Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

Molecular Plant
Research Article

CellPress
Partner Journal

Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding

Zhoulin Gu[1,2], Zhou Zhu[1,2], Zhen Li[1,3], Qilin Zhan[1], Qi Feng[1], Congcong Zhou[1], Qiang Zhao[1], Yan Zhao[1], Xiaojian Peng[1,4], Bingxin Dai[1,5], Rongrong Sun[1,2], Yan Li[1], Hengyun Lu[1], Lei Zhang[1], Tao Huang[1], Junyi Gong[6], Danfeng Lv[1], Xuehui Huang[7] and Bin Han[1,*]

[1]National Center for Gene Research, State Key Laboratory of Plant Molecular Genetics, Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200233, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]College of Life Sciences, Anhui Normal University, Wuhu, Anhui 241000, China

[4]School of Life Sciences, Anhui Agricultural University, Hefei, Anhui 230036, China

[5]School of Life Science and Technology, Shanghai Tech University, Shanghai 201210, China

[6]State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China

[7]College of Life Sciences, Shanghai Normal University, Shanghai 200234, China

***Correspondence: Bin Han (bhan@ncgr.ac.cn)**

https://doi.org/10.1016/j.molp.2021.08.007

ABSTRACT

The male sterility (MS) line is a prerequisite for efficient production of hybrid seeds in rice, a self-pollinating species. MS line breeding is pivotal for hybrid rice improvement. Understanding the historical breeding trajectory will help to improve hybrid rice breeding strategies. Maternally inherited cytoplasm is an appropriate tool for phylogenetic reconstruction and pedigree tracing in rice hybrids. In this study, we analyzed the cytoplasmic genomes of 1495 elite hybrid rice varieties and identified five major types of cytoplasm, which correspond to different hybrid production systems. As the cytoplasm donors for hybrids, 461 MS lines were also divided into five major types based on cytoplasmic and nuclear genomic architecture. Specific core accessions cooperating with different fertility-associated genes drove the sequence divergence of MS lines. Dozens to hundreds of convergent and divergent selective sweeps spanning several agronomic trait-associated genes were identified among different types of MS lines. We further analyzed the cross patterns between different types of MS lines and their corresponding restorers. This study systematically analyzed the cytoplasmic genomes of rice hybrids revealed their relationships with nuclear genomes of MS lines, and illustrated the trajectory of hybrid rice breeding and the strategies for breeding different types of MS lines providing new insights for future improvement of hybrid rice.

Key words: rice hybrid, cytoplasm, cytoplasmic male sterile, environment-sensitive genic male sterile, breeding

Gu Z., Zhu Z., Li Z., Zhan Q., Feng Q., Zhou C., Zhao Q., Zhao Y., Peng X., Dai B., Sun R., Li Y., Lu H., Zhang L., Huang T., Gong J., Lv D., Huang X., and Han B. (2021). Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding. Mol. Plant. **14**, 1–16.

INTRODUCTION

Heterosis, or hybrid vigor, is the phenomenon in which increases in yield or other agronomic traits are observed in a hybrid offspring relative to its inbred parental lines. Exploitation of heterosis in crops has achieved significant yield advantages over traditional inbred line breeding. To date, hybrid breeding that combines elite alleles from both parental lines to generate a better first filial variety remains one of the most efficient breeding approaches in rice and many other crops (Huang et al., 2016). Because cultivated rice is a self-pollinating plant or an inbreeding crop, male sterile (MS) rice varieties, which are unable to produce functional pollen grains and are used as female parents to avoid self-pollination, are a prerequisite for efficiently producing large quantities of hybrid seeds. The exploitation of rice MS lines for commercial use began in the 1970s (Li and Yuan, 2000). Two

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

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

hybrid seed production systems are widely used, the two-line system and the three-line system (Cheng et al., 2007; Qian et al., 2016; Fan and Zhang, 2018; Kim and Zhang, 2018). The two-line system comprises environment-sensitive genic male sterile (EGMS) and restorer lines. Pollen abortion in the EGMS line depends on environmental conditions. Under restrictive conditions, for example, the female line is MS under long-day growing conditions; however, under permissive conditions, the female parent is fertile. The three-line system consists of a cytoplasmic male sterility (CMS) line, a restorer, and a maintainer. The CMS line is unable to produce normal pollen, and the phenomenon is caused by a CMS-associated gene located in the mitochondrial genome. The reproduction of the CMS line relies on the maintainer line. Various types of CMS lines have been developed for hybrid seed production on a commercial scale, including wild abortive (WA), Indonesian Shuitiangu (ID), K, Gambiaca (GA), Dissi (D), dwarf abortive (DA), Maxie, Y, Honglian (HL), Boro II (BT), Dian I, and others (Luan et al., 2013). The pollen disruption mechanism has been well elucidated in some types of CMS lines. The mitochondrial gene *WA352c* is reported to confer CMS on the WA, ID, K, GA, D, DA, and Y CMS lines (Luo et al., 2013). Protein WA352c interacts with COX11 to destroy its function and triggers premature tapetal programmed cell death. The mitochondrial gene *orf79* is responsible for BT and Dian I CMS (Wang et al., 2006; Luan et al., 2013), and *orfH79* is the causative gene for HL CMS (Wang et al., 2013). There are only five base changes in coding regions between *orf79* and *orfH79*, and four of them lead to nonsynonymous mutations (Yi et al., 2002). Protein orfH79 combines with the P61 subunit to disrupt the mitochondrial electron transport chain and triggers a reactive oxygen species burst in the mitochondrion. In the nucleus, there are corresponding restoring genes that reduce the levels of toxic products to ensure normal pollen development. The pentatricopeptide repeat family protein Rf4 can degrade the transcripts of *WA352c* (Tang et al., 2014). *Rf3* also has restoring ability for WA CMS lines (Cai et al., 2013). *Rf1a* cleaves and *Rf1b* degrades transcripts of *atp6–orf79*, and both of them are used as restoring genes for BT CMS (Wang et al., 2006). *Rf1a* (also named *Rf5*) and *Rf6*, the HL CMS restoring gene, process aberrant *atp6–orfH79* transcripts by cleaving at specific loci (Huang et al., 2015a). Thus, interactions between cytoplasm and nucleus play an important role in heterosis exploitation. Based on the nucleus, numerous studies have been devoted to depicting the genomic architecture of rice heterosis (Huang et al., 2015b, 2016; Lin et al., 2020; Lv et al., 2020); however, we still lack a comprehensive understanding of the hybrid cytoplasm, as well as the relationship between the cytoplasm and the corresponding nucleus for hybrids and MS lines.

In traditional hybrid rice breeding, massive testcrosses between MS lines and elite rice inbred lines are conducted by breeders, and several excellent combinations are subsequently selected based on field phenotypic investigation. After decades of development, breeders have bred excellent MS lines and screened out a large number of extraordinary hybrid combinations. Based on valuable information provided by traditional breeding, it is feasible to summarize hybrid rice breeding rules and inform solutions for future breeding improvement. In this study, to elucidate the hybrid rice breeding trajectory, we focused on the cytoplasm because it possesses CMS genes and is suitable for pedigree tracing owing to its maternal inheritance. We analyzed the cytoplasms of 1495 hybrids, classified them into five major types, and further summarized the genomic features of each cytoplasm type. As the donors of cytoplasm to hybrids, MS lines were also divided into five major groups based on both cytoplasmic and nuclear genome variations. By identifying improving selection signals of MS lines and analyzing cross patterns between MS lines and restorers, we summarize the breeding strategies for different types of MS lines (Supplemental Figure 1).

RESULTS

Five major types of organellar genomes in 1495 rice hybrids

In our previous work, we collected and sequenced 1495 diverse rice hybrids (Huang et al., 2015b). Low-coverage whole-genome sequencing or whole-genome shotgun sequencing data (~2×) were aligned against the updated cytoplasmic genome sequence to identify single-nucleotide polymorphisms (SNPs). For the mitochondrial genome, we identified 446 high-quality SNPs among the 1495 accessions, and 77 high-quality SNPs were found in the chloroplast genome (Supplemental Data 1). Based on SNP data, the sequence diversity (π) of mtDNA was estimated at ~0.000101, which was approximately 24 times lower than that of nuclear DNA in *Oryza sativa* (Huang et al., 2012). As for chloroplast DNA (cpDNA), the sequence diversity in the 1495 hybrids was estimated at 0.000144, slightly higher than that of mtDNA, but still 16 times lower than that of *O. sativa* nuclear DNA. We investigated the population structure of rice hybrids using cytoplasmic polymorphisms. According to a neighbor-joining tree and principal component analysis (PCA) of mitochondrial SNPs, nearly all accessions could be divided into five major clades, consisting of 1115, 242, 77, 50, and 11 hybrids, respectively (Figure 1A–1C). Group 1, group 4, and group 5 hybrids were from the three-line hybrid production system, whereas group 2 and group 3 were bred from the two-line production system. Thus, hybrids bred from three-line and two-line systems contain different types of cytoplasms. This classification result was consistent with that based on chloroplast SNPs (Supplemental Figure 2B–2D). Hereafter, cytoplasms from group 1 to group 5 hybrids are referred to as group 1 to group 5 cytoplasms.

The hybrids inherited organelle and CMS-associated genes from their MS line. We therefore analyzed the distribution of three well-studied CMS genes (*WA352c*, *orf79*, and *orfH79*) among the different types of mitochondrial genomes. *WA352c*, *orf79*, and *orfH79* were cloned from WA, BT, and HL CMS lines, respectively (Wang et al., 2006, 2013; Luo et al., 2013). Approximately 99% of group 1 cytoplasms hold *WA352c*, and *WA352c* was also uniquely held by group 1 cytoplasms (Figure 1D). One hundred percent of group 4 mitochondrial genomes possessed *orf79*. As for group 5, 100% of cytoplasms contained *orfH79*. None of EGMS-derived group 2 and group 3 cytoplasms had the aforementioned CMS-associated genes. Thus, group 1, 4, and 5 cytoplasms were derived from *WA352c*-based, *orf79*-based, and *orfH79*-based three-line systems, respectively. Eleven types of CMS lines classified by agronomic production contain the three groups of cytoplasmic resources and express the three CMS genes (Figure 1E; Supplemental Data 2). WA and ID CMS lines were the

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

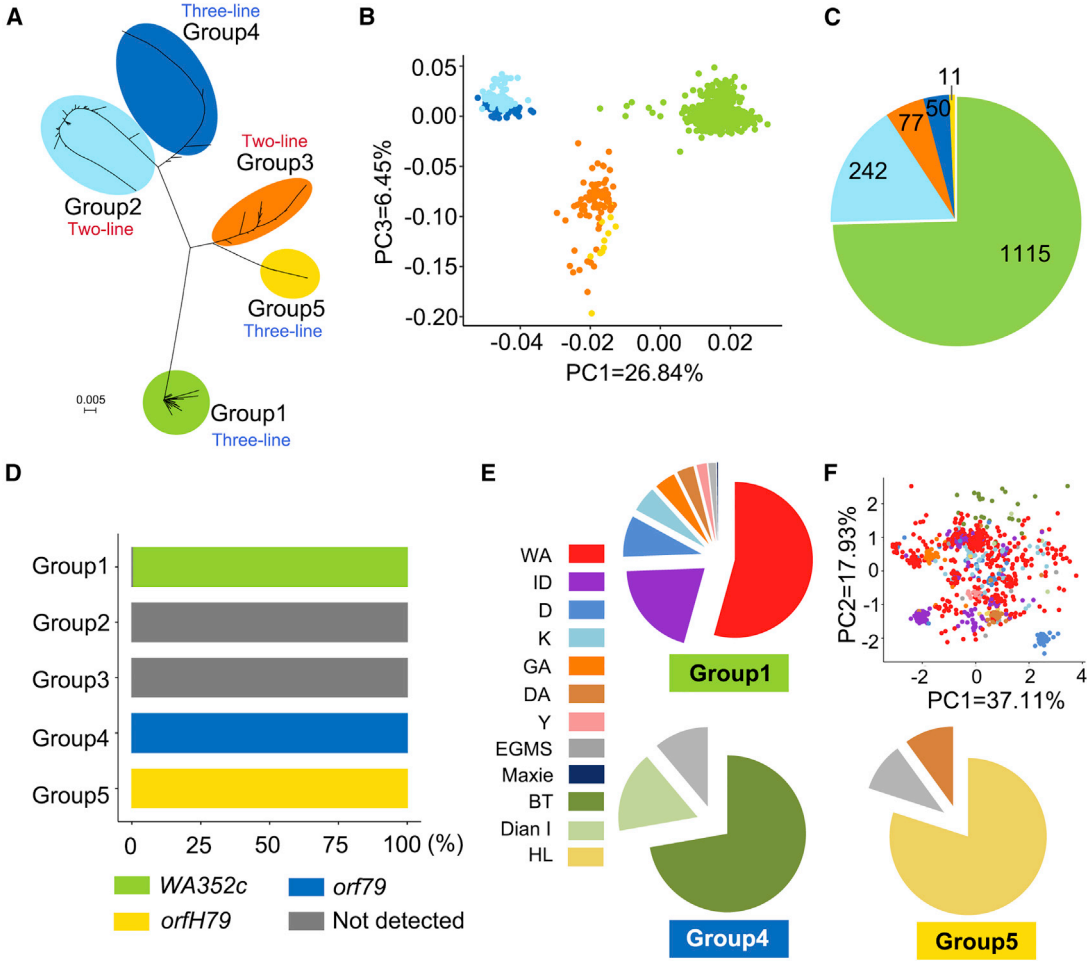

Figure 1. Population structure analysis based on hybrid mitochondrial genomes.

(A and B) SNPs from rice hybrid mitochondrial genomes were used to construct a neighbor-joining tree and perform PCA. Five major types of mitochondrial genomes were identified and marked as groups 1–5.

(C) The population sizes of hybrids containing each type of mitochondrial genome.

(D) The distribution of three known cytoplasmic male sterile causative genes (*WA352c*, *orf79*, and *orfH79*) among five types of mitochondrial genome. "Not detected" means that none of the three known genes were detected.

(E) The distribution of three types of mitochondrial genomes and known CMS-associated genes among 11 types of CMS lines classified by agronomic production.

(F) PCA based on predicted nuclear polymorphisms from 1106 CMS lines that correspond to 1106 hybrids containing group 1, 4, and 5 mitochondrial genomes. Colored dots indicate the types of male sterile lines in **(E)**.

most widely used maternal lines for the three-line system, and ~75% of group 1 hybrids were bred from them. We next conducted PCA for CMS lines corresponding to 1106 three-line hybrids based on predicted nuclear polymorphisms reported previously (Huang et al., 2015b). We did not detect nucleus differentiation for the eight CMS lines that contained group 1 cytoplasm, except for a small proportion of D CMS lines (Figure 1F). As for the two types of cytoplasm in two-line hybrids, group 2 was *japonica* and group 3 was *indica* EGMS cytoplasm. Based on the pedigree information recorded in the China Rice Data Center (https://www.ricedata.cn/variety/index.htm), we found that the EGMS lines Peiai64S Guangzhan63S and their derived lines were widely used as the female parents for group 2 hybrids (Supplemental Data 2). Nongken58S, the *japonica* EGMS line, was the common ancestor of these EGMS lines, and we inferred that it was one of the original cytoplasm donors for group 2 cytoplasm. By tracing the pedigree of the group 3 hybrids, we observed that most of these hybrids were bred from Zhu1S, Lu18S, or their derived lines. Kangluozao, the traditional *indica*-type inbred line, was the common ancestor and cytoplasm donor for group 3 cytoplasm. Therefore, we speculated that organelle differentiation in the ancestor germplasm led to differentiation of group 2 and group 3 cytoplasms.

To rapidly distinguish the type of organelle possessed by hybrids (Supplemental Figure 3A), we selected 10 highly differentiated variations from five groups of mitochondrial and chloroplast genomes (Supplemental Figure 3B and 3D). Marker-based classification based on the 10 representative molecular markers (Supplemental Figure 3C and 3E) was highly consistent with phylogenetic analysis results based on whole-genome SNPs (Figure 1A–1C; Supplemental Data 2).

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

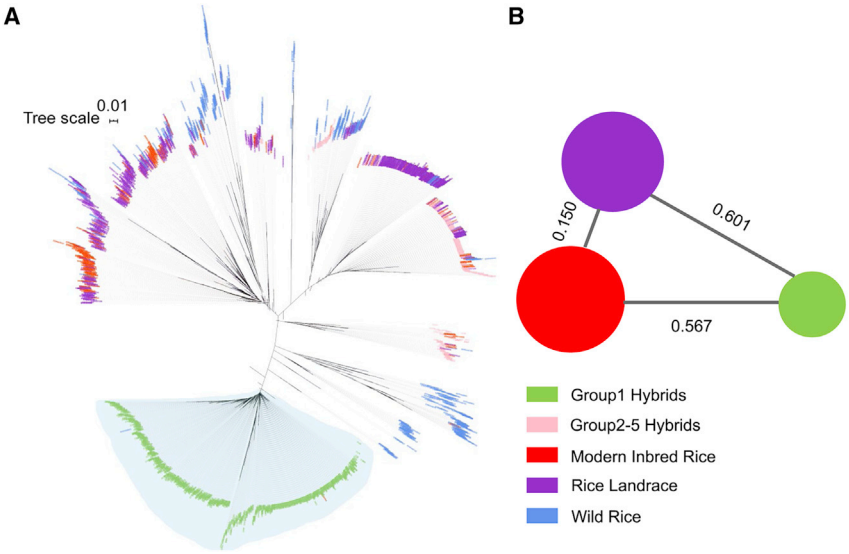

Figure 2. The mitochondria possessed by group1 hybrid varieties differed from that in inbred rice

(A) SNPs from mitochondrial genomes were detected and used to construct a neighbor-joining tree for a panel of populations, including 461 *O. rufipogon*, 620 landraces (*O. sativa*), 205 modern inbred rice varieties (*O. sativa*), and 1495 hybrid varieties (*O. sativa*). Hybrids that contained group 1 mitochondria did not cluster with landraces or modern inbred rice.

(B) Nucleotide genetic diversity (π) was calculated for mitochondrial genomes of group 1 hybrids (green), landraces (purple), and modern inbred rice cultivars (red). The size of the circles represents the level of genetic diversity. The level of genetic differentiation (F_{ST}) between populations is also shown.

Exploitation of *WA352c* during hybrid breeding has broadened the diversity of mitochondrial genomes in hybrid rice

CMS resources have extensive origins. For example, the WA CMS resource was discovered in naturally abortive wild rice, the HL CMS line inherited its organelle from red-awned wild rice (*Oryza rufipogon*), and the BT CMS line obtained its organelle from the Indian *indica* accession Chinsuran Boro II. Therefore, to evaluate the genetic diversity of hybrid cytoplasms on a large scale, we used 832 high-quality common SNPs in the mitochondrial genome from a panel of accessions that contained 461 *O. rufipogon* accessions, 620 landraces (Huang et al., 2012), 205 modern inbred rice varieties (XI-1B) (Wang et al., 2018), and 1495 hybrid rice varieties to explore their phylogenetic relationships. As shown in Figure 2A, a cluster consisting of hybrids bred from the *WA352c*-based CMS system (group 1) did not include modern inbred rice or landraces and was genetically divergent from other *O. sativa* accessions. Furthermore, we estimated that the F_{ST} (population differentiation level) between group 1 hybrids and modern inbred rice was 0.567, and it was 0.601 between group 1 hybrids and landraces (Figure 2B). It is well known that the WA CMS line was created by replacing the nucleus of a naturally abortive wild rice by backcrossing with an inbred cultivar. The WA CMS line retained the cytoplasm of the abortive wild rice. These results suggest that the exploitation of naturally abortive wild rice has brought a wild-type cytoplasm into cultivated hybrid rice and broadened cytoplasm genetic diversity.

Assemblies of five representative organelle genomes and their sequence structure analysis

To further evaluate polymorphisms and structural variations in hybrids' organelle genomes, we selected five representative hybrids from each group to perform reference-based assembly (Supplemental Figure 4A and 4B, Supplemental Table 1). Their breeding paths and organelle origins were also investigated based on records from the China Rice Data Center (Supplemental Figure 5). We adopted a hybrid assembly strategy using both whole-genome shotgun sequencing Pacific Biosciences (PacBio)/Nanopore long reads and Illumina short reads (Supplemental Figure 6). The accuracy of the hybrid assembly strategy was also evaluated (Supplemental Figure 7; Supplemental Table 3). Compared with strategies that use only next-generation sequencing reads, a hybrid assembly strategy can simultaneously retain accuracy and improve assembly continuity. Thus, we obtained organelle genome sequences of five representative accessions using the hybrid assembly pipeline. Because a hybrid assembly usually contains one copy of large repeats in the mitochondrial genome, the total lengths of the five hybrid mtDNA assemblies were approximately 350 660 bp to 364 666 bp (Supplemental Tables 4 and 5). For cpDNA, we obtained complete circular sequences for the five varieties, ranging from 134 501 bp to 134 912 bp. We could see that the sequence structure of hybrid mtDNA was variable, but that of cpDNA was conserved (Supplemental Figure 4C). Our observation in rice hybrids was consistent with previous reports on plant organelle genomes (Palmer and Herbon, 1988; Tian et al., 2006; Gualberto et al., 2014): recombination occurred at a high frequency, and genomic structure of mtDNA was variable. In addition, when we performed reference-based assembly, we detected two potential mis-assemblies (Supplemental Figure 8) in the Nipponbare mitochondrial reference sequence BA000029.3 (Notsu et al., 2002); these mis-assemblies have also been reported previously (Du et al., 2017). A 1119-bp insertion at 51 835 bp was confirmed by PacBio reads and polymerase chain reaction (PCR) (Supplemental Figure 8A–8C). A misplaced sequence from 467 565 bp to 490 520 bp was also detected (Supplemental Figure 8D and 8E). We therefore performed *de novo* assembly to update the Nipponbare organelle reference sequence (Supplemental Figure 9). The updated mtDNA was 530.03 kb in length, 39.5 kb longer than the previous reference sequence version, and had a linear structure. The major difference was the presence of two duplicated sequences that began at base 467 565 and extended to the end of the mtDNA (Supplementary Figure 10A and Supplementary Table 4). The updated mtDNA was confirmed by PacBio/Nanopore reads and Sanger sequencing (Supplemental Figure 10C–10F). The updated Nipponbare mtDNA also shared high collinearity with mtDNA of the variety Shuhui (Du et al., 2017) (Supplemental

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

Figure 10B). There were no differences in genomic structure between the newly assembled cpDNA and the published reference sequence (NC_001320.1). However, we did detect several mismatches, which were confirmed using short reads from Nipponbare (Supplemental Data 3). The organelle reference sequences (mtDNA and cpDNA) used in our work were the updated versions.

DNA polymorphisms found in five representative organelle genomes

We anchored organelle genome assemblies from five hybrids onto the updated mitochondrial/chloroplast reference sequences and identified sequence variations. In total, we identified 428 variations in mtDNA (Figure 3A and 3C; Supplemental Data 4), including 261 SNPs, 66 insertions and deletions (indels), and 101 nonhomologous sequences (NHSs). Twelve polymorphisms in protein-coding genes were selected for verification by Sanger sequencing, and all of them were confirmed (Supplemental Figure 11). We paid more attention to polymorphisms located in gene coding regions. According to annotation information for the updated Nipponbare mtDNA, there were 123 genes, including 46 coding genes for known protein products, 37 putative open reading frames, 34 tRNAs, and 6 ribosomal RNAs. There were 10 SNPs, 7 NHSs, and 3 deletions (dels) located in eight genes among the five accessions (Figure 3D and 3G). We evaluated the potential effects of these variations on coding proteins (Figure 3G). *Cox3* encoded cytochrome *c* oxidase subunit 3 and possessed three SNPs; all three mutations were located in its "four-helical bundle domain" predicted by InterProScan (Jones et al., 2014). One of them resulted in a serine to leucine change (Supplemental Figure 12B). In addition, according to the PROVEAN score (Choi and Chan, 2015), two in-frame indels in two hypothetical proteins, orf224 and orf176, were predicted to be deleterious (PROVEAN score greater than −2.5) (Figure 3G).

In chloroplast genomes, we characterized 106 variations, including 60 SNPs, 33 indels, and 13 NHSs (Figure 3B and 3E; Supplemental Data 4). An inverse repeat (IR) sequence was conserved (Figure 3B), and the variation density was only 0.096 per kilobase for the IR sequence, compared with 0.788 per kilobase for the whole chloroplast genome. Several genes, including *matK*, *psbZ*, *rpoC2*, *rpl20*, *psbB*, *rpl16*, and *ccsA*, contained nonsynonymous mutations. Among them, only substitutions in *matK* and *psbZ* were predicted to be deleterious by PROVEAN (Figure 3H), and both of them were specific to the group 4 representative variety 10You18. *MatK* encoded maturase K, and a missense mutation was located in its N-terminal domain. This mutation resulted in the substitution of the alkaline amino acid arginine with the nonpolar amino acid glycine (Supplemental Figure 12C). *PsbZ* encoded photosystem II protein Z, and its substitution replaced the nonpolar amino acid alanine with the polar amino acid threonine.

Cytoplasm genome associated with nuclear genome in rice MS lines

After summarizing cytoplasm characteristics, we performed a population structure analysis for MS lines based on both nuclear and cytoplasmic genome variations and dissected the relationships between them. We sequenced a collection of 156 rice MS lines with an average depth of 30×. Integrating with the genome sequencing data from 305 MS lines reported by Lv et al. (Lv et al., 2020), we identified 4 821 211 high-quality SNPs. According to the previously defined cytoplasm types, 270 lines contained group 1 organelles, 128 lines had group 2 organelles, and 41 lines possessed group 3 organelles; 11 lines had group 4 cytoplasm, and 11 lines had group 5 cytoplasm (hereafter referred to as group 1, group 2, group 3, group 4, and group 5 MS lines for convenience) (Figure 4A). Because of the relatively limited exploitation of *orf79* and *orfH79* CMS resources for commercial seed production, we collected only 11 accessions that contained group 4 or group 5 cytoplasm. We performed phylogenetic analysis and PCA based on nuclear SNPs and revealed that MS lines with the five types of cytoplasm could be separated, but with clear overlaps (Figure 4B and 4C). Admixture analysis based on nuclear SNPs showed that group 1 CMS lines were genetically distinct from group 2 EGMS lines, whereas the majority of group 3 EGMS lines had both group 1 and group 2 feature components (Figure 4D). Group 4 CMS lines could also be distinguished from the other four types of MS line based on phylogenetic analysis. Our results indicated that rice MS lines with different cytoplasms had genetically divergent nuclei. According to the findings in Figure 1D, different types of cytoplasm contained different CMS-associated genes; thus, rice MS lines that used different MS-associated genes were genetically divergent. Next, we performed kinship analysis for female lines and identified 2968 comparisons (of 108 812 pairs) with kinship coefficients greater than 0.92, involving 381 female lines (Figure 4E). All comparisons formed 23 multiple-member clusters, and the top three clusters accounted for 96.9% of all comparisons. Most of the accessions that contained the same type of organelle had a close relationship and were located in the same network, with key sterile lines as network hubs. For example, the WA-CMS line Zhenshan97A was related to 67 accessions that contained group 1 organelles, and it was the center of a network comprising group 1 varieties (Figure 4E). The WA-CMS lines V20A and Jin23A were also the hub of the same network of group 1 accessions, whereas the EGMS lines Guangzhan63S and XinanS were related to 38 and 37 varieties with group 2 cytoplasm, respectively. Both of them were kernel lines for the cluster that consisted of group 2 EGMS lines. We speculated that several backbone strains determined the genomic architecture of rice MS lines during the hybrid rice breeding process. Specific kernel accessions and derived lines were widely used as founder parents and providers of sterility-associated genes for different hybrid production systems. Thus, MS lines from the same sterility maintenance and fertility restoration system had close genetic relationships.

Comprehensively considering their nuclear and cytoplasmic polymorphisms, we classified the rice MS lines into five major types (Supplemental Table 7).

Differentiation of the flanking regions of fertility-associated genes in different types of MS lines

Because different types of rice MS lines belonged to different breeding systems and expressed specific fertility-associated genes, we sought to investigate linkage drags and sequence differentiation brought about by fertility-associated genes in the nucleus. The majority of group 2 and group 3 EGMS lines

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

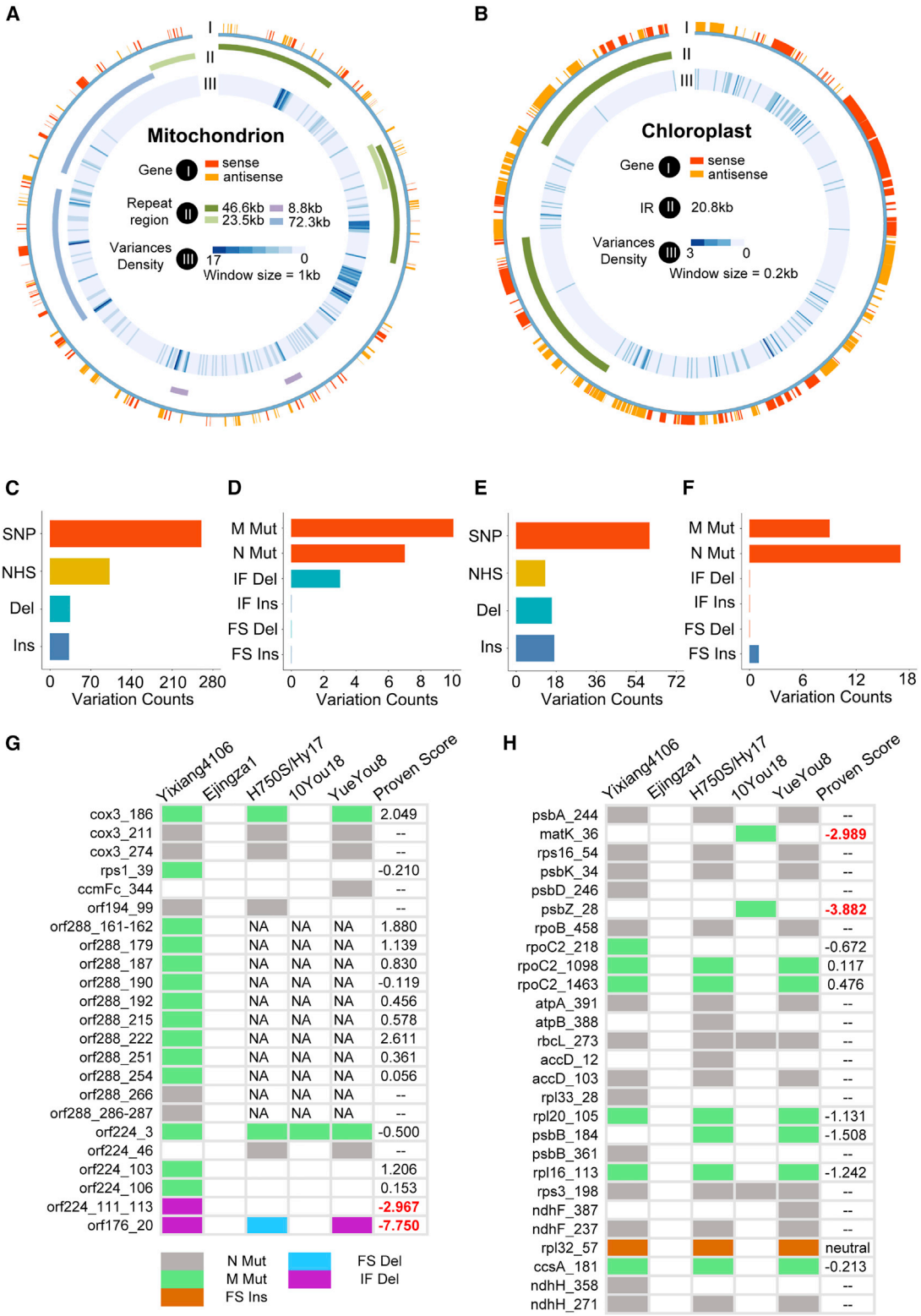

Figure 3. Variations in five representative organellar genomes.

(A and B) Variations (including SNPs, NHSs, and indels) were distributed across the whole genome of the mitochondrion **(A)** and chloroplast **(B)**. From outer to inner circle: annotation information, major repeat sequences, and variant density.

(C–F) Counts for different types of variations across the whole genome **(C and E)** and the coding regions **(D and F)** for mitochondrion **(C and D)** and chloroplast **(E and F)** genomes. Del, deletion; Ins, insertion; N Mut, nonsense mutation; M Mut, missense mutation; FS Ins, frameshift insertion; IF Ins, in-frame insertion; FS Del, frameshift deletion; IF Del, in-frame deletion.

(legend continued on next page)

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

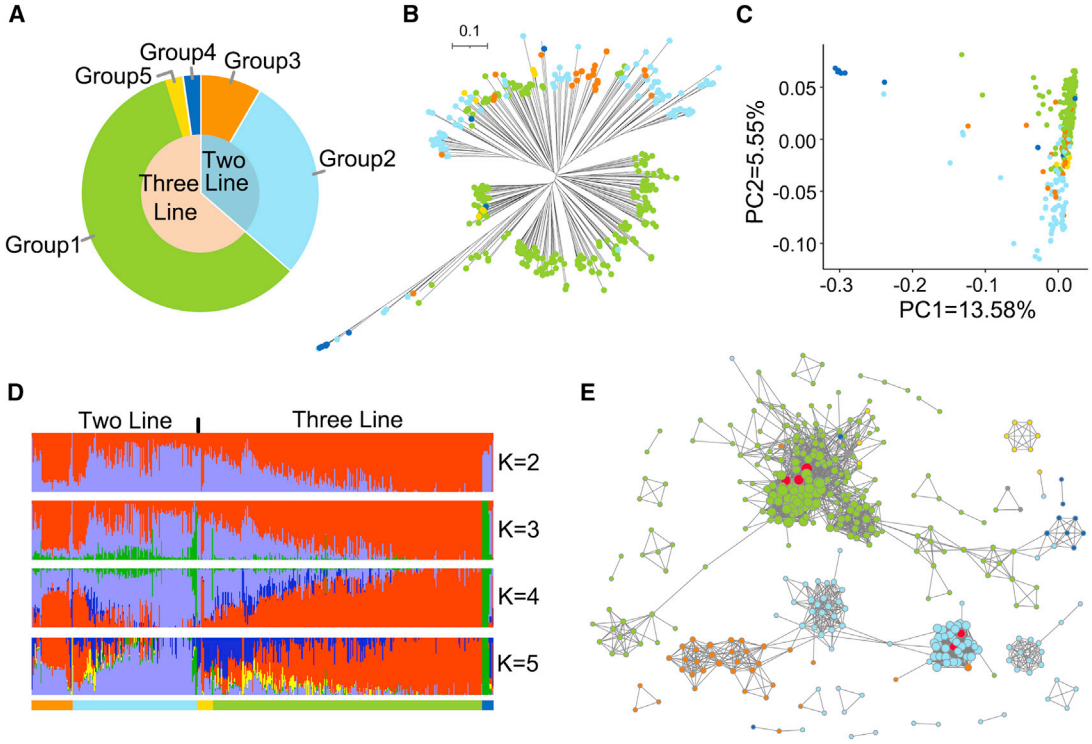

Figure 4. Association between nucleus and cytoplasm of rice MS lines.

(A) Distribution of five types of cytoplasm in 461 female lines.

(B) Neighbor-joining tree constructed from 4 821 211 nuclear SNPs in 461 female lines. Female lines containing different types of organelles are indicated in different colors (groups 1–5 are indicated in green, light blue, orange, dark blue, and yellow, respectively).

(C) PCA plot of the first and second principal components based on nuclear SNPs.

(D) Admixture analysis with different numbers of clusters (K = 2, 3, 4, and 5). The y axis quantifies cluster proportions, and the x axis lists different female accessions ordered by their organelle type. The bottom bar indicates organelle type.

(E) Kinship relationships for 381 female lines related with at least one accession. Circles represent accessions and are filled according to their organelle type, and the size of the circle indicates the count of accessions related to it. Related accessions are connected by lines. Three key wild abortive sterile lines (Zhenshan97A, Jin23A, V20A) and two environment-sensitive genic male sterile lines (Guangzhan63S, XinanS) are highlighted by red dots.

used the TGMS-associated (thermo-sensitive genic male sterility) gene *tms5* (Zhou et al., 2014), and we observed decreased sequence diversity around *tms5* in both group 2 and group 3 MS lines compared with landrace rice (Supplemental Figure 13B and 13E). In addition, *Rf3* and *Rf4* were fertility-restoring genes for group 1 MS lines (Cai et al., 2013; Tang et al., 2014). We observed lower sequence diversity around genes *rf3* and *rf4* (the loss-of-function alleles) in group 1 MS lines compared with landraces (Figure 5A; Supplemental Figure 13A). Furthermore, we noticed that the gene *D2*, which controls plant architecture, was in the flanking region of *rf3* (Figure 5A). A natural mutation in the 3′ UTR of *D2* has been reported to lead to variation in tiller angle and plant height (Dong et al., 2016; Wei et al., 2021). The "G" (*D2*) to "T" (*d2*) substitution resulted in a significantly smaller tiller angle and more erect plant architecture. Thus, *d2* was the favorable allele type for plant architecture improvement. We found that the *D2* allele was highly linked with *rf3* in *O. rufipogon* accessions and rice landraces (Figure 5B). Two hundred seventy-eight of 280 landraces and wild rice accessions contained *rf3* and *D2* simultaneously, whereas only two varieties contained *rf3* and *d2*. Thus, the fixation of allele *rf3* resulted in a high frequency of the *D2* allele in group 1 MS lines (Figure 5B, 5C, and 5F). In group 2 EGMS lines, without the need to fix the *rf3* allele, the frequency of the favorable *d2* allele was as high as 80% (Figure 5F). The cross-population composite likelihood ratio test (XP-CLR) and a population branch statistic (pbs) analysis indicated that the region containing *d2* was under artificial selection (Figure 5D and 5E) in group 2. In addition, the population differentiation level (F_{ST}) around *D2* between group 1 and group 2 MS lines was estimated to be 0.58 (Figure 5G). Therefore, we inferred that linkage drag brought about by fertility-associated genes could limit the exploitation of favorable genes for hybrid rice breeding and could also drive flanking sequence differentiation in different types of MS lines. Interestingly, we observed a complementary pattern for *D2* and its flanking sequence in group 1 MS lines and their corresponding restorer lines. Approximately

(G and H) Potential effects on proteins of variations in the mitochondrion **(G)** and chloroplast **(H)** sequences. Different types of sequence variations are indicated by different colors. Labels at the left of the matrix indicate the position of the mutation. For example, cox3_185 indicates the mutation located at amino acid 185 in the cox3 protein. "NA" in columns 3–5 represents missing data. The PROVEAN score is marked for each missense mutation or indel at the right of the matrix to evaluate its potential effect (cutoff −2.5; score > −2.5, neutral; score < −2.5, deleterious).

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

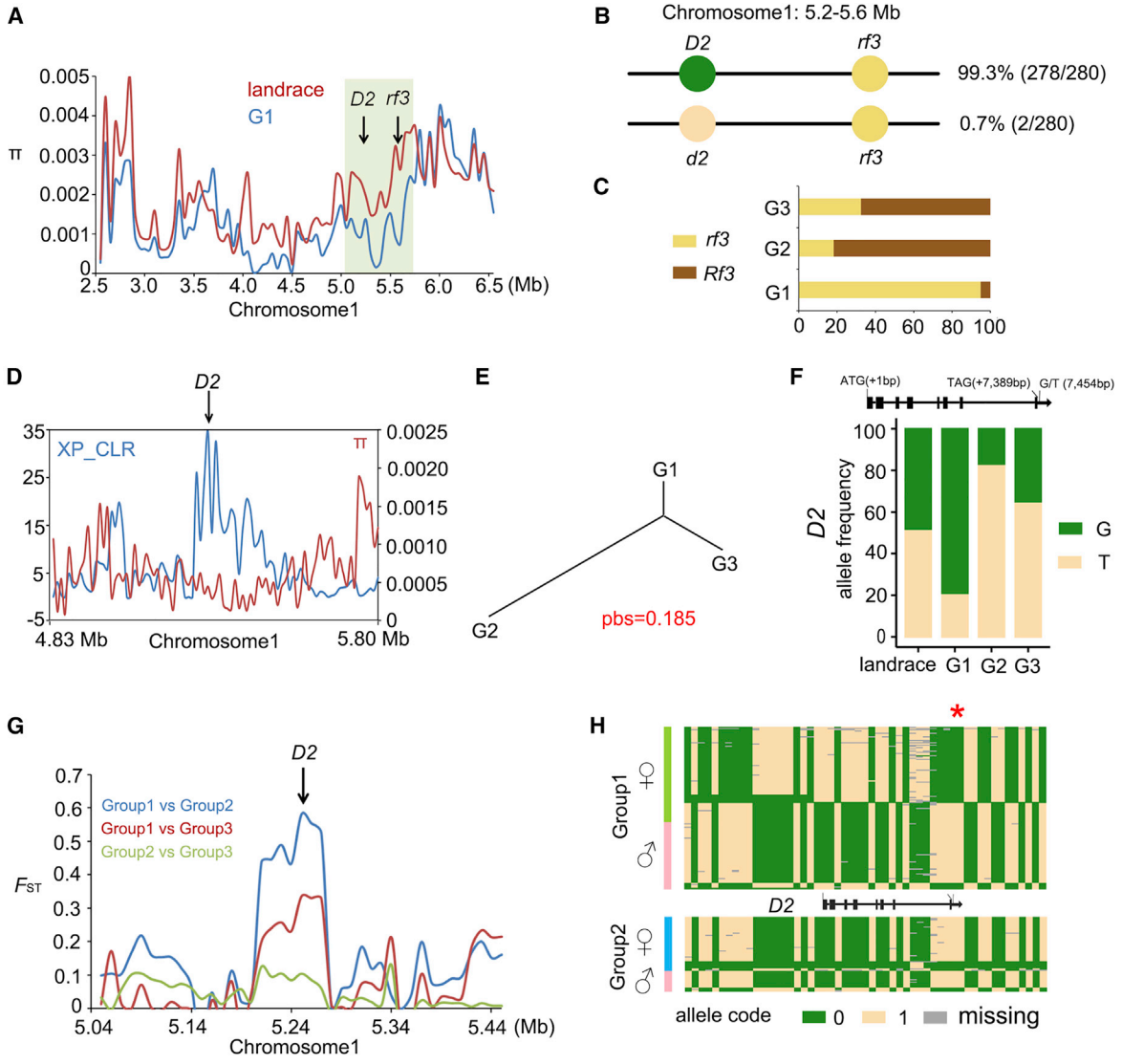

Figure 5. The genetic drags brought to group 1 MS lines by the fertility-associated gene *rf3*.

(A) Nucleotide diversity (π) around the fertility-restoring gene *rf3* was calculated for rice landrace and group 1 male sterile lines.

(B) Distribution of two types of *D2* alleles in 280 *rf3*-containing and 83 *Rf3*-containing accessions (wild rice and landrace). The allele *D2* was closely linked to allele *rf3*.

(C) The distribution of two types of *Rf3* allele in three types of male sterile lines.

(D) XP-CLR value (blue) and nucleotide diversity (red) around gene *D2* were plotted for a group 2 male sterile subpopulation.

(**E**) Tree based on pbs values for the 100-kb region containing gene *D2*. The pbs value for group_2 is estimated as 0.185.

(**F**) The gene structure and allele frequency of two types of alleles for gene *D2* in rice landrace and three female-parent subpopulations.

(G) Genetic differentiation (F_{ST}) for three pairs of MS line comparisons around gene *D2*.

(H) Haplotypes around gene *D2*. The x axis corresponds to 53 high-quality polymorphisms located at 5234 kb to 5254 kb on chromosome 1. The y axis lists samples for group 1 female lines (green) and corresponding male lines (pink), as well as group 2 female lines (blue) and corresponding male lines (pink).

90% of restorers crossing with group 1 female lines contributed *d2* alleles to the hybrids (Figure 5H).

Convergent and divergent selection in three groups of MS lines during genetic improvement breeding

We detected phenotypic variations among three groups of MS lines (group 4 and group 5 MS lines were excluded owing to small population size). Six agronomic traits were differentiated among the three groups of MS lines: total grain number, panicle length, tiller angle, leaf angle, leaf length, and leaf width (Figure 6A). To identify breeding footprints in different MS lines, we calculated the XP-CLR values and the ratio of genetic diversity in landraces to that in MS lines (Figure 6B). We detected 168 selective sites in group 1 MS lines covering 15.81% of the rice genome, 169 in group 2 (18.41%), and 167 in group 3 (17.09%). Sixty-four sites were shared by the three groups. In addition, group 1 contained 60 unique selection sites, group 2 contained 72, and group 3 contained 48 (Figure 6C). Differentiated selective sites had significantly higher sequence-differentiation (F_{ST}) levels

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

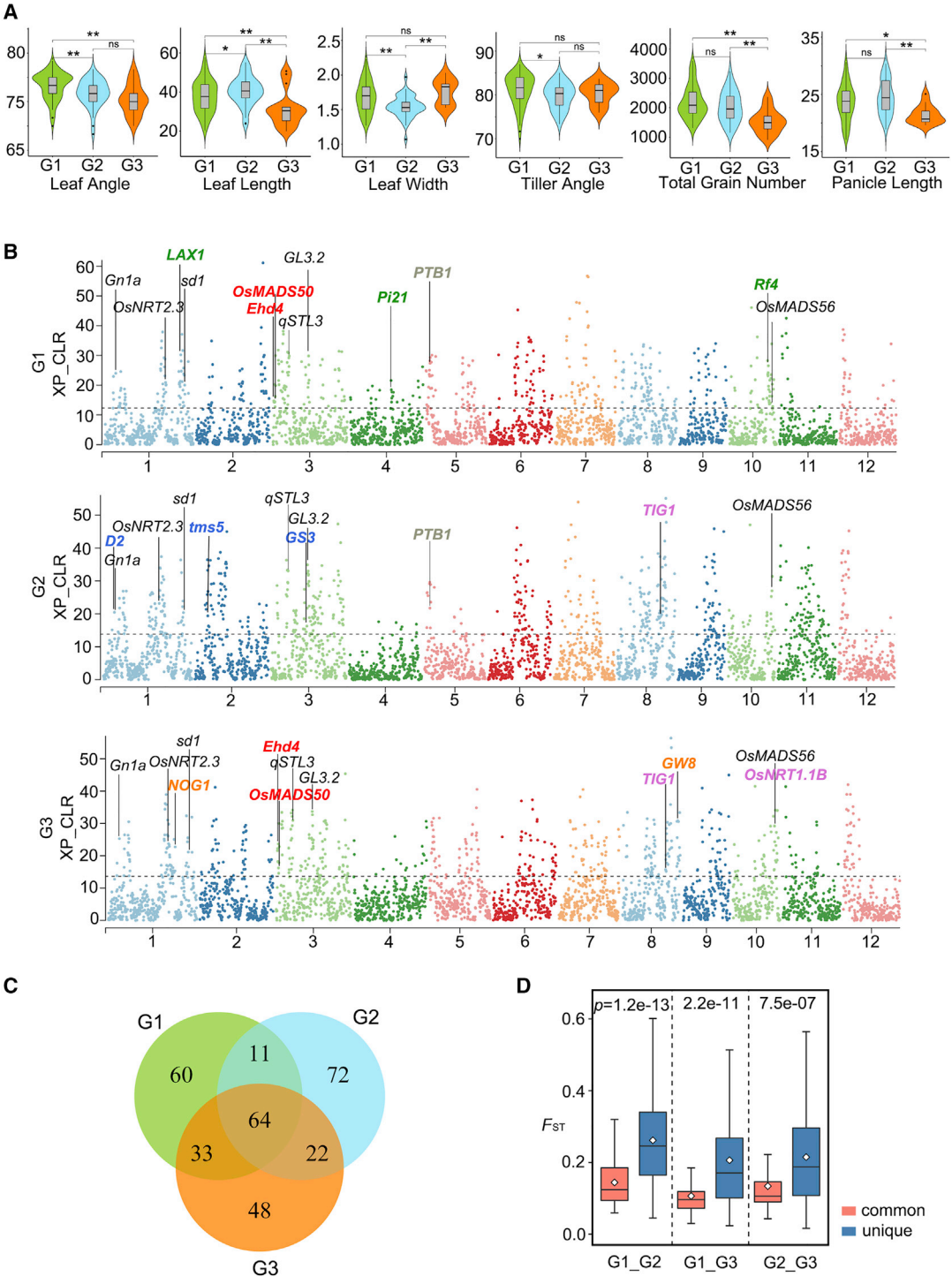

Figure 6. Profiling the selective signal across nuclear genomes of three female-parent subpopulations during genetic improvement breeding.

(A) Phenotypic variations among three types of MS lines for six agronomic traits. The Kruskal–Wallis test was performed to show phenotypic differences. *$p < 0.05$, **$p < 0.01$; ns, not significant.

(B) Distribution of XP-CLR values across the whole nuclear genome of group 1 (G1), group 2 (G2), and group 3 (G3) MS lines, using a panel of rice landrace as the reference population. Known genes located in selective sweeps are highlighted. Group 1-specific genes are marked in green, group 2-specific genes in blue, and group 3-specific genes in orange. Loci shared by group 1 and group 2 are in gray, group 1 and group 3 common loci are in red, group 2 and group 3 common loci are in purple, and common genes for all three groups are displayed in black.

(C) Venn diagram showing the numbers of convergent and divergent selective sites among three groups of MS lines.

(D) The level of sequence differentiation (F_{ST}) was computed for both convergent and divergent selective sites in three pairs of MS lines and compared. The p value was calculated by the Wilcoxon test.

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

compared with shared sites in all three comparisons (Figure 6D). This result indicated that different MS lines experienced convergent and divergent selection during improvement breeding. Several selective sites spanned known genes that control agriculturally important traits (Figure 6B). We further analyzed the functional allele frequencies of these genes and observed that some functional alleles were dominant in the three groups, whereas others were exclusive to a certain group. Protein OsMADS56 (Ryu et al., 2009) is homologous to *Arabidopsis* SOC1 and controls heading date in rice. Almost all MS lines (>95%) in the three groups contained the allele *osmads56* (allele code 1) that functions as a positive regulator of flowering under long-day conditions (Supplemental Figure 14). *Gn1a* (Akter et al., 2005) regulates rice grain production, and the allele that increases grain numbers was widespread in the three groups. Divergent loci were also identified among the three groups. *Pi21* (Fukuoka et al., 2009) encodes a proline-rich protein and is associated with rice blast resistance. The allele type (code 2) associated with improved resistance was dominant (48 of 78) in group 1 MS lines; however, it was rare in landraces (8 of 48) and group 2 (2 of 44) MS lines (Supplemental Figure 14A). *NOG1* (Huo et al., 2017) can increase yield by increasing the grain numbers per panicle, and almost all (95%) accessions in group 3 contained the superior allele type (code 1), whereas only 71.95% and 55.81% of accessions in group 1 and group 2 contained the favorable allele (Supplemental Figure 14A). *TIG1* (Zhang et al., 2019) encodes a TCP transcriptional activator, and the allele *tig1* (code 1) is related to erect tiller growth. Over 90% of MS lines in both group 2 and group 3 contained the *tig1* allele (Supplemental Figure 14A), whereas approximately 70% of group 1 MS lines contained it. We also investigated the allele frequency of four heterosis genes reported by Huang (Huang et al., 2016) (Supplemental Figure 14B), and our observation supported the Huang group's opinion. Approximately 39% and 67% of group 1 MS strains possessed the favorable alleles *hd3a* (Kojima et al., 2002) and *tac1* (Yu et al., 2007), whereas only 4% and 13% of group 2 MS lines contained *hd3a* and *tac1*, and 6% and 5% of group 3 MS strains contained *hd3a* and *tac1*. Thus, both of the favorable alleles were present preferentially in group 1 MS lines. The breeding-favorable allele *GW3p6/qLGY3* was exclusively used by 27% of group 2 MS lines. Figure 1 indicates that group 2 MS lines contained *japonica* organelles and were of *japonica* descent. Our findings are consistent with the report that *GW3p6/qLGY3* exists mainly in tropical *japonica* germplasm (Liu et al., 2018; Yu et al., 2018; Wang et al., 2019). In addition, the allele frequency of *OsMADS51* (Song et al., 2007), associated with early flowering, was higher in all three groups of MS lines compared with landrace rice (Supplemental Figure 14B).

Genomic structure of restorer lines and cross patterns between MS lines and restorers

To better understand the breeding strategies for different MS lines and further analyze the use of candidate genes in selection signals for hybrid rice breeding, we collected 367 rice restorers that were sequenced at a depth of 30×, and we identified 5 537 758 high-quality nuclear SNPs. Based on the SNPs, we performed phylogenetic analysis. According to the neighbor-joining tree and admixture analysis, we classified the restorers into five types, designated male I, male II, male III, male IV, and male V (Figure 7A and 7B). Then we investigated the cross pattern between three groups of MS lines and five types of restorers. Group 1 × male I crosses were dominant among hybrids bred from group 1 MS lines. Thus, using group 1 MS lines and male I restorers, we calculated the difference in parental allele frequency for loci in selective sites (Figure 7C). Heterotic loci (*Hd3a*, *TAC1*, and *OsMADS51*) identified by Huang (Huang et al., 2016) were also included in our analysis. The published functional variations were used to estimate superior allele frequency of candidate genes. The alleles *tac1*, associated with a smaller tiller angle, and *Pi21*, related to increased resistance to rice blast, were contributed by most of the group 1 maternal parents. The *lax1* allele (Gao et al., 2013), related to dense panicles, was under selection in group 1 MS lines (Figure 6B and Supplemental Figure 14A and Supplemental Figure 15A). However, we found that *lax1* had a negative effect on grain weight per plant (Supplemental Figure 15B and 15C), and male I restorers contributed the *LAX1* allele to hybrids. Male I restorers also contributed the superior alleles *ptb1* (Li et al., 2013), associated with seed setting rate, and *d2*, which controls plant architecture, to hybrids (Figure 7C). In addition, high stigma exertion rate is an important trait for efficient production of hybrid seeds using MS lines, but the allele *qSTL3* (Liu et al., 2015a), which is associated with high stigma exertion rate, has not been applied in any of the group 1 MS lines (Figure 7C).

Most of the group 2 MS lines were crossed with the male I, II, and V restorers to breed hybrids. Parents from group 2 × male I, group 2 × male II, and group 2 × male V combinations were chosen to estimate differences in parental allele frequencies of loci in selective sites (Figure 7D). We did not detect any functional alleles with differences in parental allele frequency. None of the alleles investigated had a difference in parental allele frequency greater than 0.5. The superior alleles *Pi21* for rice blast resistance, *qSTL3* for stigma exertion, *hd3a* for heading date, and *tac1* for tiller angle were all at low allele frequencies in both parents. The breeding-favorable alleles for *Pi21*, *Hd3a*, and *TAC1* were rarely exploited in hybrids bred from group 2 MS lines (Supplemental Figure 16G, 16J, and 16K). Group 3 MS lines were preferentially crossed with male V restorers in hybrid production. Parents from the group 3 × male V combination were chosen for further analysis (Figure 7E). The allele *LAX1* for higher grain weight was contributed by paternal parents, and the superior alleles of *qSTL3*, *hd3a*, and *tac1* were also at low allele frequencies in both parents.

DISCUSSION

Comprehensive analysis of the cytoplasmic genomes of rice hybrids has been lacking. In this study, we identified five major types of organelle genomes in cultivated hybrid rice and revealed that hybrid production systems that use different fertility recovery and sterility maintenance mechanisms have different cytoplasm types. We found that the widespread use of CMS resources from naturally abortive wild rice broadened the genetic diversity of cultivated hybrid rice. Compared with the other four cytoplasm groups, group 1 cytoplasm contained several specific polymorphisms, including missense mutations in *rps1*, *orf288* (*WA352c*), *orf224*, and *rpoB* and in-frame deletions in *orf224*.

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

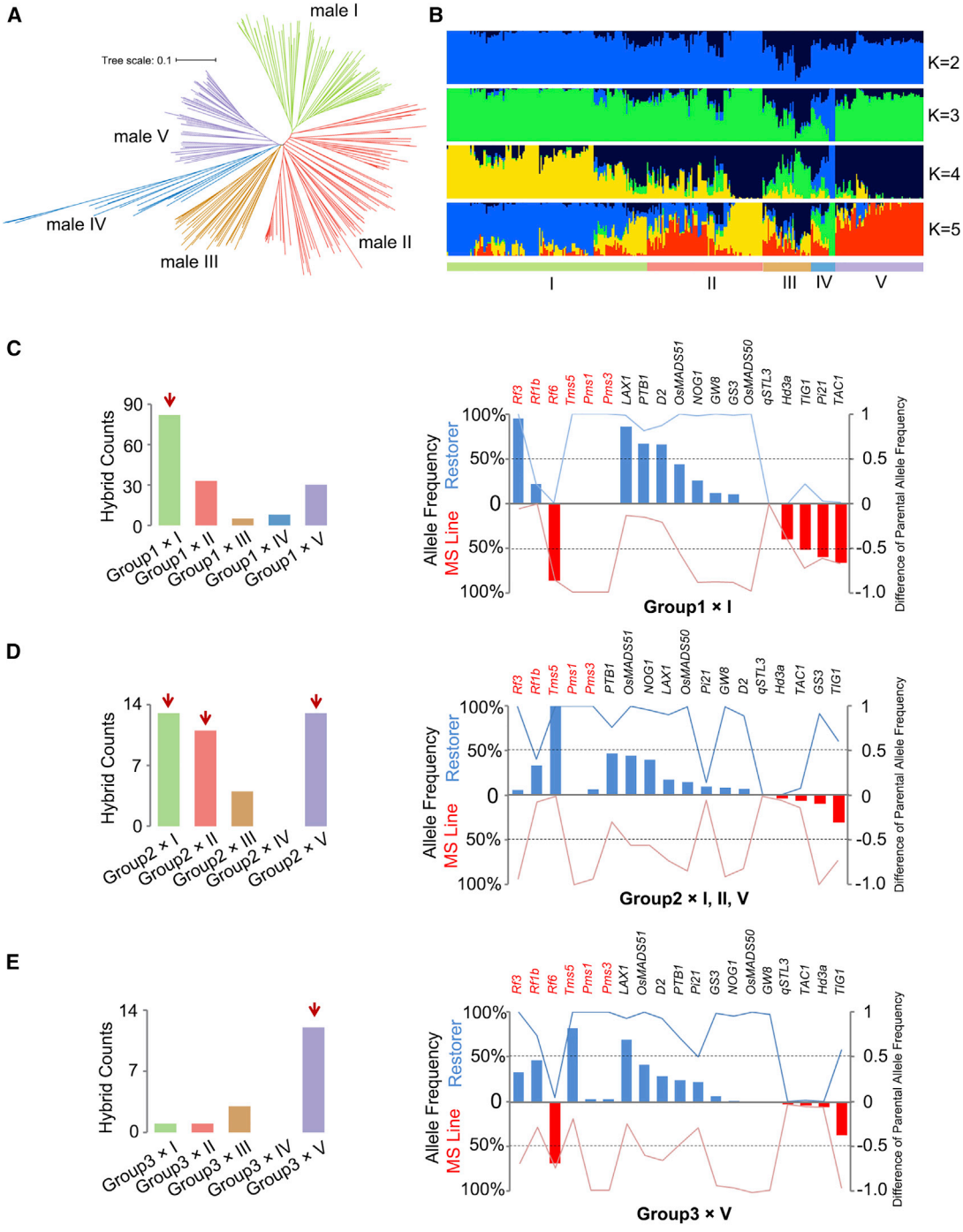

Figure 7. Genomic structure of rice restorer lines and differences in parental allele frequencies for loci in selective sites.

(A) Neighbor-joining tree constructed from 5 537 758 SNPs in 367 restorer lines.

(B) Admixture analysis with different numbers of clusters (K = 2, 3, 4, and 5). The y axis shows cluster proportions, and the x axis lists rice restorer accessions according to their phylogenetic relationships provided by the neighbor-joining tree.

(C) The numbers of crosses between group 1 MS lines and five types of restorers for hybrid production (left). Group 1 MS lines and type I restorers were used to calculate differences in parental allele frequencies for loci in selective sites (right).

(D) The numbers of hybrid crosses between group 2 MS lines and five types of restorers (left). Group 2 MS lines and restorers of types I, II, and IV were used to calculate differences in parental allele frequencies (right).

(E) The numbers of hybrid crosses between group 3 MS lines and five types of restorers (left). Group 3 MS lines and type V restorers were used to calculate differences in parental allele frequencies (right). Fertility-associated loci are marked in red, and three heterotic loci (*Hd3a*, *TAC1*, *OsMADS51*) reported by Huang (Huang et al., 2016) are also included.

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

Among these polymorphisms, deletion in the predicted protein-coding gene *orf224* was predicted to be deleterious by PROVEAN.

MS lines contribute their cytoplasm to hybrids, and we found that the nuclei of MS lines were associated with their cytoplasms. We divided MS lines into five major groups based on comprehensive consideration of the nucleus and cytoplasm. Phylogenetic and kinship analysis revealed that several kernel MS lines were adopted as backbone parents and contributors of fertility-associated genes during maternal line breeding; improved breeding based on them generated a series of MS lines that share close relationships. The fixation of fertility-associated genes in the nucleus could lower nucleotide diversity of flanking sequences and bring about linkage drags. Exploitation of CMS genes in the mitochondrion resulted in certain types of cytoplasms being retained by MS lines. Thus, different kernel accessions cooperating with different fertility-associated genes drove nuclear sequence divergence and cytoplasm differentiation of MS lines through the use of different fertility recovery and sterility maintenance mechanisms. Nuclear genomic structure differentiation between CMS lines and EGMS lines was reported by Lin and Lv (Lin et al., 2020; Lv et al., 2020). In our work, through comprehensive consideration of variations in organelles and the nucleus, we further divided MS lines into five major groups: groups 1, 4, and 5 CMS lines for the three-line system and groups 2 and 3 EGMS lines for the two-line system. Previous studies have always performed analyses using the whole set of two-line maternal populations (Lin et al., 2020; Lv et al., 2020). But we found that there were two subpopulations of two-line maternal parents, and their genomic structures and cross patterns with restorers were different. Future scientific research and hybrid rice breeding could divide the two-line maternal accessions into two subgroups for consideration. Furthermore, unexpectedly, although the two-line maternal parents used a single nuclear locus for male pollen abortion (Zhou et al., 2014), their nuclei were still associated with the cytoplasm. We speculated that the kernel accessions were responsible for this phenomenon. Breeders may prefer to use the EGMS lines as maternal parents during improvement breeding, although EGMS lines can recover fertility under certain conditions. Thus, the cytoplasms from kernel varieties were retained by most EGMS lines.

Both convergent and divergent selective sites were identified when searching for breeding footprints of the three types of MS lines. Several known genes were located in selective loci, and the frequencies of elite alleles differed among the three groups of MS lines (for example, the loci *Pi21*, *LAX1*, and *D2*). Some loci were genetically complementary between maternal lines and restorers. The crosses between group 1 MS lines and corresponding restorers contained more genetically complementary loci than crosses from group 2 and group 3 MS lines. The group 1 maternal lines contain superior alleles of *TAC1*, *Pi21*, and *TIG1*, whereas the restorers contribute *LAX1*, *PTB1*, and *D2*. This may be associated with genetic divergence of group 1 maternal lines and corresponding restorers. According to Lv's report, most *indica* three-line hybrids were crosses between *IndII* restorers and *IndI* CMS lines (Lv et al., 2020). In addition, we also found that the gene *PTB1* (Li et al., 2013), related to seed setting rate, fell within the selection signal of group 1 and group 2 MS lines. However, most group 1 and group 2 MS lines contained the genotype associated with reduced seed setting rate. We speculated that "reverse selection" might be related to the following two reasons: (1) relatively limited germplasm used for MS line breeding led to low sequence diversity in the region that contains *PTB1*. Five kernel MS accessions (Zhenshan97A, V20A, Jin23A, Guangzhan63S, and XinanS) all contain the inferior allele of *PTB1*. (2) There were other genes that controlled important agronomic traits under selection near *PTB1*, and linkage drag was responsible for the wide spread of the inferior allele in group 1 and group 2 MS lines. We also found that most of the restorers crossed with group 1 and group 2 MS lines contributed the superior allele of *PTB1* for hybrid breeding, consistent with Wei's report (Wei et al., 2021). Locus *qSTL3* (Liu et al., 2015a) was in a similar situation. It was located in the selection site of all three types of MS lines. This gene has been reported to control stigma exertion, but the three types of MS lines rarely contained the genotype with a high stigma exertion rate. Because high stigma exertion rate is key to efficient hybrid seed production, we suggest that future improvement breeding could use molecular markers to introduce alleles with high stigma exertion rate. The wide use of kernel MS lines as backbone parents led to the close relationship among MS lines and may be responsible for the wide spread of breeding-undesirable alleles. Broadening genetic resources, for example, by developing *indica–japonica* or cultivated–wild crosses, could promptly introduce more superior alleles and may be a solution for breaking the hybrid breeding bottleneck.

METHODS

Samples and sequencing

Male sterile lines

We collected 159 rice MS lines and 367 restorers from the collections preserved at the China National Rice Research Institute in Hangzhou and the varieties bred by Win-All High-Tech Seed Co. The samples were germinated and planted in the experimental field in Hangzhou, China (N30°05, E119°95). For each MS line, heading date, plant height, tiller angle, panicle number, total grain number, leaf angle, leaf length, leaf width, and panicle length were investigated; three randomly chosen individual plants were evaluated, and their means were calculated. Fresh leaf tissue was collected from each plant in the vegetative growth phase, and genomic DNA was extracted. A sequencing library with ~400-bp insert size was constructed. Sequencing was performed on the NovaSeq 6000 platform and generated 150 base pairs paired-end reads.

Hybrid rice

In our previous work, we sequenced 1495 hybrid varieties from the collections of rice accessions preserved at the China National Rice Research Institute in Hangzhou, China (Huang et al., 2015b). Five representative hybrid accessions and reference variety Nipponbare were chosen for whole-genome sequencing. Detailed information for sampled varieties is listed in Supplemental Table 1. Genomic DNA was isolated from leaves at the four-leaf seedling stage, and the plants were grown in the greenhouse of the National Center for Gene Research, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai. For Illumina short-insert size library construction, genomic DNA was fragmented, end-repaired, size-selected at ~500 bp, and ligated to Illumina PCR-free paired-end adapters. Then the library was sequenced using the Illumina HiSeq 2500 platform, producing 250 base pairs paired-end reads. Long reads from PacBio and Oxford Nanopore Technologies were produced for the reference variety Nipponbare. For five representative accessions, we produced only Oxford Nanopore data. Among them, four hybrid accessions (Z868, Z395, Z1292,

Please cite this article in press as: Gu et al., Cytoplasmic and nuclear genome variations of rice hybrids and their parents inform the trajectory and strategy of hybrid rice breeding, Molecular Plant (2021), https://doi.org/10.1016/j.molp.2021.08.007

and Z718) and Nipponbare were pooled on the PromethION platform. Nanopore data for Z45 were produced using the GridION platform.

SNP calling

Raw data with low base quality and adapter contamination were trimmed by the software Trimmomatic (version 0.38) with the parameters "ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:true" and "MAXINFO:50:0.6". Then the clean reads were aligned against the reference sequence using the software BWA version 0.7.1 (Li and Durbin, 2010) with the parameters "-M -R "@RG\tID:SampleID\tPL:Illumina\tSM:SampleName" ". The mapping files were processed with SAMtools version 0.1.19 (Li et al., 2009). We performed SNP calling using the "HaplotypeCaller, GenomicsDBImport and GenotypeGVCFs" functions in GATK (McKenna et al., 2010). To obtain high-quality SNPs, we retained SNPs with QD > 10.00, AC > 10, and FS < 15.000 for cytoplasmic variants and –cluster-size 3 –cluster-window-size 10, QD > 10.00, AC > 3, FS < 15.000, and 5 < DP < 200 for nuclear polymorphisms.

Population structure and genomic diversity analysis

Phylogenetic analysis

High-quality SNPs were used for population structure analysis. For cytoplasm, the matrix of pairwise genetic distance derived from a simple SNP matching coefficient was used to construct a neighbor-joining tree. For nuclei, distance matrices were constructed with PLINK (v.1.90b6.12 64-bit) (Purcell et al., 2007) with the parameters "–maf 0.05 –distance square 1-ibs". The software PHYLIP version 3.69 was chosen to construct the phylogenetic tree (http://evolution.genetics.washington.edu/phylip.html), and the online tool Interactive Tree of Life (https://itol.embl.de) was chosen for tree visualization.

Principal component analysis

The software GCTA version 1.92.1 beta6 (Yang et al., 2011) was used to perform PCA with default parameters. The top three principal components were chosen for PCA visualization with an R script.

Admixture analysis

For the admixture analysis, we carried out quality control with PLINK software using "–geno 0.05 –maf 0.05 –hwe 0.0001 –make-bed". We performed genomic structure analysis using the ADMIXTURE program with default parameters. Because five types of cytoplasm were identified, we obtained results for K values of 2–5 using the R package pophelper (Francis, 2017).

Kinship analysis

To measure the relationships for all comparisons among the MS lines, kinship values were computed using the emmax-kin program in EMMAX software (version emmax-beta-07Mar2010) (Kang et al., 2010) with default parameters. We set the cutoff kinship coefficient to 0.91. The results were visualized with Cytoscape software (Otasek et al., 2019). Every circle represented a sample, and the circle size indicated the number of accessions related to it. Pairs of accessions with close relationships were connected with lines.

F statistics and pbs analysis

The population differentiation index F_{ST} was estimated using VCFtools version 0.1.13 (Danecek et al., 2011) around the plant architecture-associated gene *D2* with window size 10 kilobase and window step size 10 kilobase. Three populations, group 1, group 2, and group 3, were used to perform pbs analysis. Site allele frequency likelihood was calculated for every population using the ANGSD program (Korneliussen et al., 2014) with parameters "-b bam_file_list -anc reference.fasta -dosaf 1 -gl 1 -r coordinate_for_candidate_region -fold 1". Next, the realSFS program was used to compute site frequency spectra for three pairs of groups, and F_{ST} and pbs values were calculated with realSFS for the 100-kilobase region (chr1: 5 200 000–5 300 000) comprising *D2*. We generated 500 permutation tests to derive an empirical *p* value.

Updating Nipponbare mitochondrial and chloroplast genome

The cultivar Nipponbare (*O. sativa japonica*) was used as the source of DNA. Leaf tissues from seedlings were used for cpDNA preparation, and roots from seedlings were used for mtDNA preparation. cpDNA and mtDNA were prepared using the Column Plant Chloroplast DNAout Kit (TIANDZ cat no. 120406-15) and Cell Mitochondria Isolation Kit (Beyotime cat no. C3601). The sequencing library with an insert size peaking at 450 base pairs was constructed using the Illumina TruSeq Nano DNA Sample Prep Kit and sequenced on the Illumina HiSeq platform. The PacBio sequencing library was prepared following the standard PacBio protocol and sequenced on the PacBio long read sequencing instrument RSII. Clean Illumina reads and corrected PacBio reads were used for *de novo* assembly with SPAdes v.3.10.1 (Bankevich et al., 2012; Antipov et al., 2016). Contigs with high coverage were selected as candidates, and clean Illumina reads were then aligned to candidate contigs. According to the information supplied by the paired-end reads and overlap relationships, we anchored the draft contigs and filled the gaps using GapCloser v.1.12.

For mitochondria, protein-coding genes were predicted using Exonerate (Slater and Birney, 2005), tRNA was predicted by tRNAscan-SE v.1.3.1 (Lowe and Eddy, 1997), and rRNA was identified by BLASTN. The chloroplast genome was annotated with GeSeq (Tillich et al., 2017).

The graphical maps of mitochondrial and chloroplast genomes were generated with OrganellarGenomeDRAW (OGDRAW) (Lohse et al., 2013).

Organelle genome assembly and validation

We constructed a reference-guided pipeline for organelle genome assembly using whole-genome sequencing PacBio/Nanopore data. The pipeline involved three major steps.

Step I. Constructing variety-customized reference. The quality of short reads was controlled using Trimmomatic version 0.38 (Bolger et al., 2014). Then, low-coverage short reads were supplied to SOAPdenovo2 version 2.04 (Luo et al., 2012) to perform assembly with k-mer size 127. The resulting assemblies were aligned back to the updated Nipponbare reference sequence using MUMmer version 3.22 (Kurtz et al., 2004) with the parameters "nucmer -maxmatch; show-coords -rcl; delta-filter -q; show-coords -rcl", and representative contigs were extracted for mitochondrial and chloroplast genomes with alignment identity >94% and coverage >50%. The resulting contigs were combined with the corresponding reference to extract cp/mt reads from whole-genome sequencing data using SMALT with the parameters "-i 700 -j 50 -m 60". The extracted cp/mt reads were supplied to SOAPdenovo2. The resulting contigs larger than 2 kilobase were combined with the corresponding reference sequence to produce a "variety-customized reference sequence."

Step II. Extracting PacBio/Nanopore reads based on the variety-customized reference sequence. PacBio and Nanopore long reads were aligned to the customized reference using minimap2 (Li, 2018) (2.12-r849-dirty) with default parameters, and cp/mt derived reads were extracted using customized perl scripts with read length >2 kilobase and coverage >90%.

Step III. Assembly. A hybrid assembly was generated using SPAdes v.3.13.0 (Bankevich et al., 2012; Antipov et al., 2016) with parameters "-k 49,55,67,77,83". SSPACE Basic version 2.0 (Boetzer et al., 2010) and GapFiller version 1.10 (Boetzer and Pirovano, 2012) were used for contig extension. For chloroplast assembly, manual selection of long reads to connect IRa/IRb with LSC/SSC was needed.

Polymorphism detection, validation, and evaluation

Identification of cytoplasmic genomic variations was performed as described by Zhao (Zhao et al., 2018). The potential effects of the variants located in protein-coding genes were evaluated. For missense mutations and indels, PROVEAN (http://provean.jcvi.org/seq_submit.php) was used to predict the potential impact of variations on the biological functions of proteins. In addition, the protein sequences containing nonsynonymous mutations and indels were searched for functional domain information using InterProScan version 5.22-61.0 (Jones et al., 2014) with the parameters "-f TSV; -iprlookup; -goterms; -seqtype p".

Polymorphisms located in predicted domains were viewed using IBS version 1.0.3 (Liu et al., 2015b).

Selection signal scan

A composite likelihood approach (XP-CLR) was used to scan for selection signals associated with hybrid rice breeding across the whole genome. For the XP-CLR analysis, we removed SNPs with missing rate >20% among the population and transformed the vcf format file to XP-CLR assigned format with a customized perl script. Then, the landrace population was compared as a reference with three groups of MS lines. We scanned the whole genome for selective signals with the XP-CLR program using a step size of 100 base pairs and a sliding window size of 0.05 cM. Then we combined the windows into nonoverlapping 100-kilobase genomic features, averaged the XP-CLR values of windows located in a genomic feature, and assigned the average value to the 100-kilobase region. We chose the top 20% of genomic features as candidate selective sweeps. To improve the accuracy, we filtered candidate selective signals with a π ratio ($\pi_{landrace}/\pi_{male\ sterile\ lines}$) lower than the median of the genome-wide value. We then combined the nearby candidate selective sweeps. The result was visualized using the R package CMplot (Yin et al., 2021). If the overlap length of a pair of selective sweeps from two groups was greater than half of the total length, they were identified as common selective signals; otherwise, they were considered to be unique selective signals.

DATA AVAILABILITY

All data supporting the findings are available in the paper and supplementary information files. Raw sequence data for this article were deposited in the NCBI Sequence Read Archive under study accession no. PRJEB40526. The updated Nipponbare organelle genome sequences and five representative hybrid assemblies are available at Figshare (http://figshare.com) under the following doi: 10.6084/m9.figshare.14709306.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (31788103) and the Chinese Academy of Sciences (XDB27010301).

AUTHOR CONTRIBUTIONS

B.H. conceived the project. Z.G. and B.H. designed and supervised the project and wrote the manuscript. Z.G. performed most of the data analysis. Z.Z., Q.Z., Y.Z., X.P., Y.L., H.L., L.Z., and X.H. contributed to data analysis. T.H. managed the server station. Z.L., Q.-L.Z., and J.G. were responsible for field management and phenotype information recording. Q.F. and C.Z. performed library construction and whole-genome sequencing. B.D. and R.S. conducted wet experiments, including PCR and Sanger sequencing.

ACKNOWLEDGMENTS

We thank the China National Rice Research Institute and Win-All High-Tech Seed Co., Ltd. for providing publicly available rice male sterile resources. We thank Professors Hongxuan Lin and Yijing Zhang for their valuable suggestions. No conflict of interest declared.

Received: April 7, 2021
Revised: July 6, 2021
Accepted: August 6, 2021
Published: August 10, 2021

REFERENCES

Akter, M.B., Piao, R., Kim, B., Lee, Y., Koh, E., and Koh, H.-J. (2005). Cytokinin oxidase regulates rice grain production. Science **309**:741–745.

Antipov, D., Korobeynikov, A., McLean, J.S., and Pevzner, P.A. (2016). HybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics **32**:1009–1015.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19**:455–477.

Boetzer, M., and Pirovano, W. (2012). Toward almost closed genomes with GapFiller. Genome Biol. **13**:R56.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**:578–579.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**:2114–2120.

Cai, J., Liao, Q.P., Dai, Z.J., Zhu, H.T., Zeng, R.Z., Zhang, Z.M., and Zhang, G.Q. (2013). Allelic differentiations and effects of the Rf3 and Rf4 genes on fertility restoration in rice with wild abortive cytoplasmic male sterility. Biol. Plant **57**:274–280.

Cheng, S.H., Zhuang, J.Y., Fan, Y.Y., Du, J.H., and Cao, L.Y. (2007). Progress in research and development on hybrid rice: a super-domesticate in China. Ann. Bot. **100**:959–966.

Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics **31**:2745–2747.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158.

Dong, H., Zhao, H., Xie, W., Han, Z., Li, G., Yao, W., Bai, X., Hu, Y., Guo, Z., Lu, K., et al. (2016). A novel tiller angle gene, TAC3, together with TAC1 and D2 largely determine the natural variation of tiller angle in rice cultivars. Plos Genet. **12**:e1006412.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., et al. (2017). Sequencing and de novo assembly of a near complete indica rice genome. Nat. Commun. **8**:15324.

Fan, Y., and Zhang, Q. (2018). Genetic and molecular characterization of photoperiod and thermo-sensitive male sterility in rice. Plant Reprod. **31**:3–14.

Francis, R.M. (2017). POPHELPER: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. **17**:27–32.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., et al. (2009). Loss of function of a proline-containing protein confers durable disease resistance in Rice. Science **325**:998–1001.

Gao, Z.Y., Zhao, S.C., He, W.M., Guo, L.B., Peng, Y.L., Wang, J.J., Guo, X. Sen, Zhang, X.M., Rao, Y.C., Zhang, C., et al. (2013). Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. Proc. Natl. Acad. Sci. U S A **110**:14492–14497.

Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: Dynamics and maintenance. Biochimie **100**:107–120.

Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature **490**:497–501.

Huang, W., Yu, C., Hu, J., Wang, L., Dan, Z., Zhou, W., He, C., Zeng, Y., Yao, G., Qi, J., et al. (2015a). Pentatricopeptide-repeat family protein RF6 functions with hexokinase 6 to rescue rice cytoplasmic male sterility. Proc. Natl. Acad. Sci. U S A **112**:14984–14989.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., et al. (2015b). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat. Commun. **6**:6258.

Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y., Li, W., Cheng, B., Xia, J., et al. (2016). Genomic architecture of heterosis for yield traits in rice. Nature **537**:629–633.

Huo, X., Wu, S., Zhu, Z., Liu, F., Fu, Y., Cai, H., Sun, X., Gu, P., Xie, D., Tan, L., et al. (2017). NOG1 increases grain production in rice. Nat. Commun. **8**:1497.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236–1240.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354.

Kim, Y.J., and Zhang, D. (2018). Molecular control of male fertility for crop hybrid breeding. Trends Plant Sci. **23**:53–65.

Kojima, S., Takahashi, Y., Kobayashi, Y., Monna, L., Sasaki, T., Araki, T., and Yano, M. (2002). Hd3a, a rice ortholog of the Arabidopsis FT gene, promotes transition to flowering downstream of Hd1 under short-day conditions. Plant Cell Physiol. **43**:1096–1105.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. BMC Bioinformatics **15**:356.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. **5**:R12.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094–3100.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26**:589–595.

Li, J., and Yuan, L. (2000). Hybrid rice: genetics, breeding, and seed production. Plant Breeding Reviews **17**:15–158.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, S., Li, W., Huang, B., Cao, X., Zhou, X., Ye, S., Li, C., Gao, F., Zou, T., Xie, K., et al. (2013). Natural variation in PTB1 regulates rice seed setting rate by controlling pollen tube growth. Nat. Commun. **4**:2793.

Lin, Z., Qin, P., Zhang, X., Fu, C., Deng, H., Fu, X., Huang, Z., Jiang, S., Li, C., Tang, X., et al. (2020). Divergent selection and genetic introgression shape the genome landscape of heterosis in hybrid rice. Proc. Natl. Acad. Sci. U S A **117**:4623–4631.

Liu, Q., Qin, J., Li, T., Liu, E., Fan, D., Edzesi, W.M., Liu, J., Jiang, J., Liu, X., Xiao, L., et al. (2015a). Fine mapping and candidate gene analysis of qSTL3, a stigma length-conditioning locus in rice (Oryza sativa L.). PLoS One **10**:e0127938.

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z., Lahrmann, U., Zhao, Q., Zheng, Y., Zhao, Y., et al. (2015b). IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics **31**:3359–3361.

Liu, Q., Han, R., Wu, K., Zhang, J., Ye, Y., Wang, S., Chen, J., Pan, Y., Li, Q., Xu, X., et al. (2018). G-protein βγ subunits determine grain size through interaction with MADS-domain transcription factors in rice. Nat. Commun. **9**:852.

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. **41**:575–581.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955–964.

Luan, J., Liu, T., Luo, W., Liu, W., Peng, M., Li, W., Dai, X., Liang, M., and Chen, L. (2013). Mitochondrial DNA genetic polymorphism in thirteen rice cytoplasmic male sterile lines. Plant Cell Rep. **32**:545–554.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience **1**:18.

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., Fang, C., Zhang, Q., Bai, M., Yao, N., et al. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. Nat. Genet. **45**:573–577.

Lv, Q., Li, W., Sun, Z., Ouyang, N., Jing, X., He, Q., Wu, J., Zheng, J., Zheng, J., Tang, S., et al. (2020). Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. Nat. Commun. **11**:4778.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. (2002). The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: Frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol. Genet. Genomics **268**:434–445.

Otasek, D., Morris, J.H., Bouças, J., Pico, A.R., and Demchak, B. (2019). Cytoscape Automation: Empowering workflow-based network analysis. Genome Biol. **20**:185.

Palmer, J.D., and Herbon, L.A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. Mol. Evol. **28**:87–97.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575.

Qian, Q., Guo, L., Smith, S.M., and Li, J. (2016). Breeding high-yield superior quality hybrid super rice by rational design. Natl. Sci. Rev. **3**:283–294.

Ryu, C.H., Lee, S., Cho, L.H., Kim, S.L., Lee, Y.S., Choi, S.C., Jeong, H.J., Yi, J., Park, S.J., Han, C.D., et al. (2009). OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)-dependent flowering in rice. Plant Cell Environ. **32**:1412–1427.

Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6**:31.

Song, L.K., Lee, S., Hyo, J.K., Hong, G.N., and An, G. (2007). OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1, OsMADS14, and Hd3a. Plant Physiol. **145**:1484–1494.

Tang, H., Luo, D., Zhou, D., Zhang, Q., Tian, D., Zheng, X., Chen, L., and Liu, Y.G. (2014). The rice restorer Rf4 for wild-abortive cytoplasmic male sterility encodes a mitochondrial-localized PPR protein that functions in reduction of WA352 transcripts. Mol. Plant **7**:1497–1500.

Tian, X., Zheng, J., Hu, S., and Yu, J. (2006). The rice mitochondrial genomes and their variations. Plant Physiol. **140**:401–410.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. **45**:W6–W11.

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., Su, D., Chen, Y., Guo, J., Luo, D., et al. (2006). Cytoplasmic male sterility of rice with Boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell **18**:676–687.

Wang, K., Gao, F., Ji, Y., Liu, Y., Dan, Z., Yang, P., Zhu, Y., and Li, S. (2013). ORFH79 impairs mitochondrial function via interaction with a subunit of electron transport chain complex III in Honglian cytoplasmic male sterile rice. New Phytol. **198**:408–418.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**:43–49.

Wang, C., Tang, S., Zhan, Q., Hou, Q., Zhao, Y., Zhao, Q., Feng, Q., Zhou, C., Lyu, D., Cui, L., et al. (2019). Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy. Nat. Commun. **10**:2982.

Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., Liu, J., Wang, Q., Olsen, K.M., Han, B., et al. (2021). A quantitative genomics map of rice provides genetic insights and guides breeding. Nat. Genet. **53**:243–253.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**:76–82.

Yi, P., Wang, L., Sun, Q., and Zhu, Y. (2002). Discovery of mitochondrial chimeric-gene associated with cytoplasmic male sterility of HL-rice. Chin. Sci. Bull. **47**:744–747.

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., and Liu, X. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for Genome-Wide Association Study. Genomics Proteomics Bioinformatics https://doi.org/10.1016/j.gpb.2020.10.007.

Yu, B., Lin, Z., Li, H., Li, X., Li, J., Wang, Y., Zhang, X., Zhu, Z., Zhai, W., Wang, X., et al. (2007). TAC1, a major quantitative trait locus controlling tiller angle in rice. Plant J. **52**:891–898.

Yu, J., Miao, J., Zhang, Z., Xiong, H., Zhu, X., Sun, X., Pan, Y., Liang, Y., Zhang, Q., Abdul Rehman, R.M., et al. (2018). Alternative splicing of OsLG3b controls grain length and yield in japonica rice. Plant Biotechnol. J. **16**:1667–1678.

Zhang, W., Tan, L., Sun, H., Zhao, X., Liu, F., Cai, H., Fu, Y., Sun, X., Gu, P., Zhu, Z., et al. (2019). Natural variations at TIG1 encoding a TCP transcription factor contribute to plant architecture domestication in rice. Mol. Plant **12**:1075–1089.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. **50**:278–284.

Zhou, H., Zhou, M., Yang, Y., Li, J., Zhu, L., Jiang, D., Dong, J., Liu, Q., Gu, L., Zhou, L., et al. (2014). RNase Z S1 processes Ub L40 mRNAs and controls thermosensitive genic male sterility in rice. Nat. Commun. **5**:4884.