CellPress
PARTNER JOURNAL

Molecular Plant
Research Report

Artificial Selection in Domestication and Breeding Prevents Speciation in Rice

Jiaming Mi[1,2], Guangwei Li[1,2], Conghao Xu[1,2], Jiangyi Yang[1], Huihui Yu[1], Gongwei Wang[1], Xianghua Li[1], Jinghua Xiao[1], Huazhi Song[1], Qifa Zhang[1] and Yidan Ouyang[1,*]

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

[2]These authors contributed equally to this article.

***Correspondence: Yidan Ouyang (diana1983941@mail.hzau.edu.cn)**

https://doi.org/10.1016/j.molp.2020.01.005

ABSTRACT

Speciation has long been regarded as an irreversible process once the reproductive barriers had been established. However, unlike in natural populations, artificial selection might either accelerate or prevent speciation processes in domesticated species. Asian cultivated rice is a target crop for both domestication and artificial breeding; it contains two subspecies of *indica* and *japonica*, which usually produce sterile inter-subspecific hybrids due to reproductive barriers. In this study, we constructed the evolutionary trajectory of a reproductive isolation system *S5*, which regulates fertility in *indica–japonica* hybrids via three adjacent genes, based on the data of 606 accessions including two cultivated and 11 wild rice species. Although hybrid sterility haplotypes at *S5* lead to establishment of a killer-protector reproductive barrier, origin of wide-compatibility haplotypes by complex hybridization and recombination provides an opposing force to reproductive isolation and thus prevents speciation during domestication. Analysis in a diallel set of 209 crosses involving 21 parents showed that the wide-compatibility genotypes largely rescued fertility of *indica–japonica* hybrids, indicating that the wide-compatibility gene would enable gene flow to maintain species coherence. This counteracting system indicates that combined effects of natural evolution and artificial selection may result in reversible processes of speciation in rice, which may also have implications for genetic improvement of rice.

Key words: Rice, reproductive isolation, speciation, hybrid sterility, wide compatibility, artificial selection

Mi J., Li G., Xu C., Yang J., Yu H., Wang G., Li X., Xiao J., Song H., Zhang Q., and Ouyang Y. (2020). Artificial Selection in Domestication and Breeding Prevents Speciation in Rice. Mol. Plant. **13**, 650–657.

INTRODUCTION

Speciation has long been regarded as an irreversible evolutionary process in natural populations once the intrinsic reproductive barriers had been established, inevitably leading to the splitting of the populations. According to the classical evolutionary genetic model, intrinsic reproductive barriers are difficult to undo (Muller, 1939), and these barriers increase as quickly as the square of the time after the divergence of two taxa (Orr, 1995). Therefore, speciation has become irreversible because the probability of reversing reproductive isolation quickly approaches zero (Coyne and Orr, 2004). In domesticated species, artificial selection would probably accelerate reproductive isolation depending on the purposes of the breeders, which generate a number of variants with huge phenotypic and genetic differences (Schluter and Pennell, 2017). However, artificial selection may also prevent speciation due to intervention from humans, although little is known about such a reversible process of speciation during the evolution.

As a primary force and maintaining mechanism of speciation, reproductive isolation is widely observed and, from the perspective of both molecular and evolutionary biology, its understanding is updated (Fishman and Sweigart, 2018; Sweigart et al., 2019). Reproductive barriers can arise at either prezygotic or postzygotic developmental stages showing substantial effects on population structure and speciation. Hybrid sterility is a major form of postzygotic reproductive barrier that drives speciation. The Asian cultivated rice (*Oryza sativa*) contains two subspecies, *indica/xian* and *japonica/geng*. Hybrid sterility is commonly observed in inter-subspecific crosses, leading

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

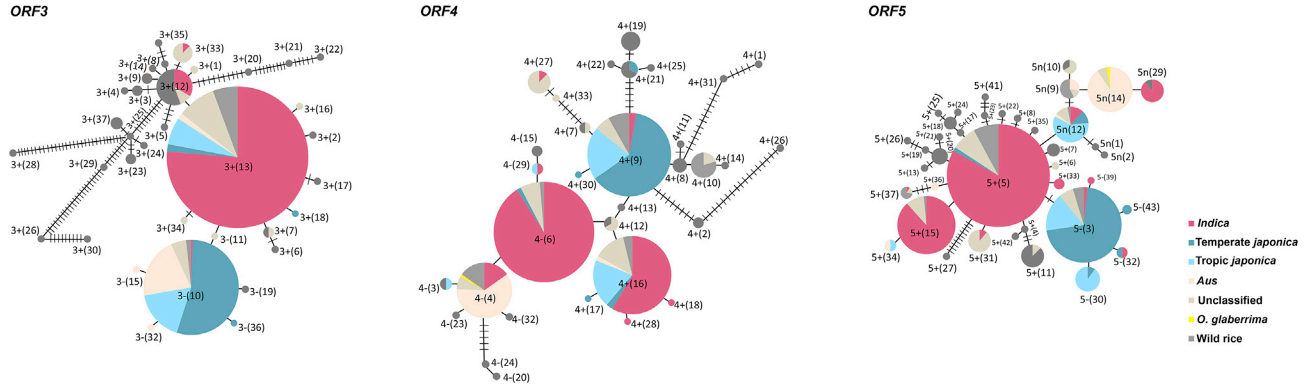


Figure 1. A Flowchart Showing the Evolutionary Process of the Allelic Groups of *ORF3*, *ORF4*, and *ORF5* of the *S5* Locus.
The steps of changes were deduced using the 172 SNPs and 30 insertions/deletions (InDels) identified in the coding sequences of *ORF3*, *ORF4*, and *ORF5*. Each circle represents a unique variant. *ORF3*, *ORF4*, and *ORF5* are abbreviated as "3," "4," and "5." For example, the description code "3+(13)" in the circle indicates the 13th variant of *ORF3*. The sizes of circles are proportional to the numbers of samples for given variants. Colors in the circles indicate taxonomic subgroups of the rice accessions. Each solid line represents one mutational step that interconnects two variants, and a thread perpendicular to the line indicates an additional mutation event. Seven variants, *ORF3*+(27), *ORF3*−(31), *ORF4*+(5), *ORF5*+(38), *ORF5*−(16), *ORF5*−(28), and *ORF5*/(40) are not included because of the complex process of variation. Unclassified indicates the accessions with unclear classification.

to reduced fitness along with segregation distortion in the offspring, and several cases regulating hybrid sterility were characterized within *O. sativa* or between *O. sativa* and other rice species (Chen et al., 2008; Long et al., 2008; Mizuta et al., 2010; Yamagata et al., 2010; Yang et al., 2012; Ouyang and Zhang, 2013, 2018; Kubo et al., 2016; Yu et al., 2016, 2018b; Li et al., 2017; Nguyen et al., 2017; Shen et al., 2017; Xie et al., 2017; Koide et al., 2018). In the cases of *O. sativa*, there exists a special group of rice germplasm, referred to as wide-compatibility varieties, that produce fertile offspring when crossed with both *indica* and *japonica* varieties (Ikehashi and Araki, 1984). Therefore, such wide-compatibility varieties enable hybridization between the two subspecies.

We previously found that a killer-protector system *S5* regulates *indica*–*japonica* hybrid sterility via three tightly linked genes, *ORF3*, *ORF4*, and *ORF5* (Chen et al., 2008; Yang et al., 2012; Zhu et al., 2017). Typical *indica* varieties possess the genotype of *ORF3*+*ORF4*−*ORF5*+, while *japonica* varieties contain *ORF3*−*ORF4*+*ORF5*−. The *ORF5*+ and *ORF4*+ constitute a killer, which selectively eliminates female gametes that do not carry the protector *ORF3*+. The non-functional alleles of *ORF3*− and *ORF4*− encode truncated proteins due to a 13-bp and an 11-bp deletion, whereas *ORF5*− had two nucleotide variations resulting in two amino acid substitutions. It was deduced that such a reproductive barrier originated through *Helitron* transposition and was enhanced by natural selection (Ouyang et al., 2016).

Here, we show that diverse forms of wide-compatibility haplotypes at *S5* locus originated and evolved owing to the complex hybridization and recombination during domestication; they enabled gene flow and provided an opposing driving force to reproductive isolation, which maintained subspecies identity and species coherence in the evolution. This counteracting system at *S5* resulted from a combination of evolutionary forces including natural evolution, domestication, and artificial selection, providing a unique example of a reversible process of speciation in crop plants.

RESULTS

Origin and Divergence of Natural Variations of *ORF3*, *ORF4*, and *ORF5* at *S5*

We expanded the investigation of the natural variations of *ORF*s *3–5* using the sequence data of 606 accessions with representatives of two cultivated and 11 wild rice species (Supplemental Table 1). The analysis identified 37 variants of *ORF3* (30 for *ORF3*+ and 7 for *ORF3*−), 33 variants of *ORF4* (24 for *ORF4*+ and 9 for *ORF4*−), and 43 variants of *ORF5* (28 for *ORF5*+, 7 for *ORF5*−, 7 for *ORF5n*, and one recombinant) (Supplemental Table 2). A haplotype flowchart was constructed to illustrate the process of how these natural variants were generated (Figure 1). The two distinct allelic groups *ORF3*+ and *ORF3*− corresponded to the *indica* and *japonica* groups, respectively. A high percentage of *ORF3*+(12) (55.6%) existed in the wild species (Supplemental Table 2); it had generated numerous derivatives in the wild rice with AA, BB, BBCC, and EE genome types. Such wide distribution of *ORF3*+(12) and its derivatives in the wild rice suggested that *ORF3*+(12) might be the most ancestral variant of *ORF3*+. The *ORF3*+(12) might also have generated a critical variant *ORF3*+(13) by one nucleotide mutation before the divergence of the cultivated and the wild species. Although *ORF3*+(13) was the most predominant variant identified in 96.1% of the sampled *indica* accessions, its frequency was only 5.7% in the wild rice. This suggested that *ORF3*+(13) originated from *ORF3*+(12) in a more recent time. Subsequently, the *ORF3*−(11) variant originated directly from *ORF3*+(13) by a 13-bp deletion, and the *ORF3*−(10) variant, which accounted for 80.5% of the accessions in *japonica*, arose from *ORF3*−(11) after a 1-bp substitution before the divergence of the cultivated and the wild species. Afterward, *ORF3*+(13) and *ORF3*−(10) became the major variants in *indica* and *japonica* rice groups, respectively.

The evolutionary process of *ORF4* could be traced based on the *ORF4*+ allelic groups in the *japonica* rice and the *ORF4*− allelic groups in the *indica* rice (Figure 1). *ORF4*+(9) existed in 16.7% of the wild species and 79.2% of the *japonica* rice; this allele was likely the most ancestral type that generated other clades

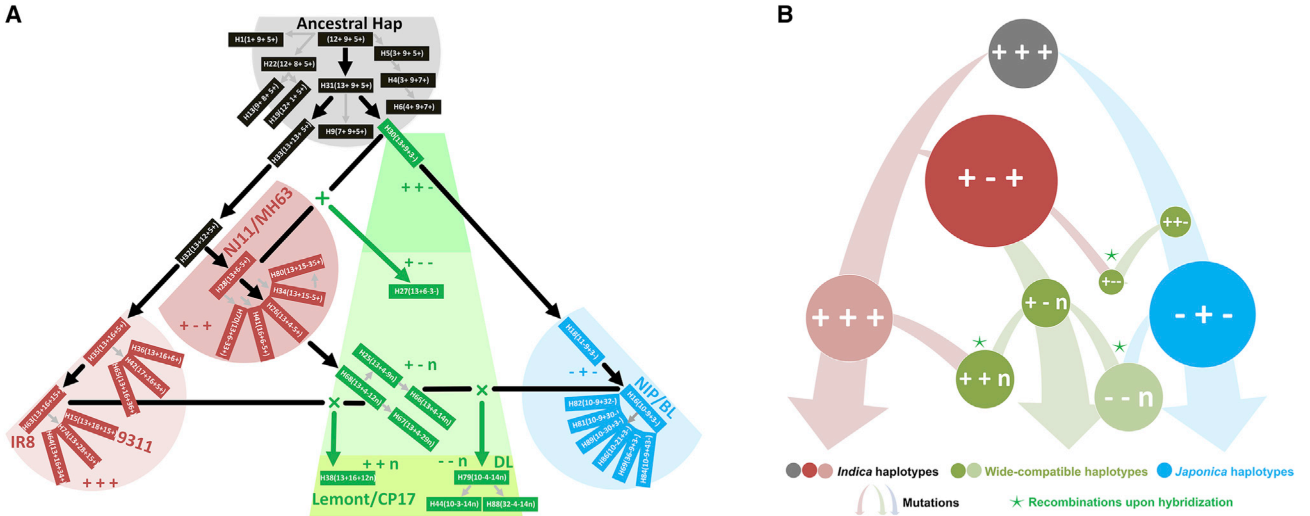


Figure 2. A Flowchart of the Evolutionary Steps at the *S5* Locus.
A flowchart of the evolutionary steps underlying the main haplotypes **(A)** and different genotypes **(B)** of the *S5* locus. The evolutionary relationships and steps are determined using 43 informative haplotypes. Each rectangle represents a unique haplotype of *S5*. For example, the description code "H31 (13+9+5+)" in the rectangle indicates the haplotype H31 from the combination of *ORF3*+(13), *ORF4*+(9), and *ORF5*+(5). Five distinct clusters are revealed in different colors. The first class of ancestral types with gray highlight is represented by 11 haplotypes that are found in most of the wild rice accessions. The gray, light-red, and red colors indicate haplotypes that are favored in *indica* cultivars. The blue color indicates the incompatible haplotypes of *ORF3*−*ORF4*+*ORF5*− in *japonica* rice, while the green color represents the wide-compatibility haplotypes. The solid arrows represent mutational and recombination events which interconnect two haplotypes. The names of the representative accessions for some haplotypes are labeled in the groups. BL, Balilla; NIP, Nipponbare; NJ11, Nanjing 11; MH63, Minghui 63; CP17, Cplso 17; DL, Dular.

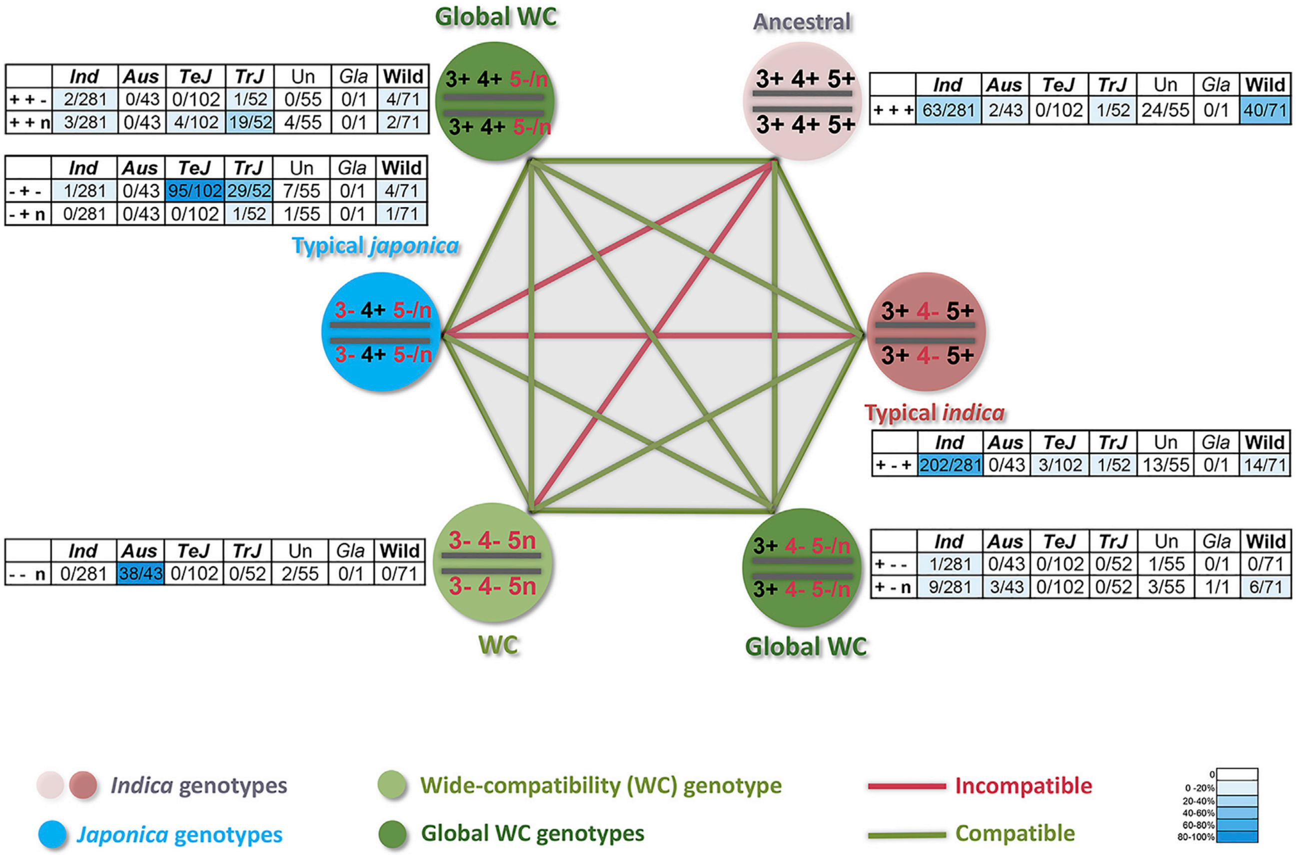
of variants during the evolution. Most of the derivatives of *ORF4*+(9) [*ORF4*+(2), *ORF4*+(7), *ORF4*+(8), *ORF4*+(13), *ORF4*+(21), and their related variants] occurred before the domestication of the cultivated rice. *ORF4*+(12) had two single-nucleotide polymorphisms (SNPs) relative to the ancestral *ORF4*+(9), and generated the *ORF4*+(16) clade and the *ORF4*− allelic groups. *ORF4*+(16) originated before the divergence of the cultivated and the wild species, and its frequency increased from 5.6% in the wild rice to 19.1% in the cultivated rice (Figure 1 and Supplemental Table 2). A 11-bp deletion, which was the most important event, occurred in *ORF4*+(12) before the divergence of the cultivated and the wild species, resulting in the *ORF4*− allelic group. The frequency of *ORF4*−(6) increased drastically in the *indica* subspecies, which made this variant the most frequent type in the *ORF4*− allelic group.

There were three allelic groups in *ORF5* that also revealed the evolutionary process. The first group was made up of the most frequent and most ancestral *ORF5*+(5), which existed in 26.4% of the accessions from the wild species and 71.2% of the *indica* rice, producing various forms of derivatives (Figure 1 and Supplemental Table 2). *ORF5*−(3) was generated from *ORF5*+(5) by two mutations, which were sequence features of the *ORF5*− group. Seven variants had the characteristic 115-amino-acid deletion in the N terminus of the predicted protein. These variants were assumed to be non-functional with respect to the growth and development of the plants per se and did not cause female sterility in the hybrids; they were regarded as belonging to the *ORF5n* allelic group (Chen et al., 2008). The *ORF5n*(12) in the *ORF5n* group was also generated from *ORF5*+(5) by a 136-bp deletion. Both of these two mutational events occurred before the divergence of the cultivated and the wild species.

Origin and Evolution Processes of Wide-Compatibility Haplotypes at *S5*

Although different haplotype groups of *ORF*s *3–5* were featured in *indica* and *japonica* groups, these natural variants were not fixed in the respective populations, suggesting that gene flow and exchange still occur between *indica* and *japonica* subspecies despite the reproductive barriers. A reasonable explanation might be the presence of the compatible haplotypes that prevent speciation, which can serve as a bridge for hybridization between the different rice groups. To learn how divergent natural variants of *ORF*s *3–5* evolved into compatible haplotypes at *S5*, we analyzed the mutational steps and evolutionary process deduced from 94 observed haplotypes (Figure 2A and Supplemental Figure 1). The *ORF3*+(12)*ORF4*+(9)*ORF5*+(5) containing the most ancestral variants of *ORF*s *3–5* was obviously the most ancestral haplotype at *S5*, although it was absent in the surveyed accessions. The haplotype H31 [*ORF3*+(13)*ORF4*+(9)*ORF5*+(5)], which originated directly from *ORF3*+(12)*ORF4*+(9)*ORF5*+(5) after one SNP substitution in *ORF3*+, behaved as a key component leading to the origin of *S5* system. Subsequently, independent mutational events occurred in H31 and generated H33 [*ORF3*+(13) *ORF4*+(13)*ORF5*+(5)] and H30 [*ORF3*+(13)*ORF4*+(9)*ORF5*−(3)], respectively, leading to different fates of the genetic combinations.

In one lineage, H32 [*ORF3*+(13)*ORF4*+(12)*ORF5*+(5)] originated from H33 by one SNP mutation in *ORF4*+, and played a critical role in the origin of the *indica* haplotypes (Figure 2A and 2B). The *ORF3*+*ORF4*+*ORF5*+ genotype was retained in one subgroup of *indica* rice. However, in the other subgroup, an 11-bp deletion mutation occurred in H32, which led to the functional differentiation from *ORF4*+ to *ORF4*−, giving rise to an *indica* genotype *ORF3*+*ORF4*−*ORF5*+. This in turn generated the predominant



	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
+ + -	2/281	0/43	0/102	1/52	0/55	0/1	4/71
+ + n	3/281	0/43	4/102	19/52	4/55	0/1	2/71

	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
+ + +	63/281	2/43	0/102	1/52	24/55	0/1	40/71

	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
- + -	1/281	0/43	95/102	29/52	7/55	0/1	4/71
- + n	0/281	0/43	0/102	1/52	1/55	0/1	1/71

	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
+ - +	202/281	0/43	3/102	1/52	13/55	0/1	14/71

	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
- - n	0/281	38/43	0/102	0/52	2/55	0/1	0/71

	Ind	*Aus*	*TeJ*	*TrJ*	Un	*Gla*	Wild
+ - -	1/281	0/43	0/102	0/52	1/55	0/1	0/71
+ - n	9/281	3/43	0/102	0/52	3/55	1/1	6/71

Figure 3. Summary of the Compatibility Spectra of Main Genotypes at *S5* Locus.

The red and green lines indicate incompatible and compatible crosses, respectively, based on our prediction. The numbers of the accessions in each combination are provided, and the intensity of the blue color corresponds to frequency of the genotypes. The genotype ++/ was not included in the tables. *Ind*, *indica*; *TeJ*, temperate *japonica*; *TrJ*, tropical *japonica*; *Gla*, *O. glaberrima*; Un, unclassified; Wild, wild rice.

haplotype of H28 [*ORF3*+(13)*ORF4*−(6)*ORF5*+(5)], which was observed in 70.8% of the *indica* rice surveyed.

In another lineage, an intermediate haplotype H30 [*ORF3*+(13) *ORF4*+(9)*ORF5*−(3)] originated from H31 after two single-nucleotide mutations in *ORF5*, which resulted in the functional differentiation from *ORF5*+ to *ORF5*− and, thus, breakdown of the killer (Figure 2A and 2B). This haplotype became the common ancestor of the *japonica* and the wide-compatibility haplotypes. A subsequent 13-bp deletion of *ORF3*+(13) to *ORF3*−(11) occurred in H30 after the emergence of non-functional killer, giving rise to the ancestral *japonica* haplotype H18. This in turn generated H16 [*ORF3*−(10)*ORF4*+(9)*ORF5*−(3)] spreading to a high frequency of 69.5% in the sampled *japonica* rice. Hence, the emergence and subsequent increased frequency of *ORF3*+*ORF4*−*ORF5*+ and *ORF3*−*ORF4*+*ORF5*− led to the establishment of the reproductive barrier, and in hybrids, the killer *ORF5*+ with an added protector *ORF3*+ provides double assurance for the *indica* rice to kill *japonica* gametes while at the same time protecting its own.

In contrast, the wide-compatibility haplotypes containing the non-functional killer or the functional protector would rescue the fertility and recover the allele and genotype ratio to Mendelian segregation in the offspring. A more complex evolutionary pattern was observed in the formation of these compatible haplotypes (Figure 2A and 2B). The intermediate haplotype H30 [*ORF3*+(13) *ORF4*+(9)*ORF5*−(3)] was compatible to both *indica* and *japonica* haplotypes, and produced a compatible haplotype of H27 [*ORF3*+(13)*ORF4*−(6)*ORF5*−(3)] by recombination, presumably after hybridization with a typical *indica* haplotype H28 [*ORF3*+(13)*ORF4*−(6)*ORF5*+(5)]. *ORF3*+*ORF4*−*ORF5n* was observed in both the wild rice and cultivated rice species (Figure 3); H68 [*ORF3*+(13)*ORF4*−(4)*ORF5n*(12)] originated directly from the *indica* haplotype H26 [*ORF3*+(13)*ORF4*−(4) *ORF5*+(5)] by a 136-bp deletion in *ORF5*. This wide-compatibility genotype of *ORF3*+*ORF4*−*ORF5n* facilitated the hybridization with both the *indica* genotype *ORF3*+*ORF4*+*ORF5*+ and the *japonica* genotype *ORF3*−*ORF4*+*ORF5*−. Consequently, a new wide-compatibility haplotype H38 [*ORF3*+(13)*ORF4*+(16) *ORF5n*(12)] was likely generated from a cross between H68 [*ORF3*+(13)*ORF4*−(4)*ORF5n*(12)] and H63 [*ORF3*+(13) *ORF4*+(16)*ORF5*+(15)]; this haplotype was observed in 36.5% of the tropical *japonica* rice surveyed. Likewise, another wide-compatibility haplotype H79 [*ORF3*−(10)*ORF4*−(4) *ORF5n*(14)] might originate as a result of recombination from hybridization between H66 [*ORF3*+(13)*ORF4*−(4)*ORF5n*(14)] and H16 [*ORF3*−(10)*ORF4*+(9)*ORF5*−(3)]. The genotype of *ORF3*−*ORF4*−*ORF5n* was almost fixed (95%) in the *aus* gene pool. Besides, the genotype of *ORF3*−*ORF4*+*ORF5n* might also arise from the recombination between *ORF3*+*ORF4*+*ORF5n*

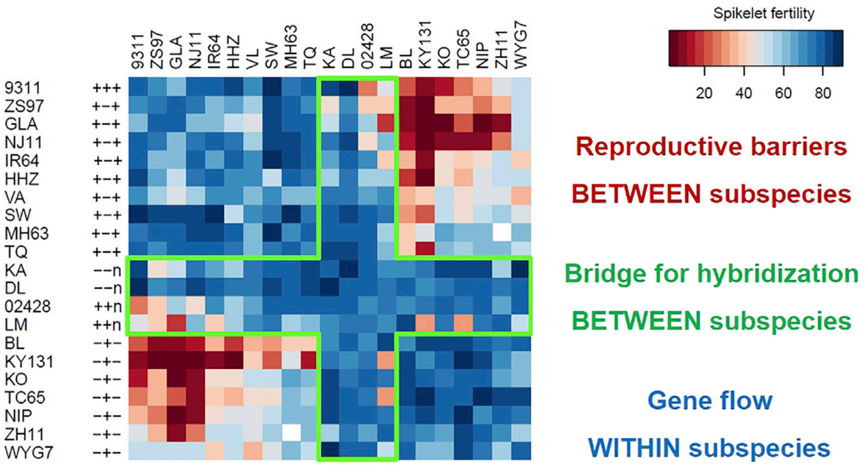


Figure 4. The Spikelet Fertility of F_1 Hybrids Based on a Diallel Set of 209 Crosses from 21 Parents.

Color bar at top right represents the values of spikelet fertility (the white color indicates missing values), and all materials used for hybridization are mentioned in Supplemental Table 6. ZS97, Zhenshan 97; MH63, Minghui 63; NJ11, Nanjing 11; TQ, Teqing; SW, Swarna; HHZ, Huanghuazhan; KA, Kasalath; GLA, Guangluai 4; DL, Dular; BL, Balilla; NIP, Nipponbare; ZH11, Zhonghua 11; WYG7, Wuyungeng 7; KY131, Kongyu 131; KO, Koshihikari; TC65, Taichung 65; VL, Varylaval; LM, Lemont.

and *ORF3−ORF4+ORF5−*, as supported by the observed low frequencies of the related haplotypes (Figure 3). Therefore, the creation of these wide-compatibility haplotypes occurred early by occasional natural hybridization before the divergence of cultivated and wild rice, but their increased frequencies might have been pushed subsequently by artificial selection in cultivated rice in the early domestication process. If the proportion of wide-compatibility haplotypes was pushed up to a certain level in rice varietal groups, these wide-compatibility genotypes would be able to maintain the population coherence at the species level.

Distributions and Proportions of Diverse Forms of Wide-Compatibility Haplotypes

We thus investigated the distributions and proportions of diverse forms of *S5* haplotypes. The ancestral genotypes *ORF3+ ORF4+ORF5+*, which was also a major genotype in *indica* rice, was identified in 34 different countries, with higher frequency in low-latitude regions of the Indian subcontinent, Southeast Asia, and Africa. The typical *indica* genotype of *ORF3+ORF4−ORF5+* was observed in 26 countries, with geographical preference in Southwest and South China. The *japonica* genotype of *ORF3−ORF4+ORF5−* was distributed in 37 countries, and was enriched in high-latitude regions of North Pacific, Northeast China, Russia, and Western Europe. Such differential geographical distributions had largely contributed in maintaining the reproductive barrier conferred by *S5*, thus enhancing the genetic differentiation of the rice populations.

Meanwhile, the 103 accessions carrying wide-compatibility haplotypes dispersed in 31 countries; they were observed in almost all regions showing coexistence of *indica* and *japonica* genotypes (Supplemental Figure 2 and Supplemental Table 1). This result supported the inference that the origin of wide-compatibility genotypes was due to the hybridization and recombination during domestication. The emergence of wide-compatibility genotypes further maintained the coherence of the rice species. *ORF3+ ORF4−ORF5n* (21.4% of all wide-compatibility accessions), *ORF3+ORF4+ORF5n* (32.0%), and *ORF3−ORF4−ORF5n* (38.8%) can be widely observed, while the other two wide-compatibility genotypes, *ORF3+ORF4+ORF5−* (0.07%) and *ORF3+ORF4−ORF5−* (0.02%), were in low frequencies (Figure 3). We speculate that four of the genotypes, *ORF3+ORF4+ORF5−*, *ORF3+ORF4+ORF5n*, *ORF3+ORF4−ORF5−*, and *ORF3+ORF4−ORF5n*, may be global wide-compatibility genotypes at *S5* locus, which can produce fertile and Mendelian segregation offspring when crossed with any other genotypes. The *ORF3−ORF4−ORF5n* genotype may produce fertile hybrids when crossed with both typical *indica* (*ORF3+ ORF4−ORF5+*) and *japonica* (*ORF3−ORF4+ORF5−*) genotypes, but not with *ORF3+ORF4+ORF5+* as it is likely that gamete of the genotype *ORF3−ORF4−ORF5n* in the hybrid would be killed by the killer of *ORF4+* and *ORF5+* without the protector *ORF3+*. Thus, *ORF3−ORF4−ORF5n* may also be defined as a wide-compatibility genotype with the only exception being the cross with the *ORF3+ORF4+ORF5+* genotype.

Phenotypic Effects of the Wide-Compatibility Genotypes in Fertility and Segregation Ratio in Offspring

To confirm the speculated effects of the described wide-compatibility genotypes, we investigated the fertility effect of *S5* locus using crosses from near-isogenic lines (Supplemental Table 3). Under the genetic background of Balilla, the spikelet fertility of three homozygotes, Balilla (*ORF3−ORF4+ORF5−*), Nanjing 11 (*ORF3+ORF4−ORF5+*), and Dular (*ORF3−ORF4−ORF5n*), were normal. In a near-isogenic line heterozygous for *S5* from *japonica* and *indica* genotypes (*ORF3−ORF4+ORF5−/ ORF3+ORF4−ORF5+*), the spikelet fertility was typically semi-sterile, showing an average 52.6% of seed-setting rate. However, under the same genetic background, the *S5* heterozygotes from either *japonica* and wide-compatibility genotype (*ORF3−ORF4+ORF5−/ORF3−ORF4−ORF5n*, 86.0%) or *indica* and wide-compatibility genotype (*ORF3+ORF4−ORF5+/ ORF3−ORF4−ORF5n*, 81.8%) showed a drastic increase of the fertility. Similarly, under the genetic background of 02428, the wide-compatibility genotype 02428 (*ORF3+ORF4+ORF5n*) could rescue the fertility to 81.3% and 77.0% when crossed with either *japonica* or *indica* genotypes. Therefore, these two types of wide-compatibility genotypes showed a high compatibility with both *indica* and *japonica* rice.

We further analyzed the fertility of hybrids at the population level in crosses involving a diallel set of 209 crosses from all possible pairs of 21 parental varieties (Figure 4 and Supplemental Table 4). The fertility of hybrids between varieties within subspecies was normal. The fertility of hybrids from 36 crosses between nine typical *indica* varieties with *ORF3+ORF4−ORF5+* showed an average fertility of 72.1%, while the spikelet fertility of the *indica* variety 9311 (*ORF3+ORF4+ORF5+*) was 77.7%. The hybrids from nine crosses between 9311 (*ORF3+ORF4+ORF5+*) and typical *indica* varieties with *ORF3+ORF4−ORF5+* varied from 60.0% to 85.9%, with an average value of 76.5%. Similarly, the average fertility of hybrids from 21 crosses between seven typical *japonica* varieties with *ORF3−ORF4+ORF5−* was 76.7%.

Reduced fertility was observed in crosses between subspecies (Figure 4 and Supplemental Table 4). The fertilities of hybrids from 62 crosses between nine typical *indica* (*ORF3+ORF4−ORF5+*) and seven typical *japonica* varieties (*ORF3−ORF4+ORF5−*) varied from 2.5% to 64.2%, with an average of 36.3%. Taking 9311 as an example, the hybrids of crosses between 9311 (*ORF3+ORF4+ORF5+*) and seven typical *japonica* varieties (*ORF3−ORF4+ORF5−*) also showed greatly reduced fertility (28.9%).

In contrast, the average spikelet fertilities of the hybrids from crosses between wide-compatibility varieties with either *indica* or *japonica* lines were significantly higher. The wide-compatibility varieties with *ORF3−ORF4−ORF5n* can increase the fertility from 36.3% to 73.1% when crossed with typical *indica* varieties with *ORF3+ORF4−ORF5+*. Besides, the average spikelet fertility of the hybrids from crosses between *ORF3−ORF4−ORF5n* and *ORF3+ORF4+ORF5+* was up to 84.9%; this fertility was higher than expected, possibly because of the existence of suppressors of hybrid sterility in *indica* genetic background. The wide-compatibility varieties with *ORF3−ORF4−ORF5n* were also compatible with *japonica* varieties with *ORF3−ORF4+ORF5−*, showing an average spikelet fertility of 76.6%. Similarly, the wide-compatibility varieties with *ORF3+ORF4+ORF5n* can also increase the fertility by 15.3% and 35.0% when crossed with *indica* and *japonica* varieties, respectively, compared with the fertility observed in *indica*–*japonica* crosses. These results suggested that the *S5* locus largely accounts for the *indica*–*japonica* hybrid sterility observed in the diallel crosses, indicating that the wide-compatibility gene enables hybridization and gene flow between the subspecies, which may reverse the process of speciation.

The transmission ratio of the gametes carrying different genotypes of *S5* locus was also investigated, using six inter-subspecific crosses (Supplemental Table 5). In the offspring from crosses between *indica* rice with either *ORF3+ORF4−ORF5+* or *ORF3+ORF4+ORF5+* and *japonica* rice with *ORF3−ORF4+ORF5−*, the *japonica* allele was in severe deficiency, also leading to deviations in genotype frequencies. Slight segregation distortion was observed in the cross of Dular (*ORF3−ORF4−ORF5n*) × Akitakomachi (*ORF3−ORF4+ORF5−*), whereas no distortion of transmission was observed in the cross between Teqing (*ORF3+ORF4−ORF5+*) and 02428 (*ORF3+ORF4+ORF5n*). Taken together, wide-compatibility haplotypes do provide a suppression force against reproductive isolation by recovering the fertility and segregation ratio in the offspring.

DISCUSSION

S5 is identified as a major and perhaps one of the most important loci contributing to inter-subspecific hybrid sterility and *indica*–*japonica* differentiation as demonstrated by a large number of studies using many crosses (Zheng et al., 1992; Yanagihara et al., 1995; Liu et al., 1997; Zhang et al., 1997; Wang et al., 1998; Song et al., 2005; Ikehashi and Araki, 2008; Matsubara et al., 2011; Li et al., 2017). In addition, heterozygosity plots of whole-genome SNPs in 1439 *indica* hybrids showed extremely low heterozygosity at *S5* locus, suggesting that this locus controlled reproductive isolation and thus prevented gene flow between subspecies (Huang et al., 2015). However, exchanges of genetic information were still widely observed at the gene and the genome levels during the process of rice domestication (Mather et al., 2010; Zhao et al., 2018). A possible explanation is the reversible process of speciation due to wide-compatibility haplotypes driven by artificial selection during rice breeding and domestication. The origin of wide-compatibility genotypes at *S5* locus began from the disruption of either *ORF5+* or *ORF4+*, arose due to recombination upon hybridization during evolution, and finally was preserved by the breeding and domestication process by human activities. Wide-compatibility genes also existed at other hybrid sterility loci such as *Sa* and *Sc* in rice (Long et al., 2008; Shen et al., 2017), suggesting that this might represent a general strategy to recover fertility and suppress transmission ratio distortion in the progenies, which serve as a bridge of gene flow between reproductively isolated *indica* and *japonica* populations and thus lead to the reversible process of speciation. Such a counteracting dynamic system might be a unique phenomenon in crops resulting from combined effects of natural evolutionary driving force and artificial selection during the course of domestication and breeding.

In addition, hybrid sterility has also been an obstacle in preventing the utilization of strong inter-subspecific heterosis for achieving higher rice yield; overcoming the fertility barriers in hybrids remains a major challenge in high-yield *indica–japonica* hybrid rice breeding. Therefore, the germplasm containing wide-compatibility genotypes are useful for breaking the reproductive barrier between *indica* and *japonica* subspecies in rice breeding, and thus provide a feasible strategy for utilization of heterosis between the two subspecies (Ikehashi, 2009; Mi et al., 2016; Yu et al., 2018a).

METHODS

Plant Materials

A rice germplasm panel of 606 accessions used for sequence analysis had been described in a previous study, including 533 accessions of Asian cultivated rice *O. sativa*, one accession of African cultivated rice *Oryza glaberrima*, 25 accessions of *Oryza rufipogon*, 26 accessions of *Oryza nivara*, seven accessions of *Oryza barthii*, and 14 additional accessions representing eight other wild rice species (Ouyang et al., 2016). A complete list of the accessions used in this study is provided in Supplemental Table 1.

Six *indica–japonica* populations were used for analysis in deviation of genotype frequency from the Mendelian segregation at the *S5* locus, amounting to a total of 828 lines or individuals. The genotype data from two RIL populations, 9311 × Nipponbare (Huang et al., 2009; Wang et al., 2011) and 02428 × Teqing (Mao et al., 2011), and two F_2 populations, Nipponbare × Zhenshan 97 and Nipponbare × Minghui 63 (Li et al., 2017), were obtained from previous studies. The genotype of *S5* locus in two F_2 populations, Akitakomachi × Dular and Dourado Precoce × Sadu-cho, was determined using SSR markers.

A diallel set of 210 crosses was obtained by pairwise crossing of 21 varieties representing a broad range of cultivated rice germplasm including landraces, modern elite cultivars, and parents of hybrid rice (Supplemental Table 6). All of these 21 varieties have been used extensively in rice breeding. Hybrid seeds were obtained for 209 crosses (except Zhonghua 11 × Minghui 63). Seeds of the 209 hybrids and their parents (230 lines in total) were planted in the field in normal rice growing season in 2013 following a randomized complete block design with two replications in the Experimental Station of Huazhong Agricultural University in Wuhan. Each cross contained 20 seedlings planted in two rows. True hybrid plants were detected by visual examination in the growing season. At maturity, true hybrid plants in the middle of each row were harvested. The spikelet fertility for each plant was scored as the ratio of the number of filled grains to the total spikelet. The primers for genotype investigation of *ORF3*, *ORF4*, and *ORF5* at *S5* locus are listed in Supplemental Table 7.

Analysis of Segregation Distortion at the *S5* Locus

The genotype frequency at *S5* locus was investigated using a linked bin/marker in six inter-subspecific populations. The deviation of each bin/marker from the Mendelian segregation (1:2:1 ratio for genotype frequency in F_2 populations or 1:1 in RIL populations) was investigated by chi-square test.

Sequence Analysis

Multiple sequence alignment was performed using Clustal X version 1.83 (Thompson et al., 1997) and was further adjusted manually with GeneDoc 2.7 (Nicholas and Nicholas, 1997). Aligned sequences were imported into DnaSP 5.0 for extracting the nucleotide polymorphism of the haplotypes (Librado and Rozas, 2009). The haplotype networks were generated using the computer program Hap Star 7.0, followed by manual editing (Teacher and Griffiths, 2011).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This research was supported by grants from the National Natural Science Foundation of China (31991223, 30800678, and 31821005), the Hubei Provincial Natural Science Foundation of China (2019CFA061), and the National Program for Support of Top-notch Young Professionals.

AUTHOR CONTRIBUTIONS

Y.O. and Q.Z. conceived and designed the experiments; Y.O., J.M., G.L., C.X., J.Y., and H.Y. performed the experiments; Y.O., J.M., G.L., and C.X. analyzed the data; G.W., X.L., J.X., and H.S. contributed reagents/materials/analysis tools; Y.O., Q.Z., and J.M. wrote the paper.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: January 13, 2019
Revised: December 13, 2019
Accepted: January 13, 2020
Published: January 18, 2020

REFERENCES

Chen, J., Ding, J., Ouyang, Y., Du, H., Yang, J., Cheng, K., Zhao, J., Qiu, S., Zhang, X., Yao, J., 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. U S A **105**:11436–11441.

Coyne, J.A., and Orr, H.A. (2004). Speciation (Sunderland, MA: Sinauer Associates).

Fishman, L., and Sweigart, A.L. (2018). When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. Annu. Rev. Plant Biol. **69**:707–731.

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., et al. (2009). High-throughput genotyping by whole-genome resequencing. Genome Res. **19**:1068–1076.

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

Ikehashi, H. (2009). Why are there indica type and japonica type in rice?—History of the studies and a view for origin of two types. Rice Sci. **16**:1–13.

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.

Ikehashi, H., and Araki, H. (1986). Genetics of F_1 sterility in remote crosses of rice. Rice Genet. 119–130.

Koide, Y., Ogino, A., Yoshikawa, T., Kitashima, Y., Saito, N., Kanaoka, Y., Onishi, K., Yoshitake, Y., Tsukiyama, T., Saito, H., et al. (2018). Lineage-specific gene acquisition or loss is involved in interspecific hybrid sterility in rice. Proc. Natl. Acad. Sci. U S A **115**:E1955–E1962.

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

Li, G., Li, X., Wang, Y., Mi, J., Xing, F., Zhang, D., Dong, Q., Li, X., Xiao, J., Zhang, Q., et al. (2017). Three representative inter and intra-subspecific crosses reveal the genetic architecture of reproductive isolation in rice. Plant J. **92**:349–362.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**:1451–1452.

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., Su, J., Wu, H., Chen, Y., Zhang, Q., Guo, J., Zhuang, C., Mei, M., et al. (2008). Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. Proc. Natl. Acad. Sci. U S A **105**:18871–18876.

Mao, D., Liu, T., Xu, C., Li, X., and Xing, Y. (2011). Epistasis and complementary gene action adequately account for the genetic bases of transgressive segregation of kilograin weight in rice. Euphytica **180**:261–271.

Mather, K.A., Molina, J., Flowers, J.M., Rubinstein, S., Rauh, B.L., Lawton-rauh, A., Caicedo, A.L., Mcnally, K.L., and Purugganan, M.D. (2010). Migration, isolation and hybridization in island crop populations: the case of Madagascar rice. Mol. Ecol. **19**:4892–4905.

Matsubara, K., Ebana, K., Mizubayashi, T., Itoh, S., Ando, T., Nonoue, Y., Ono, N., Shibaya, T., Ogiso, E., Hori, K., et al. (2011). Relationship between transmission ratio distortion and genetic divergence in intraspecific rice crosses. Mol. Genet. Genomics **286**:307–319.

Mi, J.M., Li, G.W., Huang, J.Y., Yu, H.H., Zhou, F.S., Zhang, Q.F., Ouyang, Y.D., and Mou, T.M. (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. U S A **107**:20417–20422.

Muller, H.J. (1939). Reversibility in evolution considered from the standpoint of genetics. Biol. Rev. Camb. Philos. Soc. **14**:261–280.

Nguyen, G.N., Yamagata, Y., Shigematsu, Y., Watanabe, M., Miyazaki, Y., Doi, K., Tashiro, K., Kuhara, S., Kanamori, H., Wu, J., et al. (2017). Duplication and loss of function of genes encoding RNA polymerase III subunit C4 causes hybrid incompatibility in rice. G3 (Bethesda) **7**:2565–2575.

Nicholas, K., and Nicholas, H. (1997). GeneDoc: a tool for editing and annotating multiple sequence alignments, Distributed by the author. http://wwwpscedu/biomed/genedoc.

Orr, H.A. (1995). The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139**:1805–1813.

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

Ouyang, Y., and Zhang, Q. (2018). The molecular and evolutionary basis of reproductive isolation in plants. J. Genet. Genomics **45**:613–620.

Ouyang, Y., Li, G., Mi, J., Xu, C., Du, H., Zhang, C., Xie, W., Li, X., Xiao, J., Song, H., et al. (2016). Origination and establishment of a trigenic reproductive isolation system in rice. Mol. Plant **9**:1542–1545.

Schluter, D., and Pennell, M.W. (2017). Speciation gradients and the distribution of biodiversity. Nature **546**:48–55.

Shen, R., Wang, L., Liu, X., Wu, J., Jin, W., Zhao, X., Xie, X., Zhu, Q., Tang, H., Li, Q., et al. (2017). Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. Nat. Commun. **8**:1310.

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.

Sweigart, A.L., Brandvain, Y., and Fishman, L. (2019). Making a murderer: the evolutionary framing of hybrid gamete-killers. Trends Genet. **35**:245–252.

Teacher, A.G.F., and Griffiths, D.J. (2011). HapStar: automated haplotype network layout and visualization. Mol. Ecol. Resour. **11**:151–153.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

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, L., Wang, A., Huang, X., Zhao, Q., Dong, G., Qian, Q., Sang, T., and Han, B. (2011). Mapping 49 quantitative trait loci at high resolution through sequencing-based genotyping of rice recombinant inbred lines. Theor. Appl. Genet. **122**:327–340.

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.

Yamagata, Y., Yamamoto, E., Aya, K., Win, K.T., Doi, K., Sobrizal, Ito, T., Kanamori, H., Wu, J., Matsumoto, T., et al. (2010). Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. Proc. Natl. Acad. Sci. U S A **107**:1494–1499.

Yanagihara, S., McCouch, S.R., Ishikawa, K., Ogi, Y., Maruyama, K., and Ikehashi, H. (1995). Molecular analysis of the inheritance of the S-5 locus, conferring wide compatibility in Indica/Japonica hybrids of rice (*O. sativa* L.). Theor. Appl. Genet. **90**:182–188.

Yang, J., Zhao, X., Cheng, K., Du, H., Ouyang, Y., Chen, J., Qiu, S., Huang, J., Jiang, Y., Jiang, L., 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., Tian, H., Liu, L., Bian, X., Xu, Y., Zheng, X., Gan, L., Shen, Y., et al. (2016). Hybrid sterility in rice (*Oryza sativa* L.) involves the tetratricopeptide repeat domain containing protein. Genetics **203**:1439–1451.

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

Yu, X., Zhao, Z., Zheng, X., Zhou, J., Kong, W., Wang, P., Bai, W., Zheng, H., Zhang, H., Li, J., et al. (2018b). A selfish genetic element confers non-Mendelian inheritance in rice. Science **360**:1130–1132.

Zhang, Q., Liu, K.D., Yang, G.P., Maroof, M.A.S., Xu, C.G., and Zhou, Z.Q. (1997). Molecular marker diversity and hybrid sterility in indica-japonica rice crosses. Theor. Appl. Genet. **95**:112–118.

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.

Zheng, K., Shen, P., Qian, H., and Wang, J. (1992). Tagging genes for wide compatibility in rice via linkage to RFLP markers. Chinese J. Rice Sci. **6**:145–150.

Zhu, Y., Yu, Y., Cheng, K., Ouyang, Y., Wang, J., Gong, L., Zhang, Q., Li, X., Xiao, J., and Zhang, Q. (2017). Processes underlying a reproductive barrier in *indica-japonica* rice hybrids revealed by transcriptome analysis. Plant Physiol. **174**:1683–1696.