the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **92,** 349–362 doi: 10.1111/tpj.13661

Three representative inter and intra-subspecific crosses reveal the genetic architecture of reproductive isolation in rice

Guangwei Li†, Xiaoting Li†, Yuan Wang, Jiaming Mi, Feng Xing, Dahan Zhang, Qiyan Dong, Xianghua Li, Jinghua Xiao, Qifa Zhang and Yidan Ouyang*

National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China

Received 5 May 2017; revised 25 July 2017; accepted 7 August 2017; published online 14 August 2017.
*For correspondence (e-mail diana1983941@mail.hzau.edu.cn).
†These authors contributed equally to this work.

SUMMARY

Systematic characterization of genetic and molecular mechanisms in the formation of hybrid sterility is of fundamental importance in understanding reproductive isolation and speciation. Using ultra-high-density genetic maps, 43 single-locus quantitative trait loci (QTLs) and 223 digenic interactions for embryo-sac, pollen, and spikelet fertility are depicted from three crosses between representative varieties of *japonica* and two varietal groups of *indica*, which provide an extensive archive for investigating the genetic basis of reproductive isolation in rice. Ten newly detected single-locus QTLs for inter- and intra-subspecific fertility are identified. Three loci for embryo-sac fertility are detected in both Nip × ZS97 and Nip × MH63 crosses, whereas QTLs for pollen fertility are not in common between the two crosses thus leading to fertility variation. Five loci responsible for fertility and segregation distortion are observed in the ZS97 × MH63 cross. The importance of two-locus interactions on fertility are quantified in the whole genome, which identify that three types of interaction contribute to fertility reduction in the hybrid. These results construct the genetic architecture with respect to various forms of reproductive barriers in rice, which have significant implications in utilization of inter-subspecific heterosis along with improvement in the fertility of *indica*–*indica* hybrids at single- and multi-locus level.

Keywords: *Oryza sativa* (rice), hybrid sterility, reproductive isolation, speciation, genetic basis, fertility improvement.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.13721.

INTRODUCTION

Reproductive isolation is the indication and maintaining force of speciation, which allows populations to evolve independently. During the process of speciation, mutations have arisen and fixed in diverging populations, leading to accumulation of genetic changes that confer deleterious interactions, thus resulting reproductive barriers and formation of new species. Despite the recent progress in understanding the hybrid incompatibility genes and mechanisms responsible for reproductive isolation in model organisms (Herrmann *et al.*, 1999; Bauer *et al.*, 2005, 2007; Brideau *et al.*, 2006; Masly *et al.*, 2006; Bomblies *et al.*, 2007; Chen *et al.*, 2008; Lee *et al.*, 2008; Long *et al.*, 2008; Seidel *et al.*, 2008; Bikard *et al.*, 2009; Chou *et al.*, 2010; Mizuta *et al.*, 2010; Yamagata *et al.*, 2010; Maheshwari and Barbash, 2011; Yang *et al.*, 2012; Ouyang and Zhang, 2013; Alcazar *et al.*, 2014; Chae *et al.*, 2014; Kubo *et al.*, 2015; Lafon-Placette and Kohler, 2015; Yu *et al.*, 2016), it remains to be investigated regarding the evolutionary dynamics and the underlying genetics of various forms of barriers leading to the formation of new species.

Rice maintains high genetic diversity and significant population structures due to adaptation to a wide range of geographic distributions and domestication under artificial selection (Huang *et al.*, 2011). The Asian cultivated rice (*Oryza sativa*. L) contains two major groups of *indica* and *japonica* and several other sub-types including *Aus*, *aromatic/basmati*, and intermediate groups (Garris *et al.*, 2005; Huang *et al.*, 2010; Xu *et al.*, 2012; Rice Genomes

Project, 2014; Civan *et al.*, 2015). The *japonica* subspecies can be further classified into temperate and tropical groups, whereas two major varietal groups of *indica* subspecies, *indica I* (*IndI*) and *indica II* (*IndII*), were identified based on whole genome sequencing data, which showed substantial differentiation due to geographic adaptation and accumulation of divergent selections in distinct breeding programs (Xie *et al.*, 2015). The *indica* and *japonica* groups have been also referred to as '*hsien*' and '*keng*' since the Han Dynasty in China (>2000 years ago) (Ting, 1949a,b), and they were recognized based on the reproductive barriers observed. Hybrid sterility between *indica* and *japonica* represents one of the best-characterized examples of postzygotic reproductive isolation in plants. Reproductive isolation between the *indica* and *japonica* subspecies results in barriers causing sterility in various forms, impeding gene flow between different populations (Ouyang *et al.*, 2010). The *Aus* varieties gave relatively fertile hybrids when crossing either to *indica* or *japonica* varieties, which conferred a wide spectrum and high level of wide-compatibility (Morinaga and Kuriyama, 1958). The existence of such group was subsequently demonstrated by Ikehashi and Araki (Ikehashi and Araki, 1984) and was also confirmed by a series of studies (Yanagihara *et al.*, 1992; Liu *et al.*, 1996; Wang *et al.*, 1998; Mi *et al.*, 2016), which enables hybridization and gene flow between the two subspecies, thus maintaining species coherence and providing an opposing force for reproductive isolation. The divergent populations in rice have provided an excellent model system for understanding the evolutionary genetic mechanisms of reproductive barriers.

Using different mapping populations, a number of QTLs that contribute to hybrid sterility have been identified, leading to both female gamete abortion and pollen sterility (Ouyang *et al.*, 2009). The genetic basis of hybrid sterility has been investigated mainly based on biparental crosses and their derivative mapping populations (Wang *et al.*, 2005; Zhu *et al.*, 2005a,b; Zhao *et al.*, 2006; Mizuta *et al.*, 2010; Zhang *et al.*, 2011). Such populations are also used to identify the genetic factors underlying the mechanisms of segregation distortion (Reflinur *et al.*, 2014). In addition, three-way crosses are frequently used to assay the genetic effect of wide-compatibility alleles, which enable hybridization and produce fertile hybrids between the two subspecies (Yanagihara *et al.*, 1992; Wan *et al.*, 1993, 1996; Wan and Ikehashi, 1995; Liu *et al.*, 1997; Wang *et al.*, 1998; Song *et al.*, 2005). Seven loci/locus pairs for hybrid embryo-sac or pollen sterility have been cloned in rice, which caused reproductive barriers in diverse inter-specific or inter-subspecific populations (Chen *et al.*, 2008; Long *et al.*, 2008; Mizuta *et al.*, 2010; Yamagata *et al.*, 2010; Du *et al.*, 2011; Yang *et al.*, 2012; Ouyang and Zhang, 2013; Kubo *et al.*, 2016a; Ouyang *et al.*, 2016; Yu *et al.*, 2016; Xie *et al.*, 2017). Recent progress in gametophyte development in plants would further facilitate understandings of the molecular and biochemical mechanisms of these hybrid sterility genes in gamete development and reproductive barriers (Chen and Liu, 2014; Tucker and Koltunow, 2014; Zhang and Yang, 2014; Zhang and Yuan, 2014; Shi *et al.*, 2015).

Previous genetic works still have three major limitations in detection of hybrid sterility QTLs in terms of germplasm representation. The populations derived from either biparental or three-way crosses only allow detection of loci with allelic differentiation in the two parents, thus hindering the study in the extent of genetic architecture. Meanwhile, the resolution of the QTLs is limited owing to the number of recombinant plants and the density of molecular markers. In addition, understanding the genetic basis of hybrid incompatibility requires an investigation in the phenotypic effects from interacting locus pairs, which remains to be elucidated especially at the whole genome level. Therefore, genome-wide detection of hybrid incompatibility loci at single- and multi-locus level among representative varieties of rice germplasm would be essential to understand the complexity of mechanisms that lead to reproductive isolation.

In this study, we dissected the genetic basis of hybrid sterility by investigating the pollen fertility, embryo-sac fertility, and spikelet fertility in three crosses from Nipponbare (Nip) belonging to *japonica*, Zhenshan 97 (ZS97), and Minghui 63 (MH63), belonging to two different varietal groups of *indica* subspecies (Xie *et al.*, 2015), which cover the germplasm representation of cultivated rice. Comprehensive data from crosses using these three parents would explore in depth the genetic architecture of different reproductive barriers in terms of rice diversity, which could also explain the mechanism concerning decreasing of fertility in Shanyou 63, the most widely cultivated hybrids in China from cross between ZS97 and MH63. Interpreting these QTLs and interaction data in light of the genetics and evolutionary dynamics would facilitate the understanding in driving force of speciation, which also establish a framework for mapping and cloning of new hybrid sterility genes in future.

RESULTS

Fertility segregation in crosses of the three rice varieties

The three crosses between Nip, ZS97, and MH63 showed significant differences in hybrid fertility (Table S1). Spikelet fertility of the three parents and the hybrid from two *indica* varieties (ZS97 × MH63) was normal. However, the F_1 hybrids from two inter-subspecific crosses showed reduced spikelet fertility, 25.54% for the Nip × ZS97 hybrid and 63.04% for the Nip × MH63 hybrid.

Spikelet fertility of Nip × ZS97 F_2 individuals varied from 0% to 90.74%, with an average of 42.20% (Figure S1). Similarly, pollen fertility showed continuous distribution in F_2 population, ranging from 0% to 98.83%. A normal embryo-sac contained an egg apparatus with one egg cell and two

synergids at the micropylar end, two obvious polar nuclei locating above the egg apparatus, and a group of antipodal cells at the chalazal end. Seven types of abnormal embryo-sacs were observed in Nip × ZS97 F_2 population, including degeneration of embryo-sac, small embryo-sac, none polar nucleus, abnormal location of polar nuclei, single polar nucleus, and multiple polar nuclei locating either dispersive or centralized (Figure S1a). The embryo-sac fertility was estimated as the proportion of the embryo-sacs with normal structure, which varied from 1.30% to 98.51%.

The overall fertility in Nip × MH63 F_2 population was higher. The spikelet fertility varied from 0.40% to 92.62%, with the average of 52.56%. The average pollen fertility was 81.74%, ranging from 2.76% to 99.20%. The embryo-sac showed an average fertility of 67.53%, varying from 0% to 100.00%.

Although the average spikelet fertility of ZS97 × MH63 F_2 population was close to normal (80.17%), semi-sterile individuals were also observed. The fertility of the male and female gametes in F_2 individuals ranged from 43.25% to 99.51%, and 37.93% to 100.00%, respectively.

The spikelet fertility was significantly correlated with both embryo-sac fertility and pollen fertility in both Nip × ZS97 and Nip × MH63 populations (Table 1). No significant correlation was detected between spikelet fertility and either pollen or embryo-sac fertility in the population of ZS97 × MH63 cross.

Construction of the bin maps

Using restriction site-associated DNA sequencing (RAD-seq), we obtained 20.009 Gb sequence from 157 plants of F_2 Nip × ZS97, equivalent to 128.261 Mb per plant. Similarly, 25.088 Gb and 21.843 Gb sequences were obtained from 172, and 163 individuals of Nip × MH63 and ZS97 × MH63 populations, respectively. A total of 76 866, 80 290, and 44 049 high-quality single nucleotide polymorphisms (SNPs) were identified between the parents of three populations (Table S2). Three bin maps were constructed based on the recombination breakpoints detected in these F_2 populations, which generated 2668 bins in the Nip × ZS97 population, 2643 in the Nip × MH63 population, and 1850 bins in the ZS97 × MH63 population. The average genetic distance was 0.738 cM per bin in the Nip × ZS97 cross, 0.781 cM in the Nip × MH63 cross, and 1.125 cM in the ZS97 × MH63 cross, leading to total lengths of 2074.03 cM, 1942.32 cM, and 2081.05 cM of these three maps. The physical length of each bin in the Nip × ZS97 map ranged from 20 kb to 3.68 Mb with an average of 141.22 kb, 20 kb to 5.87 Mb in the Nip × MH63 map with an average of 139.90 kb, and 20.01 kb to 4.91 Mb in the ZS97 × MH63 map with an average 201.75 kb.

Table 1 Association between embryo-sac fertility, pollen fertility, and spikelet fertility using Pearson's correlation coefficient

Population	Association between traits	Correlation	*P*-value
Nip × ZS97	Pollen and embryo-sac fertility	0.094	0.273
	Embryo-sac and spikelet fertility	0.598*	1.243E-14
	Pollen and spikelet fertility	0.487*	9.043E-10
Nip × MH63	Pollen and embryo-sac fertility	0.349*	7.561E-06
	Embryo-sac and spikelet fertility	0.602*	2.200E-16
	Pollen and spikelet fertility	0.376*	7.803E-07
ZS97 × MH63	Pollen and embryo-sac fertility	−0.125	0.139
	Embryo-sac and spikelet fertility	0.174	0.046
	Pollen and spikelet fertility	−0.073	0.389

The star indicates the significant statistics. *$P < 0.01$.

Single-locus QTLs for embryo-sac, pollen, and spikelet fertility

A total of 27 QTLs for embryo-sac, pollen, and spikelet fertility in the three F_2 populations were detected using one-way analysis of variance (ANOVA) based on the bin genotypes of each locus with $-\log_{10}P$ threshold 2.5.

Three QTLs for embryo-sac fertility were detected on chromosomes 6, 8 and 12 in the F_2 population of Nip × ZS97 (designate as *Ef6*, *Ef8*, and *Ef12*) (Figure 1, Figure S2, Table 2). At the *Ef6* locus heterozygous genotype showed reduced embryo-sac fertility. However, the homozygote for Nip allele showed lower embryo-sac fertility at *Ef8*, while at *Ef12* the homozygote for the ZS97 allele displayed a significant reduction in embryo-sac fertility. The *Ef6* locus was mapped in chr06_bin55 with 42.462 kb in length spanning *S5* (Chen *et al.*, 2008; Yang *et al.*, 2012), thus the effect is highly likely caused by the *S5* locus. The *Ef12* was mapped to a region of 23.880-25.306 Mb containing *hsa1* (Kubo and Yoshimura, 2005; Kubo *et al.*, 2016a). The *Ef8* region spanned from 4.064 to 18.328 Mb where *hsa2* was located (Kubo and Yoshimura, 2005), thus the effect of *Ef8* might be due to *hsa2*.

Five QTLs were detected for pollen fertility. Two QTLs on chromosomes 5 and 12 (designate as *Pf5.1* and *Pf12*) caused significant pollen sterility in heterozygous genotypes. *Pf5.1* was located to a 3.370 Mb interval, which overlapped a previously identified locus *f5* (Wang *et al.*, 2006) or *Sb* (Li *et al.*, 2006). *Pf12* was mapped in a 1.688 Mb interval, which corresponded to a QTL for hybrid pollen sterility detected previously (Song *et al.*, 2005; Zhang *et al.*, 2011). The other three QTLs on chromosomes 1, 5 and 10 (designate as *Pf1*, *Pf5.2* and *Pf10*) respectively were not identified previously. The genotype homozygous for the

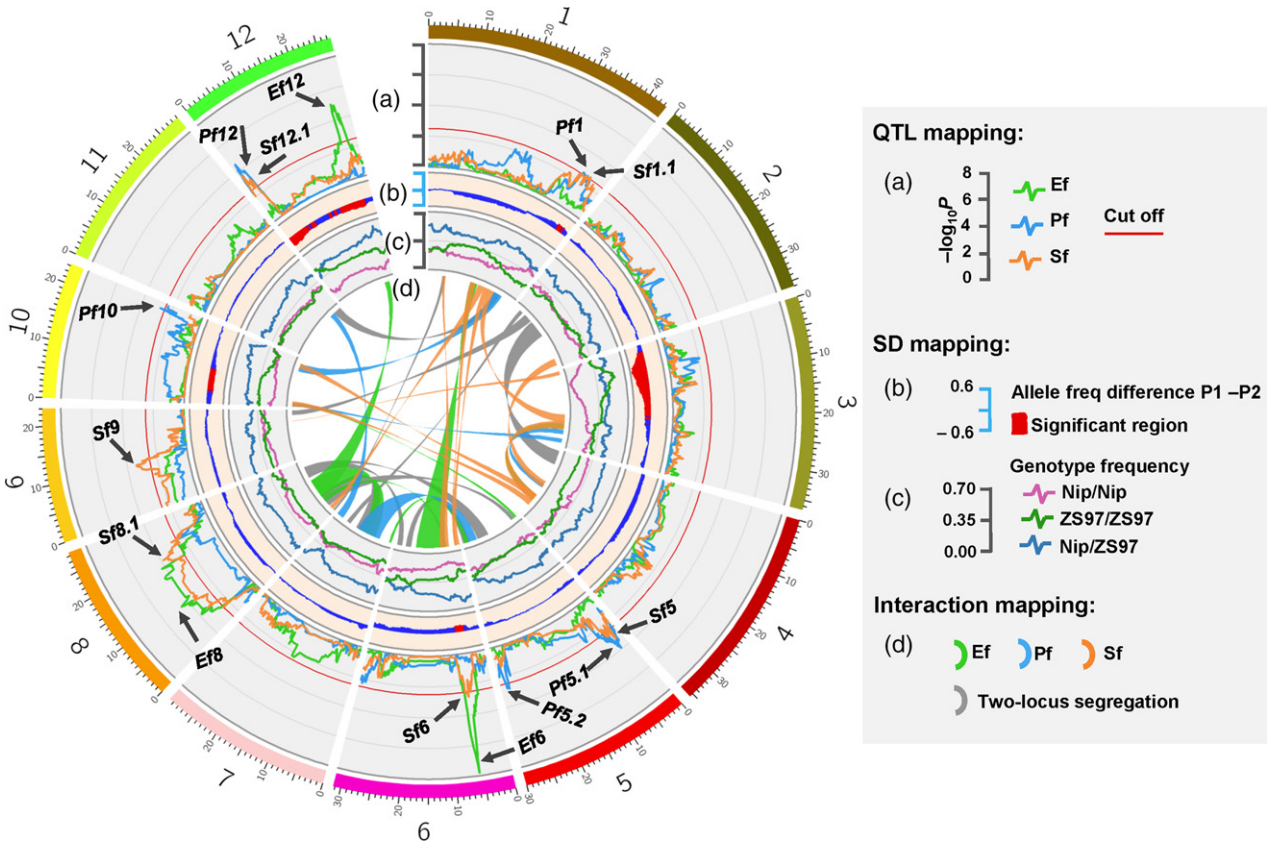

Figure 1. Distribution of QTLs and interacting loci in Nip × ZS97 F_2 population for embryo-sac, pollen, and spikelet fertility in hybrids.
ef, embryo-sac fertility; pf, pollen fertility; sf, spikelet fertility; SD, segregation distortion. Cut off suggests $-\log_{10}P$ higher than 2.5. P1 indicates Nip, and P2 indicates ZS97.

Nip allele showed significant reduced pollen fertility at *Pf1*, whereas at *Pf5.2* and *Pf10*, the homozygotes for the ZS97 alleles displayed a significant reduction in pollen fertility.

Five QTLs detected for spikelet fertility were likely due to the abortion of either female or male gametes, including *Sf1.1* caused by *Pf1*, *Sf5* caused by *Pf5.1*, *Sf6* caused by *Ef6*, *Sf8.1* caused by *Ef8*, and *Sf12.1* caused by *Pf12*. In addition, *Sf9*, a locus without a declared QTL for embryo-sac or pollen fertility, showed significant effect on spikelet fertility, possibly resulted from barriers in or post fertilization.

In the Nip × MH63 cross, we detected three QTLs (*Ef6*, *Ef8*, and *Ef12*) responsible for female gamete abortion on chromosomes 6, 8 and 12 (Figure 2, Figure S2, Table 2), all of them were also detected in the Nip × ZS97 cross. A major QTL on chromosome 6 was identified in chr06_bin27 (34.820 kb) containing *S5*, showing a major effect on embryo-sac fertility. The two minor QTLs *Ef8* and *Ef12* corresponded to the regions containing *hsa2* and *hsa1*.

Only one QTL *Pf8* for pollen fertility was identified on chromosome 8 where no QTL for pollen fertility has been reported, the genotype carrying homozygous Nip alleles showed significant reduced pollen fertility.

Six QTLs were identified for spikelet fertility. *Sf6* was the same locus as *Ef6* corresponding to *S5*, while the effect of *Sf12.2* was likely due to *Ef12* corresponding to *hsa1*. The *Sf8.2* locus spanning from 5.603 to 10.241 Mb likely resulted from the combined effects of *Ef8* and *Pf8*, which overlapped with the spikelet fertility locus *f8* (Wang *et al.*, 1998), *spf8* (Song *et al.*, 2005) or *qSS-8b* (Wang *et al.*, 2005). The genotype of *Sf4* carrying homozygous MH63 alleles showed significantly reduced spikelet fertility. This locus was mapped to a 2.370 Mb interval, which corresponded to a QTL *S9* for spikelet fertility detected previously (Zhao *et al.*, 2006). In addition, two new QTLs (*Sf1.2* and *Sf7*) were detected on chromosomes 1 and 7, and the homozygotes for the Nip alleles at both loci displayed significant reduction in spikelet fertility.

In the ZS97 × MH63 cross, *Ef1* for embryo-sac fertility and *Pf3* for pollen fertility were detected in the F_2 population (Figure 3, Figure S2, Table 2). At both loci genotypes homozygous for the ZS97 alleles showed significant disadvantage in fertility. A QTL *Sf3* for spikelet fertility was detected, and the homozygote for the MH63 allele displayed a significant reduction in fertility.

Table 2 QTLs for embryo-sac fertility, pollen fertility and spikelet fertility in three F_2 populations

Population P1 × P2	QTLs	Chr.	Bin[a]	$-\log_{10}P$	Interval (Mb)[b]	Additive[c]	Dominant	Var (%)	Overlapped locus[d]	Reference
Nip × ZS97	*Ef6*	6	chr06_bin55	7.810	5.417–6.276	−1.471	−21.822	23.694	*S5*	Chen *et al.* (2008); Yang *et al.* (2012)
	Ef8	8	chr08_bin76	4.562	4.064–18.328	−12.258	2.701	14.613	*hsa2*	Kubo and Yoshimura (2005)
	Ef12	12	chr12_bin148	5.057	23.88–25.306	15.863	−1.837	16.063	*hsa1*	Kubo *et al.* (2016a,b)
	Pf1	1	chr01_bin275	2.636	37.588–38.531	−9.026	3.350	8.481	*Sd1*	/
	Pf5.1	5	chr05_bin20	3.918	0.828–4.198	−3.650	−15.263	12.338	*f5/Sb*	Li *et al.* (2006); Wang *et al.* (2006)
	Pf5.2	5	chr05_bin213	2.859	28.064–28.664	2.192	13.984	9.162	/	/
	Pf10	10	chr10_bin145	2.769	21.605–21.719	8.139	7.728	8.886	/	/
	Pf12	12	chr12_bin15	4.498	0.165–1.853	1.332	−18.770	14.032	*pf12/qS12*	Song *et al.* (2005); Zhang *et al.* (2011)
	Sf1.1	1	chr01_bin296	2.619	38.673–39.774	−8.716	−3.490	8.426	*Sd1*	/
	Sf5	5	chr05_bin16	3.220	1.302–2.853	−3.434	−11.965	10.259	*f5/Sb*	Li *et al.* (2006); Wang *et al.* (2006)
	Sf6	6	chr06_bin56	2.838	5.756–6.39	−3.043	−10.866	9.098	*S5*	Chen *et al.* (2008); Yang *et al.* (2012)
	Sf8.1	8	chr08_bin105	3.866	5.108–21.585	−10.899	0.897	12.185	*hsa2*	Kubo and Yoshimura (2005)
	Sf9	9	chr09_bin34	3.395	9.339–11.545	−8.811	−3.307	10.785	/	/
	Sf12.1	12	chr12_bin8	3.680	0–1.768	1.171	−15.367	11.635	*pf12/qS12*	Song *et al.* (2005); Zhang *et al.* (2011)
Nip × MH63	*Ef6*	6	chr06_bin27	17.913	4.916–6.107	−1.948	−24.958	42.728	*S5*	Chen *et al.* (2008); Yang *et al.* (2012)
	Ef8	8	chr08_bin60	3.676	4.641–15.704	−9.178	3.635	10.808	*hsa2*	Kubo and Yoshimura (2005)
	Ef12	12	chr12_bin147	2.687	26.457–27.465	7.725	1.577	8.020	*hsa1*	Kubo *et al.* (2016a,b)
	Pf8	8	chr08_bin92	5.404	8.664–20.583	−8.307	−3.369	14.832	/	/
	Sf1.2	1	chr01_bin175	2.960	23.753–25.688	−8.014	−5.232	8.471	/	/
	Sf4	4	chr04_bin68	3.268	16.657–19.207	10.133	−5.135	9.310	*S9*	Zhao *et al.* (2006)
	Sf6	6	chr06_bin27	5.947	4.872–6.465	−4.256	−16.642	16.292	*S5*	Chen *et al.* (2008); Yang *et al.* (2012)
	Sf7	7	chr07_bin130	2.624	22.998–23.177	−8.801	−2.911	7.546	/	/
	Sf8.2	8	chr08_bin67	15.677	5.603–10.241	−18.279	14.615	37.425	*qSS–8b/spf8*	Wang *et al.* (2005); Song *et al.* (2005)
	Sf12.2	12	chr12_bin151	4.431	26.417–27.532	11.377	−3.837	12.410	*hsa1*	Kubo *et al.* (2016a,b)
ZS97 × MH63	*Ef1*	1	chr01_bin33	3.210	3.518–4.579	−2.407	1.896	10.088	/	/
	Pf3	3	chr03_bin22	2.949	0.396–3.477	−3.451	3.324	9.061	/	/
	Sf3	3	chr03_bin175	3.279	32.849–35.142	3.985	1.174	10.225	/	/

[a]Bins showing the most significant effect.
[b]1.5-$\log_{10}P$ support interval of the QTL.
[c]A negative additive effect indicates the allele from P1 showing reduced fertility.
[d]The confidence interval is overlapped with or close to other locus.

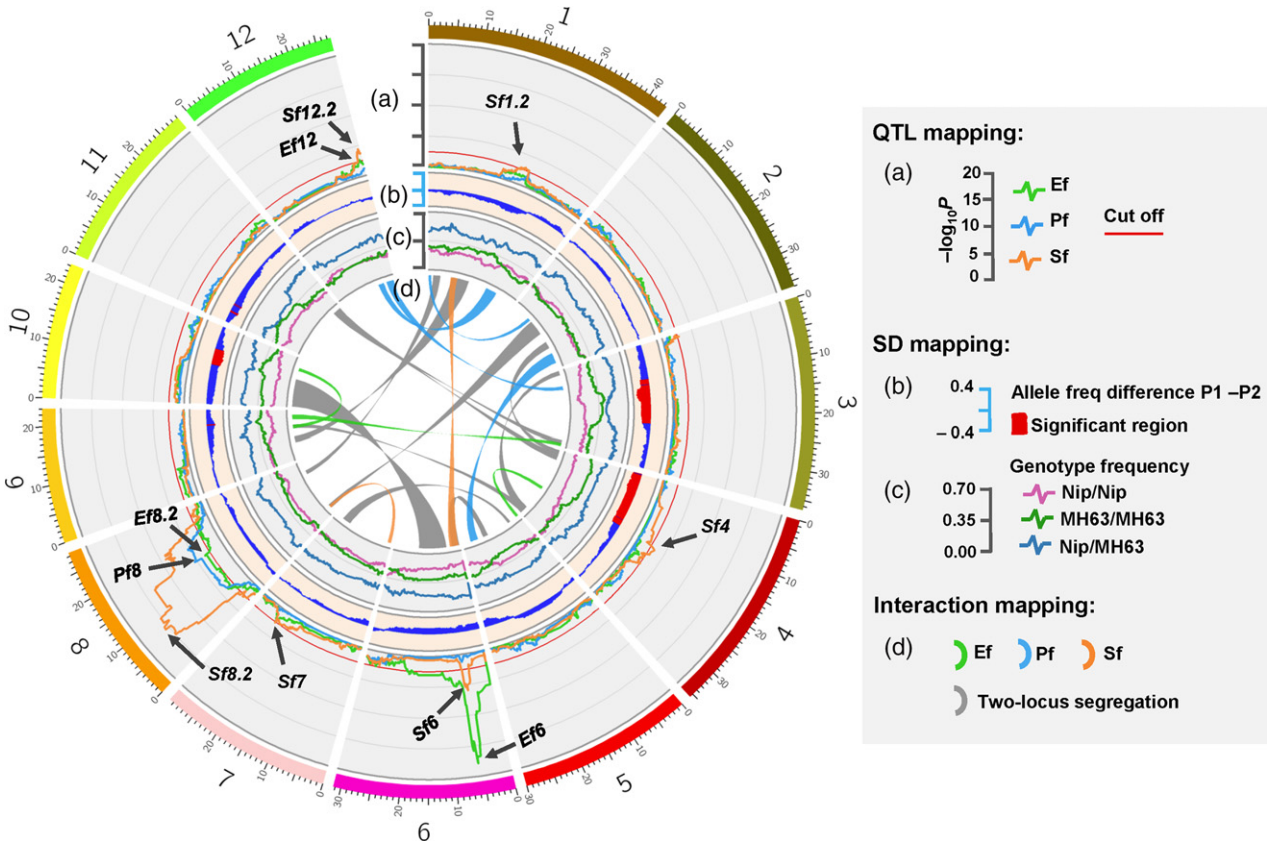

Figure 2. Distribution of QTLs and interacting loci in Nip × MH63 F_2 population for embryo-sac, pollen, and spikelet fertility in hybrids.
ef, embryo-sac fertility; pf, pollen fertility; sf, spikelet fertility; SD, segregation distortion. Cut off suggests $-\log_{10}P$ higher than 2.5. P1 indicates Nip, and P2 indicates MH63.

Prediction of candidate genes for QTLs with respect to fertility

The high-quality reference genome sequences available for all three parents allowed the prediction of the candidate genes of the mapped QTLs. The *Ef6*/*S5* locus was located in chr06_bin55 (42.462 kb) of the Nip × ZS97 cross and in chr06_bin27 (34.820 kb) of the Nip × MH63 cross (Figure S3). Overlapping region of these two bins contained LOC_Os06g11010 (*S5-ORF5*). The *ORF5* alleles from ZS97 and MH63 were identical and both differed by two nucleotides from the Nip allele, being characteristic of the functional mutations. Sequence comparison in mapped bins also revealed a 13-bp insertion in LOC_Os06g10990 (*S5-ORF3*) and an 11-bp deletion in LOC_Os06g11000 (*S5-ORF4*), respectively, in ZS97 and MH63 compared to Nip. Thus *S5-ORFs3-5* were located exactly in these two bins, which were well in accordance with our previous map-based cloning result (Chen *et al.*, 2008; Yang *et al.*, 2012).

Such comparative analysis was applied to predict candidate genes for QTLs with $-\log_{10}P > 4$. The *Ef8* locus was mapped in chr08_bin76 in the Nip × ZS97 population, embryo-sac fertility of the Nip allele of this locus was lower (Figure S4, Table S3). Sequence alignment between Nip and ZS97 revealed a 41.7 kb insertion (Chr 8: 9242940-9284596) containing 20 Nip-specific genes encoding 10 chloroplast proteins, two cytochromes, an oxidoreductase, a RNA polymerase subunit, a photosystem II reaction center protein H, a ATP-dependent Clp protease proteolytic subunit, a apocytochrome f precursor, and three uncharacterized or hypothetical proteins. Thus this insertion might be associated with some undesirable effect on embryo-sac fertility.

Pf12 (chr12_bin15, 139.95 kb) showed effect on hybrid pollen fertility in the Nip × ZS97 cross (Figure S5, Table S3). Three Nip-specific genes (LOC_Os12g02560, LOC_Os12g02570, and LOC_Os12g02750) and eight ZS97-specific genes (ZS12t0025100-5400, ZS12t0025900-6100, and ZS12t0027100) were identified in this region, providing a list of the likely candidates. ZS12t0027100 encoded a bis (5′-adenosyl)-triphosphatase, and other genes were annotated as expressed proteins or hypothetical proteins. We also found significant differences between Nip and ZS97 in predicted proteins of six genes due to either SNPs or InDels in coding regions, leading to a variance in splicing site in LOC_Os12g02589 and LOC_Os12g02620, a frame shift in LOC_Os12g02630, LOC_Os12g02640, and LOC_

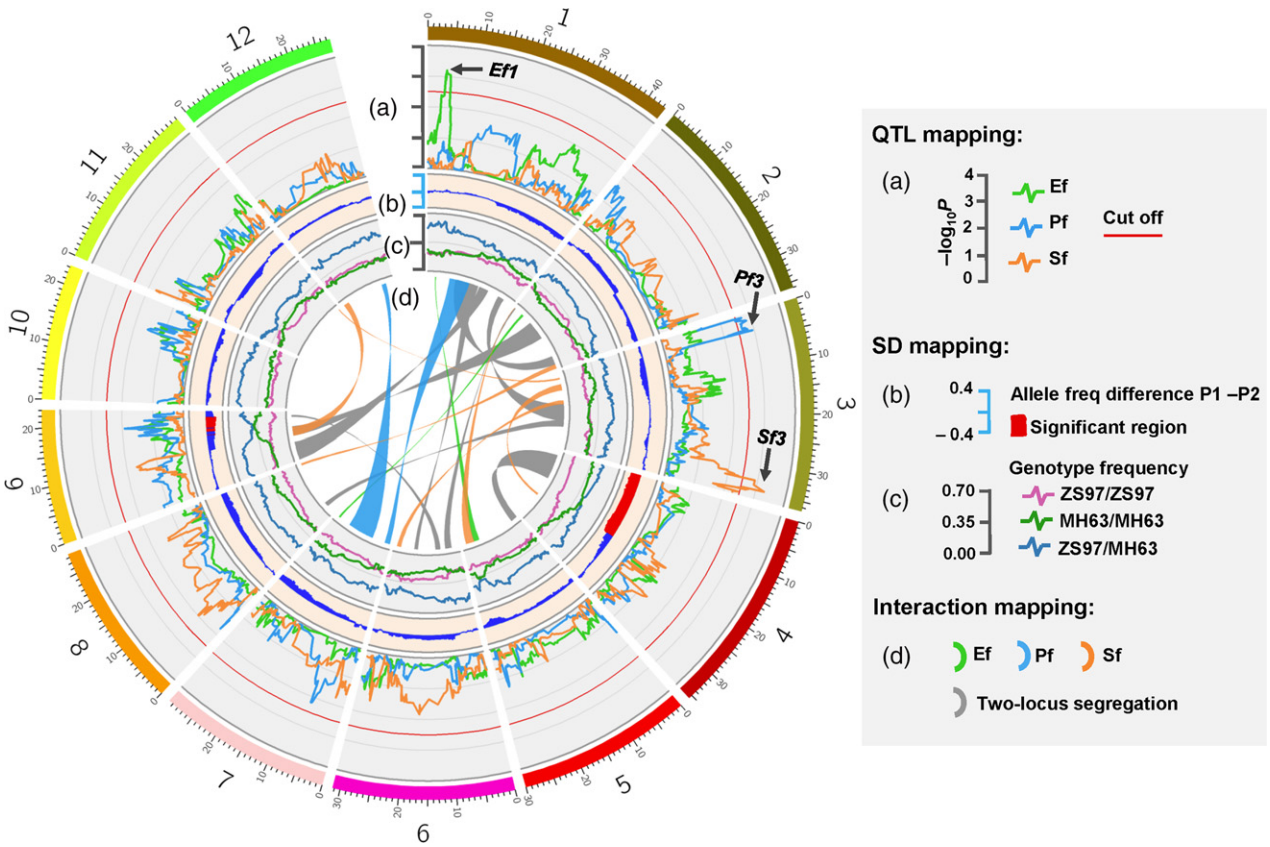

Figure 3. Distribution of QTLs and interacting loci in ZS97 × MH63 F_2 population for embryo-sac, pollen, and spikelet fertility in hybrids.
ef, embryo-sac fertility; pf, pollen fertility; sf, spikelet fertility; SD, segregation distortion. Cut off suggests $-\log_{10}P$ higher than 2.5. P1 indicates ZS97, and P2 indicates MH63.

Os12g02720, and a premature termination in LOC_Os12g02760. Based on expression analysis in these six genes, we found preferential expression of LOC_Os12g02620, LOC_Os12g02640, LOC_Os12g02720, and LOC_Os12g02760 in panicles or stamen. In addition, the transcripts of four genes, LOC_Os12g02490, LOC_Os12g02530, LOC_Os12g02540, and LOC_Os12g 02550, also accumulated in panicles or stamen, and these genes had nonsynonymous mutations relative to the ones in ZS97. However, further investigation is needed to identify the real cause.

Pf8 (chr08_bin92, 99.62 kb) was identified in the Nip × MH63 cross inducing pollen fertility (Figure S6, Table S3). When transposons/retrotransposons and the genes with identical CDS/protein sequences in two parents were ignored, sequence alignment identified six genes in MH63 (MH08t0329000, MH08t0329100, MH08t0329200, MH08t0329500, MH08t0330600, MH08t0330700) showing frame shift variations relative to the ones in Nip due to InDels in coding regions. These genes encoded an F-box domain containing protein, a pyrophosphate fructose 6-phosphate 1-phosphotransferase subunit alpha, and four hypothetical proteins. In addition, two insertions (Chr 8: 13844849-13868136, 23.3 kb; Chr 8: 13873820-13878985, 5.2 kb) expanded the segment in MH63, which added six MH63-specific genes in this bin region (MH08t0329900, MH08t0330300, MH08t0330400, MH08t0330500, MH08t0330800, and MH08t0331000). The predicted products included a cell wall protein, a glutathione S-transferase T3, a serine/threonine-protein phosphatase 7 long form homolog, and three hypothetical proteins. We also noticed that nonsynonymous mutations between LOC_Os08g25710 (in Nip) and MH08t0329400, a gene that was annotated as a cellulose synthase-like protein D3. It was specifically expressed in stamen at 1 day before flowering, which might be involved in gamete development. These 13 genes provided possible candidates for reduced pollen fertility in the Nip × MH63 F_2 population.

Segregation distortion and the effects on fertility in each population

In the F_2 population of Nip × ZS97, seven regions were detected showing significant deviations ($P < 0.001$) from the expected allelic (1:1) and genotypic (1:2:1) frequencies (Figure 1 and Table S4). In five of the seven regions (*Sd3*, *Sd6*, *Sd10.1*, *Sd12.1*, and *Sd12.2*), the alleles from Nip were

in deficiency, also leading to deviations in genotypic frequencies. Five segregation-distorted regions (*Sd1*, *Sd3*, *Sd6*, *Sd12.1*, and *Sd12.2*) caused significantly reduced fertility in heterozygotes ($P < 0.05$). The distorted region on chromosome 3 (*Sd3*) was likely due to the hybrid male sterility locus *Sc* (Yang *et al.*, 2004). The pollen fertility of the heterozygote (60.38%) was lower relative to the homozygotes (71.02% and 71.06%), although no QTL for hybrid sterility was detected with the threshold applied. Three segregation-distorted regions (*Sd6*, *Sd12.1*, and *Sd12.2*) coincided with the QTLs for hybrid sterility. Significantly reduced fertility was detected in heterozygotes at *Sd6* and *Sd12.1*, due to the effects of *Ef6* (*S5*) and *Pf12* causing hybrid female and male sterility, respectively. The embryo-sac fertility of the homozygote for ZS97 allele at *Sd12.2* was also significantly lower than the homozygote for Nip allele, likely due to the effect of *Ef12* (*hsa1*). In addition, *Sd1*, another locus corresponding to the newly identified QTL *Pf1*, also showed significant difference in pollen and spikelet fertility between heterozygote and homozygotes.

Seven regions were detected as showing segregation distortion in the Nip × MH63 population, all of which led to deficiencies in the Nip alleles (Figure 2 and Table S4). The distorted region on chromosome 3 was also observed, suggesting that *Sc* has a sizable effect in reducing the Nip gametes, although the pollen fertility was normal in heterozygote. The other six regions were different from those in the Nip × ZS97 cross. None of the seven regions were associated with lower fertility in the heterozygotes.

We found distorted regions on chromosomes 4 and 9 in the ZS97 × MH63 cross (Figure 3 and Table S4), both of which were also detected in the Nip × MH63 cross; significant preferential transmission of the MH63 alleles occurred in both regions, although the fertility of the heterozygotes at *Sd4* and *Sd9* was close to normal.

The effects of two-locus interactions on fertility

Two-way ANOVA of two-locus genotypes formed of all possible pairs of the bins in the whole genome identified a total of 22 interactions that had significant effects on fertility ($P < 0.001$) in the F_2 population of Nip × ZS97 cross (Figure 1, Table S5). Seven of the 22 interactions had effects on embryo-sac fertility, five interactions affecting pollen fertility, and 10 interactions influencing spikelet fertility.

In 12 of the interactions, the complementary genotypes in each of the two-locus pairs, i.e. homozygous for Nip alleles at one locus while homozygous for the ZS97 alleles at the other, led to significant reduction in fertility. For example, the genotype with homozygous Nip alleles at chr08_bin79 (interval: 8.775–18.328 Mb) corresponding to *hsa2* and homozygous ZS97 alleles at chr12_bin144 (interval: 23.027–25.213 Mb) corresponding to *hsa1* caused abortion of female gametes, similar to the previous result (Kubo and Yoshimura, 2005; Kubo *et al.*, 2016a). This kind of complementary genotypes would contribute to fertility reduction in the F_1 hybrid between the two parents.

In six of the interactions, genotypes at one locus would affect the fertility of the heterozygote at the other locus. For example, homozygote of Nip alleles at chr08_bin78 (interval: 8.659–18.328 Mb) corresponding to *hsa2* led to a significant reduction in embryo-sac fertility when chr07_bin90 (interval: 8.311–15.397 Mb) was heterozygous compared to the homozygotes. In another case, the homozygote of ZS97 alleles at chr01_bin73 (interval: 7.068–8.132 Mb) greatly reduced the spikelet fertility of the heterozygote at chr06_bin58 (interval: 5.12–9.452 Mb) containing *S5*. This kind of interactions is also relevant to hybrid sterility.

In the remaining four interactions, genotypes that are homozygous for alleles from the same parents at both loci produced lower fertility. For example, in the locus pairs chr01_bin255/chr10_bin114, genotype with homozygous ZS97 alleles at both loci caused significant reduction in pollen fertility. Whereas in the other three locus pairs (chr01_bin160/chr06_bin98, chr01_bin156/chr06_bin22, chr01_bin221/chr06_bin12), genotypes of homozygous Nip alleles at both loci showed significant reduction in fertility.

Nine significant interactions were detected in the Nip × MH63 cross, including three locus pairs showing effects on embryo-sac fertility, 4 on pollen fertility, and two on spikelet fertility (Figure 2, Table S5). However, none of these interactions were in common with the ones detected in the Nip × ZS97 population. In four of the interactions (chr03_bin283/chr09_bin132, chr04_bin90/chr05_bin9, chr01_bin98/chr06_bin11, chr07_bin12/chr08_bin32), significant reduction in fertility occurred in genotypes heterozygous at one locus and homozygous at the other. In three of the interactions (chr09_bin86/chr10_bin93, chr01_bin4/chr02_bin73, chr03_bin67/chr12_bin139), genotypes with homozygous Nip alleles at both loci led to significant reduction in fertility, whereas in the remaining two interactions (chr01_bin219/chr12_bin101, chr02_bin235/chr05_bin180), genotypes with homozygous MH63 alleles at both loci caused significant fertility reduction.

Ten significant interactions were detected in the ZS97 × MH63 cross, including two showing effects on embryo-sac fertility, 2 on pollen fertility, and six on spikelet fertility (Figure 3, Table S5). In five interactions, the complementary homozygous genotypes, i.e. homozygous ZS97 alleles at one locus and homozygous MH63 alleles at the other, showed significant reduction in fertility. In four of the interactions, genotypes with homozygous alleles at both loci from the same parents showed reduced fertility. In the remaining case, homozygote for the MH63 allele at one locus (chr11_bin169, interval: 28.646–29.021 Mb) would affect the spikelet fertility of the heterozygote at the other (*Pf3*, chr03_bin20, interval: 2.642–3.506 Mb). But

overall the effects of these interactions on fertility reduction were much smaller than in the other two crosses.

Deviations of genotype frequencies of two-locus combinations from expectations based on independent segregation of individual loci (two-locus segregation distortion) may provide another measure for interaction between different loci (Corbett-Detig *et al.*, 2013). Such significant interactions ($P < 0.001$) were detected in all three F_2 populations, including 69 locus pairs in the Nip × ZS97 cross, 61 in the Nip × MH63 cross, and 52 in the ZS97 × MH63 cross (Figures 1, 2 and 3, Table S6). Combined with two-way ANOVA for embryo-sac, pollen, and spikelet fertility, 14 segregation-distorted two-locus combinations also showed significant effects on fertility in the Nip × ZS97 cross, 9 in the Nip × MH63 cross, and 12 in the ZS97 × MH63 cross. In addition, three interacting two-locus combinations were detected in both Nip × ZS97 and Nip × MH63 populations that involved known loci for hybrid sterility. The first case involved the segregation-distorted locus pairs, chr02_bin136/chr05_bin62 in the Nip × ZS97 cross and chr02_bin108/chr05_bin95 in the Nip × MH63 cross, which is in agreement with a previous finding that the region on chromosome 2 contained an factor epistatic to *f5/Sb/S24* on chromosome 5 (Kubo *et al.*, 2016b). In the second case, chr11_bin174/chr12_bin153 in the Nip × ZS97 cross and chr11_bin191/chr12_bin148 in the Nip × MH63 cross also caused abortion of female gametes, in which the locus on chromosome 12 in both populations corresponded to *hsa1* on chromosome 12 (Kubo *et al.*, 2016a). The third case occurred in two-locus genotypes of chr05_bin41/chr08_bin62 in the Nip × ZS97 cross and chr05_bin17/chr08_bin62 in the Nip × MH63 cross, suggesting possible interaction between *hsa2* and the region on chromosome 5.

DISCUSSION

The genetic architecture of reproductive isolation in rice based on three representative inter and intra-subspecific crosses

Hybrid sterility between *indica* and *japonica* subspecies is complicated due to the contribution of multiple genetic elements regulating various types of reproductive barriers. Although approximately 30 loci are recognized so far responsible for the fertility of *indica-japonica* hybrids, the data is still limited due to a wide range of differentiation in rice germplasms (Ouyang, 2016). By taking advantage of three pairwise crosses from representative varieties of *japonica* and two *indica* varietal groups with well sequenced reference genomes, 43 single-locus QTLs and 223 two-locus combinations are detected for showing effects in fertility at the whole genome level, which provided an extensive archive for investigating the genetic basis of reproductive barriers in rice (Table S7). The three parents of the mapping populations used in this study represent a wide range of diversity in the cultivated rice germplasms. Nip belongs to temperate *japonica*, while ZS97 and MH63 from two *indica* varietal groups are also used as the parents of Shanyou 63, the most widely cultivated hybrid in China. Therefore, the hybrid sterility loci identified for inter-subspecific and intra-subspecific fertility in these three crosses have comprehensive implications for understanding the genetic architecture and diverse variation of reproductive isolation in rice.

In total, 27 loci were detected for embryo-sac fertility, pollen fertility, and spikelet fertility in three F_2 populations, including 14 loci showing effects in the Nip × ZS97 cross, 10 in the Nip × MH63 cross, and three in the ZS97 × MH63 cross. Along with five consequential segregation distortion regions affecting the fertility of either homozygotes or heterozygote, these loci described a framework for reproductive barriers in rice. Ten QTLs were newly detected for fertility that would provide fresh understanding for reproductive isolation, including four loci for the Nip × ZS97 cross, three for the Nip × MH63 cross, and three in the ZS97 × MH63 cross. Three loci (*S5*, *hsa1*, and *hsa2*) responsible for embryo-sac fertility were detected in both Nip × ZS97 and Nip × MH63 crosses, whereas QTLs for pollen fertility were not in common between the two crosses, suggesting that these loci might be responsible for fertility variation in crosses from various rice germplasms. Three QTLs responsible for embryo-sac, pollen, and spikelet fertility were identified in the ZS97 × MH63 cross. Two segregation distortion regions (*Sd4* and *Sd9*) were also identified, leading to distortion of gamete transmission in hybrids within *indica* subspecies. These loci for hybrid fertility might have distinct molecular and biochemical functions, and their mechanisms are in close relation with the involved pathway in male and female gametophyte development. These genes might result in reduced fertility in hybrids due to abnormal reproductive process in various stages, from the formation of sporophytic anther or ovule, and subsequent meiosis process, to the development of mature pollen or embryo-sac (Chen and Liu, 2014; Tucker and Koltunow, 2014; Zhang and Yang, 2014; Zhang and Yuan, 2014; Shi *et al.*, 2015). The genes for hybrid fertility might interplay with other reproductive developmental genes and constitute the network of gamete development thus leading to partial reproductive barriers in hybrids.

We failed to detect the effect of two cloned hybrid sterility loci/gene pairs, *Sa* (Long *et al.*, 2008) and *DPL1/DPL2*, in our populations. Sequence comparison revealed that ZS97 carried the neutral haplotype of $SaF^{+}SaM^{-}$ at *Sa* locus, while the *SaF* and *SaM* genes were lost in MH63. Thus both ZS97 and MH63 showed compatibility to Nip (haplotype of $SaF^{-}SaM^{-}$). Meanwhile, both *DPL1* and *DPL2* did not show functional difference in Nip, ZS97, and MH63,

which was in consistent with our result. Other discrepancies in QTLs were also observed in this study as compared with previous reports (Ouyang *et al.*, 2009). This might be due to either lack of variation or environmental fluctuations.

Detection of two-locus interaction of all possible pairs of the bins at the whole genome level also provided the first glimpse for investigating the contribution of epistasis in reproductive isolation in rice. Forty-one-two-locus interactions for three types of reproductive barriers were detected, including 12 locus pairs showing effects in embryo-sac fertility, 11 in pollen fertility, and 18 in spikelet fertility. In addition, the fertility was influenced by 35 two-locus combinations showing significant deviation from the expected genotype frequencies. Interaction might exist between different QTLs involved in hybrid incompatibility. We have detected interaction between *hsa2* and *hsa1* underlying female fertility, which was in agreement with interactions between *hsa1*, *hsa2*, and *hsa3* identified in previous study (Kubo and Yoshimura, 2005). Duplicated interacting loci also caused genetic incompatibility in rice (Mizuta *et al.*, 2010; Yamagata *et al.*, 2010). In addition, epistatic interactions between hybrid sterility genes and other unlinked loci also contributed significantly to reproductive barriers in rice. Two independent genetic pathways of *EFS-S24* (*f5*/*Sb*) and *INK-S35* (*Sd*/*f1*) were identified in *indica-japonica* cross, both of which were responsible for hybrid pollen fertility (Kubo *et al.*, 2016b). The effect of *S24* on hybrid pollen sterility would be suppressed by the *indica* allele of *EFS* (Kubo *et al.*, 2011), whereas the *indica* allele of *INK*, which located adjacent to *S24*, was required to induce *S35*-dependent pollen sterility. The interacting locus pairs and their phenotypic effects observed in three F_2 populations in our study confirmed significant contribution of complex interactions in reproductive isolation.

The genes responsible for hybrid fertility contribute to reproductive isolation and evolution of different rice populations

The origin and divergence of hybrid fertility and segregation distortion loci have comprehensive implications for understanding the evolution and differentiation of different rice populations. Three loci for hybrid female fertility were shared among inter-subspecific crosses between *japonica* and different varietal groups of *indica*, suggesting genetic divergence at the early stage of speciation. However, the loci for hybrid male fertility were different among rice lineages, possibly behaved as a more rapidly evolving form of incompatibility. Consistently, most hybrid male sterility loci are not shared in two mouse subspecies pairs (White *et al.*, 2012). Therefore, the genetics and evolutionary patterns of hybrid male and female fertility loci differ substantially along with speciation. Two segregation distortion regions, *Sd4* and *Sd9*, might have evolved along the *IndII* branches, leading to distortion of gamete transmission within *indica* subspecies. Three new loci responsible for hybrid fertility were identified between two *indica* varieties, suggesting that the two varietal groups of *indica* might have diverged in the recent past, which was likely due to the combined effects of ongoing reproductive isolation and rice breeding process. Our genetic data along with information on phylogenic relationships of subspecies and varietal groups provide understandings in evolutionary dynamics of reproductive isolation and speciation.

Implications in rice genetic improvement

Our finding might have significant implications in rice genetic improvement. Based on present result, the genetics of hybrid sterility loci differs substantially in diverse inter-subspecific crosses. Varieties from either *IndI* or *IndII* would be incompatible with *japonica* rice in hybrid female fertility, whereas different *indica* varieties might be able to produce semi-fertile male gametes at different loci for hybrid pollen fertility. Utilization of these single-locus hybrid sterility loci for development of '*japonica*-compatible *indica* lines' through introgression would produce fertile hybrids when crossed with targeted *japonica* lines. Pyramiding of multiple wide-compatibility alleles at these QTLs into elite parents with strong heterosis would also facilitate breaking the reproductive barriers.

In this study, we also quantified the importance of two-locus interactions on fertility in the whole genome, and found that three types of interaction contributed significantly to segregation of fertility in the F_2 populations. Genetic dissection identified that 17 interactions with complementary genotypes resulted in fertility reduction in the hybrid; one locus pairs corresponded to *hsa2* and *hsa1* in previous result (Kubo and Yoshimura, 2005). Genotypes with homozygous alleles at both loci from the same parents led to reduced fertility at 13 two-locus pairs, and 11 interactions were relevant to hybrid sterility with the homozygote at one locus and the heterozygote at the other locus. These two-locus interactions result in effects of reducing the fertility in the F_2 populations, and such unfavorable consequence in fertility might be improved via breaking of the deleterious allelic combinations.

In particular, the population from ZS97 $\times$ MH63 cross exhibited a wide range segregation in embryo-sac fertility, pollen fertility, and spikelet fertility. Genotypes homozygous for the MH63 alleles at *Ef1* and *Pf3*, and homozygote for the ZS97 allele at *Sf3* contributed to increasing fertility the hybrids. Further improvement in the fertility of ZS97 $\times$ MH63 cross would be possible using the better alleles at single-locus level and allelic combinations at multi-locus level, which might have general implications for other *indica*–*indica* hybrids.

EXPERIMENTAL PROCEDURES

Rice materials and growth conditions

Three F_2 populations, with the plants of 157, 172, and 163, respectively, were obtained from crosses of Nip × ZS97, Nip × MH63, and ZS97 × MH63 followed by self-fertilization of the F_1 hybrids. Nip belongs to *japonica* subgroup (*O. sativa* L. ssp. *japonica*), while ZS97 and MH63 are two elite rice lines of *indica* subspecies (*O. sativa* L. ssp. *indica*). All plants were field grown in the rice growing season in Wuhan in 2014.

Fertility examination

Each plant in F_2 population was used to investigate the embryo-sac fertility, pollen fertility, and spikelet fertility. During flowering time, florets from the middle and upper parts of three panicles per plant were sampled for examining embryo-sac and pollen fertility. The florets were fixed in FAA (formaldehyde: acetic acid: 50% ethanol = 5:6:89) for 24 h at room temperature, after which they were washed with 50% ethanol and stored in 70% ethanol at 4°C. Approximately 150 ovaries were investigated using the whole stain-clearing method (Song *et al.*, 2005).

Pollen fertility was examined with the I_2-KI staining method, using mixed pollen grains from six to eight florets from each plant. Three independent microscope fields were observed, with at least 100 pollen grains per field. Pollen fertility was estimated from the mean proportion of the stained pollen.

Three panicles per plant were harvested for examination in spikelet fertility, which was scored as the percentage of filled grains.

DNA extraction and genotyping using RAD-Seq

Genomic DNA was extracted from fresh flag leaves (Doyle and Doyle, 1987), and 3 μg DNA extracted from each individual was subjected to BGI (Huada Genomics Institute Co. Ltd, China) for RAD-Seq. RAD library construction, sample indexing and pooling were performed according to the Manufacturer's instructions (Illumina). The DNA library was sequenced using the Illumina HiSeq 2000 platform, and a total of 45.08 GB paired-end 100-bp reads sequence data were obtained. All reads were mapped against the rice reference genome (Nipponbare MSU release version 7.1) using the software SOAP2 (version 2.20) (Li *et al.*, 2009). SNP calling was performed with CaSFS (version 0.983), and input data was prepared by SAMtools (version 0.1.8) based on the Bayesian estimation of site frequency at every site (Duan *et al.*, 2013).

Bin map construction and QTL scanning

The SNP data identified in three F_2 populations were converted into another format to simplify the genotype calling analysis, based on the reference genome sequences from three parents. The Nip alleles were coded as 'a', and the SNPs from ZS97 or MH63 were coded as 'b' in Nip × ZS97 and Nip × MH63 populations. The ZS97 alleles were coded as 'a', and the SNPs from MH63 were coded as 'b' in ZS97 × MH63 population. The heterozygotes were coded as 'h', while missing data were coded as '-'. Based on SNP locations on the reference genome and genotypes identified, a modified sliding window approach was adopted, in order to evaluate a group of consecutive SNPs for genotyping and recombination breakpoint determination (Huang *et al.*, 2009). For each scanning, a window of 15 SNPs without missing data was used for genotyping calling. An a/b ratio of 12:3 or higher was recognized as 'a', 3:12 or lower as 'b' and anything in between as 'h'. The crossover of transforming between a homozygous genotype and a heterozygous genotype in F_2 population was defined between two adjacent blocks with different genotypes. Another type of crossover was transforming between the homozygous genotype from the two parents. There existed several temporary 'h' in this type of crossover before transformation into another homozygous genotype, and the third changed SNP was chosen for determination of the breakpoint. The interval between two adjacent crossovers in the entire F_2 population was defined as a recombination bin. After initial construction of the bin map, genotyping errors were identified using R/qtl package function *calc.errorlod* with error.prob = 0.001. Genotypes with error LOD scores >4 were set missing and were imputed using R/qtl package function *fill.geno* with the 'argmax' method (Broman *et al.*, 2003). The genetic linkage map based on the bins was constructed using the R/qtl package function *est.map* with Kosambi map method and error.prob = 0.0001.

QTL analysis was performed using one-way ANOVA to test the association of the genotype of each bin with the observed trait. A 1.5 $-\log_{10}P$ drop support interval was used for each QTL. The variation explained by each bin was determined using *lm* function of R fitted liner model.

Identification for region with segregation distortion

To investigate the deviation of each bin from the Mendelian segregation (1:1 radio for allele frequency and 1:2:1 for genotype radio), a chi-squared test was performed using the R function *chisq.test*. The adjacent bins showing segregated distortion ($P < 0.001$) were defined as the segregated distortion region.

Test for independence of marker segregation

To identify the independence of marker segregation, a chi-square test was employed to scan the independence of all bin pairs along the whole genome. The bin pairs showing significant association ($P < 0.001$) indicated possible two-locus interactions. The *P*-value for the most significant bin pairs in two associated chromosome fragments was assigned to the 2D significant interaction regions, and $-\log_{10}P$ 2-drop support interval was used as the confidence interval for every one-dimensional of the chromosome segment.

Identification for two-locus epistatic interactions of fertility-related traits

Two-way ANOVA was performed between all bin pairs across the whole genome, in order to detect epistatic interactions for embryo-sac, pollen, and spikelet fertility. Permutation tests were used to define the genome-wide significant *P*-value threshold. For each trait of the three F_2 populations, we reshuffled the phenotype 1000 times; 1000 pairs of bins were random selected at each time to perform two-way ANOVA analysis, and the minimum *P*-value was recorded. Among the 1000 times permutation, the 50th minimum *P*-value was taken as the genome-wide threshold. The epistatic interaction regions were detected as described above.

Haplotype analysis for bins showing the most significant effect in fertility QTLs

The SNP data from rice accessions of *O. sativa* was download from the database of RiceVarMap (http://ricevarmap.ncpgr.cn/) (Xie *et al.*, 2015; Zhao *et al.*, 2015). The SNP data from 98 accessions of *IndI*, 105 accessions of *IndII*, and 91 accessions of temperate *japonica* (*TeJ*) was used for haplotype analysis with respect to the bins showing the most significant effect in fertility QTLs.

Sequence and expression analysis for candidate genes

We compared the sequence of annotated genes within the bins showing the most significant effect in hybrid sterility, using the reference genome of Nip (http://rice.plantbiology.msu.edu), ZS97, and MH63 (Zhang *et al.*, 2016).

Expression profiles of candidate genes in 29 tissues for ZS97 and MH63 were extracted from the CREP database (http://crep.ncpgr.cn). The microarray data have been submitted into the NCBI Gene Expression Omnibus (GEO) under the accession number of GSE19024 (Wang *et al.*, 2010). After normalization and variance stabilization, the average signal value of two biological replicates for each sample, except for samples 7, 8, 17, 18 and 19 (three biological replicates and two technical replicates) was used for analysis. Wherever more than one probe set was available for one gene, a randomly selected probe set was used for analysis. Quantitative differences in gene expression levels were plotted using heatmap.2 function of R library 'gplots'. For data analysis, Student's *t*-test was performed to identify differentially expressed genes. The genes that were up- or down-regulated by more than two-fold and with *P*-values <0.05 were considered to be differentially expressed.

AUTHOR CONTRIBUTIONS

YO and QZ conceived and designed the experiments; GL, XL, YW, JM, DZ, and QD performed the experiments; YO, GL, XL, and FX analyzed the data; XL and JX contributed reagents/materials/analysis tools; YO, GL, and QZ wrote the paper.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Natural Science Foundation of China (31371599 and J1103510), the Outstanding Young Talents Program, and the Fundamental Research Funds for the Central Universities (2662016Q002, 2662015PY037, and 2014BC020).

CONFLICT OF INTEREST

All the authors declare that they have no conflict of interests.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phenotype and distribution of embryo-sac fertility, pollen fertility, and spikelet fertility in Nip × ZS97, Nip × MH63, and ZS97 × MH63 F_2 populations.

Figure S2. The phenotypes of the QTLs responsible for embryo-sac fertility, pollen fertility and spikelet fertility in three F_2 populations.

Figure S3. Fine mapping of *S5* locus in the Nip × ZS97 and Nip × MH63 F_2 populations.

Figure S4. Prediction of candidate genes for *Ef8* locus in the Nip × ZS97 F_2 population.

Figure S5. Prediction of candidate genes for *Pf12* locus in the Nip × ZS97 F_2 population.

Figure S6. Prediction of candidate genes for *Pf8* locus in the Nip × MH63 F_2 population.

Table S1. The spikelet fertility of Nip, ZS97, and MH63, and their F_1 hybrids.

Table S2. The information of bin maps from three F_2 populations.

Table S3. Prediction of the candidate genes in bins showing significant effect on hybrid fertility.

Table S4. Segregation distortion and the effects on fertility in three F_2 populations.

Table S5. The interacting loci for embryo-sac, pollen, and spikelet fertility in hybrids in three F_2 populations.

Table S6. Interacting bin clusters showing significant χ^2 independence test values in three F_2 populations.

Table S7. Numbers of single-locus QTLs and two-locus interactions.

REFERENCES

Alcazar, R., von Reth, M., Bautor, J., Chae, E., Weigel, D., Koornneef, M. and Parker, J.E. (2014) Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature. *PLoS Genet.* **10**, e1004848.

Bauer, H., Willert, J., Koschorz, B. and Herrmann, B.G. (2005) The t complex–encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genet.* **37**, 969–973.

Bauer, H., Veron, N., Willert, J. and Herrmann, B.G. (2007) The *t*-complex-encoded guanine nucleotide exchange factor *Fgd2* reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev.* **21**, 143–147.

Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M.J. and Loudet, O. (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science*, **323**, 623–626.

Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J.L. and Weigel, D. (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-Type incompatibility syndrome in plants. *PLoS Biol.* **5**, e236.

Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X. and Barbash, D.A. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*, **314**, 1292–1295.

Broman, K.W., Wu, H., Sen, S. and Churchill, G.A. (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889–890.

Chae, E., Bomblies, K., Kim, S.T. *et al.* (2014) Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell*, **159**, 1341–1351.

Chen, L. and Liu, Y.G. (2014) Male sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* **65**, 579–606.

Chen, J., Ding, J., Ouyang, Y. *et al.* (2008) A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proc. Natl Acad. Sci. USA*, **105**, 11436–11441.

Chou, J.Y., Hung, Y.S., Lin, K.H., Lee, H.Y. and Leu, J.Y. (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol.* **8**, e1000432.

Civan, P., Craig, H., Cox, C.J. and Brown, T.A. (2015) Three geographically separate domestications of Asian rice. *Nat. Plants*, **1**, 15164.

Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. and Ayroles, J.F. (2013) Genetic incompatibilities are widespread within species. *Nature*, **504**, 135–137.

Doyle, J.J. and Doyle, J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15.

Du, H., Ouyang, Y., Zhang, C. and Zhang, Q. (2011) Complex evolution of *S5*, a major reproductive barrier regulator, in the cultivated rice *Oryza sativa* and its wild relatives. *New Phytol.* **191**, 275–287.

Duan, M., Sun, Z., Shu, L., Tan, Y., Yu, D., Sun, X., Liu, R., Li, Y., Gong, S. and Yuan, D. (2013) Genetic analysis of an elite super-hybrid rice parent using high-density SNP markers. *Rice*, **6**, 21.

Garris, A.J., Tai, T.H., Coburn, J., Kresovich, S. and McCouch, S. (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics*, **169**, 1631–1638.

Herrmann, B.G., Koschorz, B., Wertz, K., McLaughlin, K.J. and Kispert, A. (1999) A protein kinase encoded by the *t complex responder* gene causes non-mendelian inheritance. *Nature*, **402**, 141–146.

Huang, X., Feng, Q., Qian, Q. *et al.* (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068–1076.

Huang, X., Wei, X., Sang, T. *et al.* (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genet.* **42**, 961–967.

Huang, X., Zhao, Y., Wei, X. *et al.* (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genet.* **44**, 32–39.

Ikehashi, H. and Araki, H. (1984) Varietal screening of compatibility types revealed in F_1 fertility of distant crosses in rice. *Jpn. J. Breed.* **34**, 304–313.

Kubo, T. and Yoshimura, A. (2005) Epistasis underlying female sterility detected in hybrid breakdown in a *Japonica-Indica* cross of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **110**, 346–355.

Kubo, T., Yoshimura, A. and Kurata, N. (2011) Hybrid male sterility in rice is due to epistatic interactions with a pollen killer locus. *Genetics*, **189**, 1083–1092.

Kubo, T., Takashi, T., Ashikari, M., Yoshimura, A. and Kurata, N. (2015) Two tightly linked genes at the *hsa1* locus cause both F_1 and F_2 hybrid sterility in rice. *Mol. Plant*, **9**, 221–232.

Kubo, T., Takashi, T., Ashikari, M., Yoshimura, A. and Kurata, N. (2016a) Two tightly linked genes at the *hsa1* locus cause both F_1 and F_2 hybrid sterility in rice. *Mol. Plant*, **9**, 221–232.

Kubo, T., Yoshimura, A. and Kurata, N. (2016b) Pollen killer gene *S35* function requires interaction with an activator that maps close to *S24*, another pollen killer gene in rice. *G3*, **6**, 1459–1468.

Lafon-Placette, C. and Kohler, C. (2015) Epigenetic mechanisms of postzygotic reproductive isolation in plants. *Curr. Opin. Plant Biol.* **23**, 39–44.

Lee, H.Y., Chou, J.Y., Cheong, L., Chang, N.H., Yang, S.Y. and Leu, J.Y. (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*, **135**, 1065–1073.

Li, W.T., Zeng, R.Z., Zhang, Z.M., Ding, X.H. and Zhang, G.Q. (2006) Fine mapping of locus *S-b* for F_1 pollen sterility in rice (*Oryza sativa* L.). *Chin. Sci. Bull.* **51**, 675–680.

Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K. and Wang, J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, **25**, 1966–1967.

Liu, K.D., Zhou, Z.Q., Xu, C.G., Zhang, Q. and Saghai Maroof, M.A. (1996) An analysis of hybrid sterility in rice using a diallel cross of 21 parents involving *indica*, *japonica* and wide compatibility varieties. *Euphytica*, **90**, 275–280.

Liu, K.D., Wang, J., Li, H.B., Xu, C.G., Liu, A.M., Li, X.H. and Zhang, Q. (1997) A genome-wide analysis of wide compatibility in rice and the precise location of the *S5* locus in the molecular map. *Theor. Appl. Genet.* **95**, 809–814.

Long, Y., Zhao, L., Niu, B. *et al.* (2008) Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl Acad. Sci. USA*, **105**, 18871–18876.

Maheshwari, S. and Barbash, D.A. (2011) The genetics of hybrid incompatibilities. *Annu. Rev. Genet.* **45**, 331–355.

Masly, J.P., Jones, C.D., Noor, M.A., Locke, J. and Orr, H.A. (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science*, **313**, 1448–1450.

Mi, J., Li, G., Huang, J., Yu, H., Zhou, F., Zhang, Q., Ouyang, Y. and Mou, T. (2016) Stacking *S5*-n and *f5*-n to overcome sterility in *indica-japonica* hybrid rice. *Theor. Appl. Genet.* **129**, 563–575.

Mizuta, Y., Harushima, Y. and Kurata, N. (2010) Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl Acad. Sci. USA*, **107**, 20417–20422.

Morinaga, T. and Kuriyama, H. (1958) Intermediate type of rice in the subcontinent of India and Java. *Jpn. J. Breed.* **7**, 253–259.

Ouyang, Y. (2016) Progress of *indica-japonica* hybrid sterility and wide-compatibility in rice. *Chin. Sci. Bull.* **61**, 3833–3841.

Ouyang, Y. and Zhang, Q. (2013) Understanding reproductive isolation based on the rice model. *Annu. Rev. Plant Biol.* **64**, 111–135.

Ouyang, Y., Chen, J., Ding, J. and Zhang, Q. (2009) Advances in the understanding of inter-subspecific hybrid sterility and wide-compatibility in rice. *Chin. Sci. Bull.* **54**, 2332–2341.

Ouyang, Y., Liu, Y.G. and Zhang, Q. (2010) Hybrid sterility in plant: stories from rice. *Curr. Opin. Plant Biol.* **13**, 186–192.

Ouyang, Y., Li, G., Mi, J. *et al.* (2016) Origination and establishment of a trigenic reproductive isolation system in rice. *Mol. Plant*, 10.1016/j.molp.2016.1007.1008.

Reflinur, Kim, B., Jang, S.M., Chu, S.H., Bordiya, Y., Akter, M.B., Lee, J., Chin, J.H. and Koh, H.J. (2014) Analysis of segregation distortion and its relationship to hybrid barriers in rice. *Rice*, **7**, 3.

Rice Genomes Project (2014) The 3,000 rice genomes project. *GigaScience*, **3**, 7.

Seidel, H.S., Rockman, M.V. and Kruglyak, L. (2008) Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science*, **319**, 589–594.

Shi, J.X., Cui, M.H., Yang, L., Kim, Y.J. and Zhang, D.B. (2015) Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci.* **20**, 741–753.

Song, X., Qiu, S.Q., Xu, C.G., Li, X.H. and Zhang, Q. (2005) Genetic dissection of embryo sac fertility, pollen fertility, and their contributions to spikelet fertility of intersubspecific hybrids in rice. *Theor. Appl. Genet.* **110**, 205–211.

Ting, Y. (1949a) Chronological studies of the cultivation and the distribution of rice varieties, Keng and Sen (in Chinese). *Agr. Bull. Col. Agr. Sun Yat-sen Univ.* **6**, 1–32.

Ting, Y. (1949b) A preliminary report on the cultivation and distribution of *hsien* and *keng* rices in ancient China and the classification of current cultivars (in Chinese). *Mem. Coll. Agr. Sun Yatsen Univ.* **6**, 1–32.

Tucker, M.R. and Koltunow, A.M. (2014) Traffic monitors at the cell periphery: the role of cell walls during early female reproductive cell differentiation in plants. *Curr. Opin. Plant Biol.* **17**, 137–145.

Wan, J. and Ikehashi, H. (1995) Identification of a new locus *S-16* causing hybrid sterility in native rice varieties (*Oryza sativa* L.) from Tai-hu lake region and Yunnan province, China. *Breed. Sci.* **45**, 461–470.

Wan, J., Yanagihara, S., Kato, H. and Ikehashi, H. (1993) Multiple alleles at a new locus causing hybrid sterility between a Korean *indica* variety and a *javanica* variety in rice (*Oryza sativa* L.). *Jpn. J. Breed.* **43**, 507–516.

Wan, J., Yamaguchi, Y., Kato, H. and Ikehashi, H. (1996) Two new loci for hybrid sterility in cultivated rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **92**, 183–190.

Wang, J., Liu, K.D., Xu, C.G., Li, X.H. and Zhang, Q. (1998) The high level of wide-compatibility of variety 'Dular' has a complex genetic basis. *Theor. Appl. Genet.* **97**, 407–412.

Wang, C., Zhu, C., Zhai, H. and Wan, J. (2005) Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genet. Res.* **86**, 97–106.

Wang, G.W., He, Y.Q., Xu, C.G. and Zhang, Q. (2006) Fine mapping of *f5*-Du, a gene conferring wide-compatibility for pollen fertility in inter-subspecific hybrids of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **112**, 382–387.

Wang, L., Xie, W., Chen, Y. *et al.* (2010) A dynamic gene expression atlas covering the entire life cycle of rice. *Plant J.* **61**, 752–766.

White, M.A., Stubbings, M., Dumont, B.L. and Payseur, B.A. (2012) Genetics and evolution of hybrid male sterility in house mice. *Genetics*, **191**, 917–934.

Xie, W., Wang, G., Yuan, M. *et al.* (2015) Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc. Natl Acad. Sci. USA*, **112**, 5411–5419.

Xie, Y., Xu, P., Huang, J., Ma, S., Xie, X., Tao, D., Chen, L. and Liu, Y.G. (2017) Interspecific Hybrid Sterility in Rice Is Mediated by OgTPR1 at the S1 Locus Encoding a Peptidase-like Protein. *Mol. Plant*, **10**, 1137–1140.

Xu, X., Liu, X., Ge, S. *et al.* (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nature Biotechnol.* **30**, 105–111.

Yamagata, Y., Yamamoto, E., Aya, K. *et al.* (2010) Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proc. Natl Acad. Sci. USA*, **107**, 1494–1499.

Yanagihara, S., Kato, H. and Ikehashi, H. (1992) A new locus for multiple alleles causing hybrid sterility between an *Aus* variety and *javanica* varieties in rice (*Oryza sativa* L.). *Jpn. J. Breed.* **42**, 793–801.

Yang, C., Chen, Z., Zhuang, C., Mei, M. and Liu, Y.G. (2004) Genetic and physical finemapping of the *Sc* locus conferring *indica-japonica* hybrid sterility in rice (*Oryza sativa* L.). *Chin. Sci. Bull.* **49**, 1718–1721.

Yang, J., Zhao, X., Cheng, K. *et al.* (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science*, **337**, 1336–1340.

Yu, Y., Zhao, Z., Shi, Y. *et al.* (2016) Hybrid sterility in rice (*Oryza sativa* L.) involves the tetratricopeptide repeat domain containing protein. *Genetics*, **203**, 1439–1451.

Zhang, D. and Yang, L. (2014) Specification of tapetum and microsporocyte cells within the anther. *Curr. Opin. Plant Biol.* **17**, 49–55.

Zhang, D. and Yuan, Z. (2014) Molecular control of grass inflorescence development. *Annu. Rev. Plant Biol.* **65**, 553–578.

Zhang, H., Zhang, C.Q., Sun, Z.Z., Yu, W., Gu, M.H., Liu, Q.Q. and Li, Y.S. (2011) A major locus *qS12*, located in a duplicated segment of chromosome 12, causes spikelet sterility in an *indica-japonica* rice hybrid. *Theor. Appl. Genet.* **123**, 1247–1256.

Zhang, J.W., Chen, L.L., Xing, F. *et al.* (2016) Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA*, **113**, E5163–E5171.

Zhao, Z.G., Wang, C.M., Jiang, L., Zhu, S., Ikehashi, H. and Wan, J.M. (2006) Identification of a new hybrid sterility gene in rice (*Oryza sativa* L.). *Euphytica*, **151**, 331–337.

Zhao, H., Yao, W., Ouyang, Y., Yang, W., Wang, G., Lian, X., Xing, Y., Chen, L. and Xie, W. (2015) RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res.* **43**, D1018–1022.

Zhu, S.S., Jiang, L., Wang, C.M., Zhai, H.Q., Li, D.T. and Wan, J.M. (2005a) The origin of weedy rice Ludao in China deduced by a genome wide analysis of its hybrid sterility genes. *Breed. Sci.* **55**, 409–414.

Zhu, S.S., Wang, C.M., Zheng, T.Q., Ikehashi, H. and Wan, J. (2005b) A new gene located on chromosome 2 causing hybrid sterility in a remote cross of rice. *Plant Breed.* **124**, 440–445.