011).

### 2.4 Estimation of genome sizes

Nuclear genome sizes of the *O. rufipogon* and *O. longistaminata* plants were estimated using flow cytometry analysis, and *O. sativa* ssp. *japonica* cv. Nipponbare (SAT) was selected as standard with 0.794 pg (~389 Mb) genome size (Cavaler-Smith, 1985; IRGSP, 2005). The genome sizes were further estimated using k-mer frequencies, which were calculated using the pregraph module implemented in SOAPdenovo (version 1.05) (Li et al., 2010).

### 2.5 Assembly of the *O. rufipogon* and *O. longistaminata* genomes

Raw sequencing data were preprocessed using Trimomatic (version 0.33) (Bolger, Lohse, & Usadel, 2014). SOAPdenovo (Luo et al., 2012) was used to assemble the *O. rufipogon* and *O. longistaminata* genomes. Gaps within scaffolds were filled using GapCloser (version 1.12) (Luo et al., 2012). To solve heterozygous regions in the genome, we finally employed Platanus (v1.2.4) (Kajitani et al., 2014) to assemble these two genomes. All the short insert size libraries were input to assemble subprogram to construct contigs with parameters “-c 20 -k 32.” After obtaining contig sequences, PE reads and MP reads were realigned onto these contigs by Platanus with scaffold subprogram with default parameters to determine the orders of the contigs, remove bubbles and branches, and finally form scaffold sequences. To fill gaps in the assembled scaffolds, GapCloser (version 1.12) (Luo et al., 2012) was adopted with parameters “-p 25 -l 150” using PE reads. Then, the assemblies were masked, and HaploMerger (Huang, Chen, et al., 2012; Huang, Kurata, et al., 2012) was used to further remove redundant sequences.

### 2.6 Prediction of protein-coding genes and quality validation

Repetitive sequences of the *O. rufipogon* and *O. longistaminata* genome assemblies were masked prior to gene predictions. A combined strategy that integrates ab initio, protein and expressed sequence tag (EST) evidences was adopted to predict protein-coding genes of *O. rufipogon* and *O. longistaminata*. Augustus (version 3.0.3) (Stanke, Steinkamp, Waack, & Morgenstern, 2004), GlimmerHMM (version 3.0.3) (Majoros, Pertea, & Salzberg, 2004) and GeneMarkHMM (version 3.47) (Lukashin & Borodovsky, 1998) were used to detect the hypothetical gene-coding regions within genomes. The protein sequences from *O. sativa* ssp. *japonica* cv. Nipponbare (SAT), *O. nivara* (NIV), *O. glaberrima* (GLA), *O. barthii* (BAR), *O. glumaepatula* (GLU), *O. meridionalis* (MER), *O. brachyantha*, *Zea mays*, *Sorghum bicolor*, and *Brachypodium distachyon* were aligned to the genome assemblies using GenBlastA (version 1.0.1) (She, Chu, Wang, Pei, & Chen, 2009) and further refined by GeneWise (version 2.2.0) (Birney, Clamp, & Durbin, 2004). To improve the quality of gene predictions, we aligned the assembled transcripts to the genome assembly using PASA (Program to Assemble Spliced Alignments) (Haas et al., 2003) to determine the potential gene structures. We used EVidenceModeler (EVM) (Haas et al., 2008) to combine the ab initio gene predictions, protein alignments and transcription alignments described above into weighted consensus gene structures. To validate the predicted gene models, protein sequences of SAT (v 7.0) genome were downloaded from MSU database (http://rice.plantbiology.msu.edu/), and then, all these peptide sequences were aligned to gene models using BLAT.
2.7 | Annotation of repeat sequences

RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html), RECON (Bao & Eddy, 2002), and RepeatScout (Price, Jones, & Pevzner, 2005) were used to build de novo repeat elements. The LTR retrotransposons were identified by LTR_STRUC (McCarthy & McDonald, 2003), LTRHarvest (Ellinghaus, Kurtz, & Willhoef, 2008), LTR_FINDER (Xu & Wang, 2007), MGEScan (Lee et al., 2016), and LTR retriever (Ou & Jiang, 2018). Previously annotated transposons were retrieved from the collected database (release 9.1) with default parameters. We created a combined library for RepeatMasker (Chen, 2004; Smith, Hubley, & Green, 2016). Simple Sequence Repeats (SSRs) were used to create a combined library for RepeatMasker (Chen, 2004; Smith, Hubley, & Green, 2016). Simple Sequence Repeats (SSRs) were identified and located using MISA (http://pgrc.ipk-gatersleben.de/misa/). We combined SSRs from the plus and minus strands and differences caused by reading frames.

2.8 | Annotation of noncoding RNA genes

The transfer RNA (tRNA) genes, ribosomal RNA (rRNA) genes, small nucleolar RNA (snRNA) genes, small nuclear RNA (snRNA) genes, and microRNA (miRNA) genes were predicted using de novo and homology search methods. We used tRNAscan-SE algorithms (version 1.23) (Lowe & Eddy, 1997) with default parameters to identify tRNA genes. The rRNA genes (5S, 18S, and 28S rRNA genes) were predicted using RNAmmer (Lagesen et al., 2007) algorithms with default parameters. The snRNA genes were annotated using snoScan with the yeast rRNA methylation sites and yeast rRNA sequences provided by the snoScan (Lowe & Eddy, 1999) algorithm. The snRNA genes were identified using the INFERNAL (Nawrocki, Kolbe, & Eddy, 2009) software on the Rfam (Griffiths-Jones et al., 2005) database (release 9.1) with default parameters. We annotated miRNAs in two steps. First, we downloaded the existing rice miRNA entries from miRBase release 18.0 (Kozomara & Griffiths-Jones, 2011). Then, the conserved miRNAs were identified by mapping all miRBase-recorded rice miRNA precursor sequences against the assembled O. rufipogon and O. longistaminata genomes using BLASTN with cutoffs at E-value <1e-5, identity >80%, and query coverage >80%. Second, additional miRNA genes were identified by aligning all miRBase-recorded grass miRNA precursor sequences against our assembled genomes using BLASTN with cutoffs at E-value <1e-5, identity >60%, and query coverage >60%.

2.9 | Gene family clustering and evolutionary analyses

OrthoMCL (version 2.0.9) (Li, Stoeckert, & Roos, 2003) was used to identify gene families among O. sativa ssp. japonica, O. rufipogon, O. nivara, O. barthii, O. glaberrima, O. glumaepatula, O. longistaminata, O. meridionalis, and O. punctata (PUN), which were separately downloaded from MSU Rice Genome Annotation Project Database (http://rice.plantbiology.msu.edu) and Oryza AA Genomes Database (Zhang et al., 2014). Second, the filtered protein sequences from these nine rice species were compared using all-versus-all Blastp with an E-value of 1E-5. Finally, gene families among the nine rice species were clustered using a Markov cluster algorithm (MCL) with an inflation parameter of 1.5. According to the presence and absence of genes for a given species, we retrieved and classified species-specific gene families. An updated version of CAFE (version 3.1) (De Bie, Cristianini, Demuth, & Hahn, 2006) implemented with the likelihood model was used to examine expansions and/or contractions of gene families. Functional enrichment analysis of gene families with the expansion, contraction, and species-specificity was performed using Fisher's exact test with false discovery rate (FDR) corrections. PFAM domains or gene ontology (GO) terms for each gene used in functional enrichment analyses were directly extracted from the InterProScan entries.

2.10 | Phylogenomic analyses

The orthologous gene families among O. sativa ssp. japonica, O. rufipogon, O. nivara, O. barthii, O. glaberrima, O. glumaepatula, O. longistaminata, O. meridionalis, and O. punctata were constructed using OrthoMCL (version 2.0.9) (Li et al., 2003). RAxML package (version 8.1.13) (Stamatakis, 2006) was used to resolve phylogenetic relationships among these nine Oryza species. Phylogenomic tree was finally constructed using RAxML package (version 8.1.13) (Stamatakis, 2006) using O. punctata as outgroup. Divergence times among these species were estimated using the “mcmctree” program implemented in the PAML package (Yang, 1997).

2.11 | R-gene identification and classification

The identification of R-genes within the O. sativa ssp. japonica, O. rufipogon, O. nivara, O. barthii, O. glaberrima, O. glumaepatula, O. longistaminata, O. meridionalis, and O. punctata genomes was performed using a reiterative method (Zhang et al., 2014). Briefly, protein sequences of these rice genomes were first aligned against the raw Hidden Markov Model (HMM) of NB-ARC family (PF00931) using HMMER (version 3.1b1) (Finn, Clements, & Eddy, 2011) with default parameters, respectively. High-quality hits with an E-value of ≤1E-60 were retrieved and self-aligned using MUSCLE (version 3.8.31) (Edgar, 2004) to construct each rice species-specific NBS HMMs, respectively. Based on these specific HMMs, scanning whole O. sativa ssp. japonica, O. rufipogon, O. nivara, O. barthii, O. glaberrima, O. glumaepatula, O. longistaminata, O. meridionalis, and O. punctata proteomes was again conducted, and genes with each rice species-specific PF00931 domains were defined as R-genes, respectively. The identified R-genes were further classified using TIR domain (PF01582 and LRR domains (PF00560, PF07725, PF12799, PF13306, PF13516, PF13504, and PF13855). These two types of
PFAM domains could be detected using HHMNER (version 3.1b1) (Finn et al., 2011). CC domains within R-genes were identified using ncoils (Lupas, Van Dyke, & Stock, 1991) with default parameters.

2.12 Construction of the ortholog syntenic map of the eight AA-genome Oryza species

To aid evolutionary analyses, we accurately identified and aligned orthologous genomic regions from the eight assembled AA-genomes using MERCATOR (Dewey, 2007) and MAVID (Bray & Pachter, 2004). Orthologous sequence alignments were provided with the confirmed phylogenetic relationships (O. sativa ssp. japonica: 0.002022, O. rufipogon: 0.003527): 0.000373, O. nivara: 0.002659): 0.001616, (O. barthii: 0.001242, O. glaberrima: 0.001793): 0.001899): 0.001268, O. glumeapatula: 0.004631): 0.007018, O. longistaminata: 0.011113): 0.003953, O. meridionalis: 0.014768).

2.13 Identification and analysis of conserved noncoding sequences

We identified conserved noncoding sequences (CNSs) using GERP++ software (Cooper et al., 2005). The nucleotide substitution frequency of a CNS was estimated by nucleotide substitution frequency = 2 \sum_{i \neq j} \frac{d_{ij}}{(n-1)}, where \( d_{ij} \) is an estimate of the number of nucleotide substitutions per site between DNA sequence \( i \) and \( j \), and \( n \) is the number of the examined sequences. For each rapidly evolving region in CNSs, nucleotide substitution frequency was subsequently calculated for each species using an average of nucleotide substitution frequency of a species within this genomic region compared to other seven Oryza AA-genome species. We then used the nucleotide substitution frequency of this rapidly evolving region to compare with average nucleotide substitution frequency of total CNSs by Chi-squared test.

3 RESULTS

3.1 Genome sequencing, assembly and quality assessment

We sequenced nuclear genomes of two wild rice species: O. rufipogon from Asia and O. longistaminata from Africa (Table S1). We performed a whole-genome shotgun sequence (WGS) analysis with the next-generation sequencing (NGS) Illumina and 454 platforms. This generated raw sequencing datasets of ~174.85 Gb (RUF) and ~176.09 (LON), respectively (Figure S3a and Table S5). The resulting genome assembly was referred to as Oryza_rufipogon_v1.0. The N50 lengths of the assembled LON contigs and scaffolds were ~16.89 Kb and ~1.13 Mb, respectively (Table 1 and Table S4). The contig N50 and scaffold N50 sizes represented ~1.35-fold and ~3.10-fold improvement in length from the previously reported (Zhang et al., 2015) ~12.5 Kb and 364 Kb, respectively. About 96.94% of the LON assembly fell into 391 scaffolds larger than 100 Kb in length (Figure S3a and Table S5). The resulting genome assembly was referred to as Oryza_longistaminata_v1.0.

Using two orthogonal methods, we estimated genome sizes of O. rufipogon and O. longistaminata are between ~383 and 388 Mb and between ~363 and 392 Mb, respectively (Figures S1 and S2 and Table S3). These genomes were assembled using the clean reads, resulting in final assemblies of ~441.41 Mb (RUF) and ~332.08 Mb (LON), respectively (Table 1 and Table S4). The N50 lengths of the assembled RUF contigs and scaffolds were ~18.88 Kb and ~1.94 Mb, respectively (Table 1; Table S4). About 97.62% of the RUF assembly falls into 391 scaffolds larger than 100 Kb in length (Figure S3a). The resulting genome assembly was referred to as Oryza_rufipogon_v1.0. The N50 lengths of the assembled LON contigs and scaffolds were ~16.89 Kb and ~1.13 Mb, respectively (Table 1 and Table S4). The contig N50 and scaffold N50 sizes represented ~1.35-fold and ~3.10-fold improvement in length from the previously reported (Zhang et al., 2015) ~12.5 Kb and 364 Kb, respectively. About 96.94% of the LON assembly fell into 419 scaffolds larger than 100 Kb in length (Figure S3b and Table S5). The resulting assembled genome was referred to as Oryza_longistaminata_v1.0.

To validate the quality of the O. rufipogon and O. longistaminata genome assemblies, we first aligned all available DNA and protein sequences of RUF and LON from public databases and obtained mapping rates of ~86.94% and 84.50%, and ~64.43% and 66.08%, respectively (Table S6); second, we mapped high-quality reads of RUF and LON genome sequences to the assembled genome sequences, showing good alignments with average mapping rate of 94.50% and 85.31%, respectively (Table S7); third, we evaluated the quality of WGS assemblies through aligning the assembled RUF and LON genome sequences to the O. sativa ssp. japonica cv. Nipponbare genome. After removing repeat sequences, average mapping rates and sequence similarities were ~83.17% and 66.51%, and ~97.02% and 94.79%, respectively (Table S8). The assembly quality of RUF was additionally confirmed by aligning a total of the five WGS
scaffolds with lengths varying from 385 to 680 Kb against the available contigs assembled using SMRT technology. After eliminating repeat sequence-masked and gap regions, pairwise alignments yielded high sequence similarities of ~99.54% to 99.64% with 100% coverage of sequence length (Figure S4 and Table S9).

### 3.2 | Genome annotation

To aid the gene annotation, we de novo assembled the transcriptomes of *O. rufipogon* and *O. longistaminata*. A total of ~29.79 Gb and 223 Mb (Table S10) RNA-Seq data from RUF were generated on Illumina Hiseq2000 and Roche 454 platform, respectively. We assembled the Illumina data into 109,000 transcripts with a N50 length of 1,193 bp and a total length of ~112 Mb (Table S11). The transcriptome of *O. rufipogon* was also assembled based on 454 sequencing reads, resulting in 33,496 transcripts with a N50 length of 582 bp and a total sequence length of 18,905,412 bp (Table S11). We also assembled RNA-Seq data (Table S10) from LON, and this process generated 111,105 transcripts with a N50 length of 1,064 bp and a total sequence length of ~74,561,481 bp (Table S11).

In combination with ab initio prediction, protein and public EST alignments, EVM combing and further filtering, we predicted 52,997 genes and 40,014 protein-coding genes for RUF and LON, respectively (Figure S5 and Table S12). After the predicted genes were functionally annotated against InterPro, Pfam, and GO protein databases, we assembled the Illumina data into 109,000 transcripts with a N50 length of ~112 Mb (Table S11). The transcriptome of *O. rufipogon* was also assembled based on 454 sequencing reads, resulting in 33,496 transcripts with a N50 length of 582 bp and a total sequence length of 18,905,412 bp (Table S11). We also assembled RNA-Seq data (Table S10) from LON, and this process generated 111,105 transcripts with a N50 length of 1,064 bp and a total sequence length of ~74,561,481 bp (Table S11).

In combination with ab initio prediction, protein and public EST alignments, EVM combing and further filtering, we predicted 52,997 and 40,014 protein-coding genes for RUF and LON, respectively (Figure S5 and Table S12). After the predicted genes were functionally annotated against InterPro, Pfam, and GO protein databases, we aligned protein sequences of *O. sativa* ssp. *japonica* cv. *Nipponbare* and the above-mentioned RNA-Seq data of RUF and LON, representing the major tissue types and different developmental stages, to assess the quality of gene prediction. Our results showed that ~89.8% and ~81.6% gene models were supported by transcripts or proteins (identity ≥30% and coverage ≥90%) in RUF and LON, respectively (Figure S6 and Table S13).

Our annotation of repetitive sequences showed that approximately 45.18% of the RUF genome consists of TEs (Figure 1a and Table S14), slightly lower than that (~50.97%) annotated in the *Nipponbare* genome with the same methods (Figure 1a and Table S14). LTR retrotransposons and MULEs were the most abundant TE types, occupying roughly 15.26% and 11.21% of the RUF genome, respectively. The annotation of repetitive sequences revealed that approximately 36.69% of the LON genome consists of TEs (Table S14). LTR retrotransposons and MULEs were also the most abundant repeated sequences, occupying approximately 10.66% and 6.53% of the LON genome. We annotated a total of 214,337 SSRs (Table S15) in RUF and 184,773 SSRs (Table S15) in LON, they will provide valuable genetic markers to assist rice breeding programs (Jurka & Pethiyagoda, 1995).

### 3.3 | Evolutionary dynamics of rice gene families

To investigate evolutionary dynamics of gene families underlying physiological and phenotypic changes and the adaptation of wild rice species, we compared the nine predicted proteomes of *O. sativa*, *O. rufipogon*, *O. nivara*, *O. glaberrima*, *O. barthii*, *O. glumeapatula*, *O. longistaminata*, *O. meridionalis*, and *O. punctata*. A total of 357,284 protein sequences (Table S18) were included in the analysis. Finally, we generated a total of 236,506 orthologous gene families which comprised 304,101 genes (Table S19). This revealed a core set of 143,538 genes belonging to 10,100 gene clusters that were shared among all nine rice species, representing ancestral gene families in AA-genome *Oryza* species (Figure 1b). Interestingly, 650 (1,459 genes) and 313 (652 genes) gene clusters were found unique to RUF and LON, respectively (Figure 1b). Functional analyses of RUF-specific genes by both GO terms and PFAM domains revealed the enriched functional categories related to pathogenesis (GO: 0009405, p < .001), pollen allergen (PF01357, p < .001), stress upregulated Nod 19 (PF07712, p < .001), and root cap (PF06830, p < .001) (Table S20). Functional analyses of LON-specific genes further showed functional categories enriched in peroxidase activity (GO: 0004601, p < .001), plastid (GO: 0009511, p < .001), and root cap (PF06830, p < .001) (Table S20).

**FIGURE 1** Genome annotation of *O. rufipogon* and *O. longistaminata*. (a) Genome constituents of the annotated genes and repeat sequences; (b) The shared and unique gene families among the eight AA-genome *Oryza* species using *O. punctata* (BB-genome) as outgroup.
p < .001), oxidation-reduction process (GO: 0055114, p < .001), and response to oxidative stress (GO: 0006979, p < .001) (Table S21). The creation of new gene families in these two wild rice species may have contributed to the observed SI, response to biotic and abiotic stresses, and formation of reproductive separation, which are vital for reproductive success and enhance the abilities of strong adaptation in a remarkably diverse range of habitats in Asia and Africa.

To understand the expansion or contraction of rice gene families causing phenotypic diversification and speciation, we characterized gene families that underwent detectable changes and divergently evolved along different branches with a particular emphasis on those involved in phenotypic traits and environmental adaptation. Our results showed that, of the 19,539 gene families (20,9968 genes) inferred to exist in the most recent common ancestor of the nine studied rice species, 2,459 (2,426) and 1,579 (6,493) exhibited significant expansions (contractions) (p < .001) in the RUF and LON lineages, respectively (Figure 2). Remarkably, functional annotation showed that a large number of genes enriched in functional categories involved in the recognition of pollen (GO: 0048544, p < .001), disease resistance, including NB-ARC domain (PF00931, p < .001), Leucine rich repeats (PF13516, PF13855, PF12799, PF00560; p < .001) and Leucine rich repeat N-terminal domain (PF08263, p < .001), and salt stress response/antifungal (PF01657; p < .001) were significantly amplified in RUF in comparisons with the eight other rice species (Table S22). Compared with the eight other rice species, we surprisingly found that gene families in LON were significantly enriched in a number of functions related to the recognition of pollen (GO: 0048544, p < .001), salt stress response/antifungal (PF01657; p < .001), disease resistance, including NB-ARC domain (PF00931; p < .001), Leucine rich repeats (PF13516, PF13855, PF12799, PF00560; p < .001), and Leucine rich repeat N-terminal domain (PF08263, p < .001) (Table S23).

Genome-wide comparative analysis of the nucleotide-binding sites with leucine-rich repeat (NBS-LRR) genes further showed a remarkable expansion of gene families relevant to an enhanced disease resistance in RUF and LON. In total, we identified 631, 845, 489, 450, 476, 392, 768, 416, and 426 genes encoding NBS-LRR proteins in SAT, RUF, NIV, GLA, BAR, GLU, LON, MER, and PUN, respectively (Table S24). This amplification in RUF and LON is mainly attributable to an increase in CC-NBS, CC-NBS-LRR, NBS, and NBS-LRR domains, further supporting that they may have played an important role in biotic resistance and abiotic stresses. We positioned these orthologous R-genes to definite genomic locations across the SAT chromosomes (Figure 3), displaying an almost unequal distribution of the amplified NBS-encoding genes throughout the entire genome, among which Chromosome 11 harbored the utmost number of R-genes for all these nine rice species. They will evidently offer a large number of candidate loci for further functional genomic studies on disease resistance and rice breeding programs.

3.4 | Rapid evolution of CNSs

Conserved noncoding sequences are genomic regions showing a reduced mutation frequency of noncoding bases, many of which are regulatory elements that evolve under purifying selection.
To aid evolutionary analyses of CNSs, we first identified and aligned orthologous genomic regions from the eight AA-genome assemblies. In total, we obtained 8,742 orthologous genomic segments among the eight AA-genome *Oryza* species, ranging from 40.02% in RUF to 52.58% in GLA (Table S25). Through this orthologous synteny map as a framework, we identified a total of 67,154 CNSs across the eight AA-genome *Oryza* species. The total length of these CNSs was 33,258,680 bp, and the length of these CNSs was 495 bp on average. To examine sequence divergence of these CNSs among *Oryza* AA-genomes, we identified rapidly evolving regions in CNSs by comparing nucleotide substitution frequency between each CNS and the average of all CNSs. The average nucleotide substitution frequency of these regions was 3.46%, which is significantly higher than the average (1.44%, Chi-squared test, \(p < 2.2\times10^{-16}\)). A set of 3,123 CNSs was finally identified as rapidly evolving regions (Chi-squared test, \(p < 0.001\)). Genomic regions with \(p\) value < .001 were detected as species-specific rapidly evolving regions. Results indicated that the distribution of these species-specific regions that rapidly evolve were associated with upstream 2 Kb regions, 5’UTRs, introns, 3’UTRs, and downstream 2 Kb regions of protein-coding genes in the eight *Oryza* AA-genomes (Figure 4). The proportions of the rapidly evolving regions were particularly elevated in introns and downstream regions in comparisons with other genomic regions. Comparisons of species-specific rapidly evolving regions in CNSs among eight *Oryza* AA-genomes showed that nucleotide substitution frequencies of the two outcrossing species, RUF and LON, were higher than those selfing rice species (Table S26). Nucleotide substitution frequency of the rapidly evolving regions within CNSs was 5.68% in *O. longistaminata*, which is significantly higher than the average (3.28%) in *O. sativa* ssp. *japonica*, *O. nivara*, *O. glaberrima*, and *O. barthii* (Chi-squared test, \(p < 2.2\times10^{-16}\)). Similarly, nucleotide substitution frequency of the rapidly evolving regions in CNSs of *O. rufipogon* was 3.79%, which is also significantly higher than its close relatives, such as *O. sativa* ssp. *japonica* (3.06%, Chi-squared test, \(p < 2.2\times10^{-16}\)) and *O. nivara* (3.52%, Chi-squared test, \(p < 2.2\times10^{-16}\)). The results suggest that a considerable variation of CNSs has been largely reduced as a result of shifts of mating system from outcrossing to selfing. GO of genes in which upstream 2 Kb region, 5’UTR, intron, 3’UTR, and downstream 2 Kb region are associated with the rapidly evolving regions were annotated in each of eight *Oryza* AA-genomes using AgriGO (Du, Zhou, Ling, Zhang, & Su, 2010). It is intriguing that a number of GO terms related to developmental processes, such as developmental process (GO: 0032502), reproductive structure development (GO: 0048608), flower development (GO: 0009908), postembryonic development (GO: 0007971), and multicellular organisms development (GO: 0007275), were remarkably enriched in *O. longistaminata* (Figure 5). These genomic features may associate with highly valued traits in *O. longistaminata*, such as SI, but functional genomic experiments are further needed.
4 | DISCUSSION

We report and annotate the two draft genomes of *O. longistaminata* and *O. rufipogon* to supplement the currently existing rich rice genome resources. Compared to the formerly reported *O. longistaminata* genome assembly (Zhang et al., 2015), we produced another contiguous genome assembly that will greatly benefit the research community. Although *O. rufipogon* was recently sequenced (Stein et al., 2018), the generation of the typical *O. rufipogon* genome in this study has the advantage to enhance our understanding of the origin, domestication, and genome evolution of Asian cultivated rice. The genome size of *O. rufipogon* was previously estimated to be ~439 Mb (Uozu et al., 1997), which was seemingly overestimated, compared to the estimations of ~383 and 388 Mb using two orthogonal methods in this study and other representative accessions of the species (Miyabayashi, Nonomura, Morishima, & Kurata, 2007). Probably due to the nature of the typical *O. rufipogon* with a fairly high rate of genomic heterozygosity, we obtained a slightly inflated length of genome assembly (~441.41 Mb), which is larger than the published genome assembly (~339 Mb) (Stein et al., 2018), as divergent haplotypes might be individually represented as separate scaffolds or contigs in the assembled genome.

We detected a considerable number of lineage-specific gene families associated with the observed SI, response to biotic and abiotic stresses and the formation of reproductive separation in these two outcrossing wild rice species, which may account for their adaptive evolution under the remarkably diverse natural habitats in Asia, Oceania, and Africa. In comparisons with the six other AA-genome *Oryza* species, we find the rapid evolution of gene families, particularly evidenced by a noticeable fraction showing fast and/or lineage-specific expansions and contractions with the enrichment of a large number of important functional categories, such as the recognition of pollen in particular; this is obviously related to floral traits of outcrossing and SI observed in *O. rufipogon* and *O. longistaminata*, in sharp contrast with the six other mainly selfing rice species. Besides, a substantial portion of lineage-specific gene families exhibited significant expansions enriched in a number of functions related to salt stress, antifungal response, and disease resistance. Comparative analysis on whole-genome level further demonstrates an extraordinary expansion of genes encoding NBS-LRR proteins in RUF and LON, compared with the six other rice species, including SAT, NIV, GLA, BAR, GLU, and MER, which may associate with their abilities of strong adaptation to remarkably changing environments. One may have concerns about the inflation of the *O. rufipogon* genome assembly that may partially include alternative haplotype to impact our downstream data analysis. In this study, we obtained a genome assembly of ~332.08 Mb for *O. longistaminata*, which was estimated to be between ~363 and 392 Mb using two orthogonal methods. Our above-described findings indicate that the genome analyses of these two outcrossing rice exhibit similar trends of the gene family expansion contrary to the six other predominantly selfing rice species. Thus, the impact of inflated genome assembly of *O. rufipogon* should not be fully excluded but may not considerably influence subsequent data analyses of the gene family evolution.

To avoid mis-assembly that may affect the downstream data analysis of CNSs, we aligned all eight AA-genome assemblies and identified stringent orthologous genomic segments across all eight AA-genome *Oryza* species. Our comparative genomic analyses show a fairly rapid evolution of CNSs in these two outcrossing rice species (RUF and LON) relative to those selfing species, indicating a great reduction of genomic variation within CNSs owning to the switch of

---

FIGURE 5 GO annotation of genes related to the rapid evolving regions across AA-genome *Oryza* species. pv1 represents *p* value of *O. sativa* ssp. *japonica*, pv2 represents *p* value of *O. rufipogon*, pv3 represents *p* value of *O. nivara*, pv4 represents *p* value of *O. glaberrima*, pv5 represents *p* value of *O. barthii*, pv6 represents *p* value of *O. glumaepatula*, pv7 represents *p* value of *O. longistaminata*, and pv8 represents *p* value of *O. meridionalis*. The yellow-to-orange and gray scale represent GO terms that are significance and nonsignificance, respectively. GO, gene ontology
Acknowledgments

We thank International Rice Research Institute (Manila, Philippines) for kindly providing rice germplasms. This work was supported by Yunnan Innovation Team Project and Natural Science Foundation of Yunnan (2015FA030) (to L.-Z. G.).

Conflict of Interest

The authors declare no conflict of interest.

Authors’ Contributions

L.G. conceived and designed the study; C.S., C.S., L.Y., S.M., L.Z., and T.Z. contributed to the collection and preparation of the rice samples; H.H. and Y.T. performed flow cytometry experiment; T.Z., J.J., J.J., S.M., and P.X. conducted the library preparation and genome sequencing; W.L., K.L., Q.Z., Y.Z., and Y.Y. performed genome assembly; L.Z. performed RNA preparation; W.L., Y.L., E.X., D.Z., and Y.Z. performed genome annotation; L.G. wrote and revised the manuscript.

Data Availability Statement

The raw sequence data and genome assemblies reported in this paper have been deposited in the Genome Sequence Archive and Genome Warehouse in Big Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers PRJCA002637.

References

Aggarwal, R. K., Brar, D. S., & Khush, G. S. (1997). Two new genomes in the Oryza complex identified on the basis of molecular divergence analysis using total genomic DNA hybridization. *Molecular and General Genetics*, 254, 1–12. https://doi.org/10.1007/s004380050384

Bao, Z., & Eddy, S. R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Research*, 12, 1269–1276.

Binney, E., Clamp, M., & Durbin, R. (2004). GeneWise and genomewise. *Genome Research*, 14, 988–995.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brar, D. S., & Ramos, J. M. (2007). Wild species of Oryza: A rich reservoir of genetic variability for rice improvement. Manila, Philippines: International Rice Research Institute.

Bray, N., & Pachter, L. (2004). MAVID: Constrained ancestral alignment of multiple sequences. *Genome Research*, 14, 693–699.

Cavaler-Smith, T. (1985). The evolution of genome size, London, UK: Wiley.

Chen, N. (2004) Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, 5, 4–10. 11–14.10. 14.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H., & Ohtsubo, E. (2003). Polyphylectic origin of cultivated rice: Based on the intersperser pattern of SINEs. *Molecular Biology and Evolution*, 20, 67–75.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Muller, W. E., Wetter, T., & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research*, 14, 1147–1159.

Cooper, G. M., Stone, E. A., Asimenos, G., Green, E. D., Batzoglou, S., & Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, 15, 901–913.

Copetti, D., Zhang, J., El Baidouri, M., Gao, D., Wang, J., Barghini, E., ... Wing, R. A. (2015). RITE database: A resource database for genome-wide rice genomics and evolutionary biology. BMC Genomics, 16, 538. https://doi.org/10.1186/s12864-015-1762-3

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, 22, 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Dewey, C. N. (2007). Aligning multiple whole genomes with Mecerator and MAVID. *Methods in Molecular Biology*, 395, 221–236.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., ... Liang, C. (2017). Sequencing and de novo assembly of a near complete indica rice genome. *Nature Communications*, 8, 15324. https://doi.org/10.1038/scommes15324

Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, 38, W64–W70. https://doi.org/10.1093/nar/gkq310

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797.

Ellinghaus, D., Kurtz, S., & Willhoft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, 9, 18. https://doi.org/10.1186/1471-2105-9-18

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HHMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, 39, W29–37. https://doi.org/10.1093/nar/gkr367

Fuller, D. Q., Sato, Y.-I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell-Banham, E. J., ... van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological & Anthropological Sciences*, 2, 115–131. https://doi.org/10.1007/s12520-010-0035-y

Gao, L. (2004). Population structure and conservation genetics of wild rice Oryza rufipogon (Poaceae): A region-wide perspective from microsatellite variation. *Molecular Ecology*, 13, 1009–1024.
Jurka, J., & Pethiyagoda, C. (1995). Simple repetitive DNA sequences

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno,

Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice.

Kovach, M. J., Sweeney, M. T., & McCouch, S. R. (2007). New insights into

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., &

IRGSP. (2005). The map-based sequence of the rice genome.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., … Han,

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., … Briggs, S.

Glover, J. D., Reganold, J. P., Bell, L. W., Borevitz, J., Brummer, E. C.,

Hannick, L. I., … White, O. (2003). Improving the Arabidopsis genome

Liu, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., … Wang, J. (2010). De

Lin, S. C., & Yuan, L. P. (1980). Hybrid rice breeding in China. In Innovative

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., … Wang, J. (2012).

Luo, T. M., & Eddy, S. R. (1997). tRNAscan-SE: A program for improved
detection of transfer RNA genes in genomic sequence. Nucleic Acids

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. Science, 252, 1162-1164. https://doi.org/10.1126/science.252.5009.1162

Majoros, W. H., Pertea, M., & Salzberg, S. L. (2004). TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. Bioinformatics, 20, 2878-2879. https://doi.org/10.1093/bioinformatics/bth315

McCarthy, E. M., & McDonald, J. F. (2003). LTR_STRUC: A novel search and identification program for LTR retrotransposons. Bioinformatics, 19, 362–367. https://doi.org/10.1093/bioinformatics/btf878

Miyabayashi, T., Nonomura, K. I., Morishima, H., & Kurata N. (2007). Genome Size of Twenty Wild Species of Oryza Determined by Flow Cytometric and Chromosome Analyses. Breeding Science, 57(1), 73-78. https://doi.org/10.1270/jsbbs.p773

Morishima, H., Sano, Y., & Oka, H. I. (1992). Evolutionary studies in cultivated rice and its wild relatives. Taxon, 8, 135–184.

Nawrocki, E. P., Kolbe, D. L., & Eddy, S. R. (2009). Infernal 1.0: Inferrence of RNA alignments. Bioinformatics, 25, 1335. https://doi.org/10.1093/bioinformatics/btp482

Nayar, N. M. (1973). Origin and cytogenetics of rice. Advances in Genetics, 17, 153–292.

Oka, H. I. (1988). Origin of cultivated rice. Amsterdam: Japan Science Social Press.

Ou, S., & Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiology, 176, 1410–1422. https://doi.org/10.1104/pp.17.01310

Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Molecular Biology Reporter, 15, 8–15.

Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. Bioinformatics, 21(Suppl 1), i351–i358. https://doi.org/10.1093/bioinformatics/bti1018
She, R., Chu, J. S., Wang, K., Pei, J., & Chen, N. (2009). GenBlastA: Enabling BLAST to identify homologous gene sequences. Genome Research, 19, 143–149.

Shi, C., Li, W., Zhang, Q. J., Zhang, Y., Tong, Y., Li, K., ... Gao, L. Z. (2020). The draft genome sequence of an upland wild rice species, Oryza Granulata. Scientific Data, 7, 131. https://doi.org/10.1038/s41597-020-0470-2

Smit, A., Hubley, R., & Green, P. (2016) RepeatMasker Open-4.0. 2015. Google Scholar.

Song, W.-Y., Wang, G.-L., Chen, L.-L., Kim, H.-S., Pi, L.-Y., Holsten, T., ... Hu, F., Xie, M., ... Wang, W. (2011). Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. Proceedings of the National Academy of Sciences of the United States of America, 113, E5163–5171.

Zhang, Q. J., Zhu, T., Xia, E. H., Shi, C., Liu, Y. L., Zhang, Y., ... Gao, L. Z. (2014). Rapid diversification of five Oryza AA genomes associated with rice adaptation. Proceedings of the National Academy of Sciences of the United States of America, 111, E4954–E4962.

Zhang, Y., Zhang, S., Liu, H., Fu, B., Li, L., Xie, M., ... Hu, F. (2015). Genome and comparative transcriptomics of African wild rice Oryza longistaminata provide insights into molecular mechanism of rhizomatous- ness and self-incompatibility. Molecular Plant, 8, 1683–1686. https://doi.org/10.1016/j.molp.2015.08.006

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences, 13, 555–556.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., ... Yang, H. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science, 296, 79–92. https://doi.org/10.1126/science.1068037

Zhang, J., Chen, L. L., Xing, F., Kudrna, D. A., Yao, W., Copetti, D., ... Xie, W. (2016). Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. Proceedings of the National Academy of Sciences of the United States of America, 113, E5163–5171.

Zhang, Q. J., Zhu, T., Xia, E. H., Shi, C., Liu, Y. L., Zhang, Y., ... Gao, L. Z. (2014). Rapid diversification of five Oryza AA genomes associated with rice adaptation. Proceedings of the National Academy of Sciences of the United States of America, 111, E4954–E4962.

Zhang, Y., Zhang, S., Liu, H., Fu, B., Li, L., Xie, M., ... Hu, F. (2015). Genome and comparative transcriptomics of African wild rice Oryza longistaminata provide insights into molecular mechanism of rhizomatous- ness and self-incompatibility. Molecular Plant, 8, 1683–1686. https://doi.org/10.1016/j.molp.2015.08.006

Zhu, T., Xu, P. Z., Liu, J. P., Peng, S., Mo, X. C., & Gao, L. Z. (2014). Phylogenetic relationships and genome divergence among the AA-genome species of the genus Oryza as revealed by 53 nuclear genes and 16 intergenic regions. Molecular Phylogenetics and Evolution, 70, 348–361. https://doi.org/10.1016/j.ympev.2013.10.008

How to cite this article: LiW, Zhang Q-J, Zhu T, et al. Draft genomes of two outcrossing wild rice, Oryza rufipagone and O. longistaminata, reveal genomic features associated with mating-system evolution. Plant Direct. 2020;00:1–12. https://doi.org/10.1002/pld3.232