University of Massachusetts Amherst
ScholarWorks@UMass Amherst

Masters Theses 1911 - February 2014

2013

The Evolution of Four Hybrid Sterility Loci in Us Weedy Red Rice

Stephanie M. Craig
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/theses

Craig, Stephanie M., "The Evolution of Four Hybrid Sterility Loci in Us Weedy Red Rice" (2013). *Masters Theses 1911 - February 2014*. 1116.
Retrieved from https://scholarworks.umass.edu/theses/1116

This thesis is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Masters Theses 1911 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

The Evolution of Four Hybrid Sterility Loci in US Weedy Red Rice

A Thesis Presented

By

Stephanie M. Craig

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

May 2013

Plant Biology

The Evolution of Four Hybrid Sterility Loci in US Weedy Red Rice

A Thesis Presented

By

Stephanie M. Craig

Approved as to style and content by:

_____

Ana L. Caicedo, Chair

_____

Elsbeth L. Walker, Member

_____

Li-Jun Ma, Member

_____

Elsbeth L. Walker, Program Director
Plant Biology

ACKNOWLEDGEMENTS

I extend my deepest gratitude to my advisor, Ana Caicedo whose support and creativity allowed me the opportunity to work in her lab and bring this project to fruition. She has been an amazingly dedicated mentor throughout my educational career in the lab. Special thanks to my committee members, Elsbeth Walker and Li-Jun Ma for all their insight and helpful suggestions throughout this project.

I would like to thank all the members of my lab for always providing useful feedback and a laugh when needed. Thanks to Carrie Thurber for lending me her expertise throughout the years. I would especially like to thank Lauren Resnick and Mike Reagon for their intellectual contributions to this project. I would also like to thank Sherin Perera for running countless PCR reactions.

I would also like to extend my gratitude to Peter Hepler and Caleb Rounds for all their help with the microscopy work. Thank you to the Bezanilla lab for allowing me access to their microscope

I would like to thank the National Science Foundation for providing the funding to make this research possible.

A special thank you to my family for their relentless love and moral support over the years.

ABSTRACT

THE EVOLUTION OF FOUR HYBRID STERILITY LOCI IN US WEEDY RED RICE

MAY 2013

STEPHANIE M. CRAIG, B.S., UNIVERSITY OF MASSACHUSETTS AMHERST

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Doctor Ana L. Caicedo

The accumulation of independent mutations in two populations of an ancestral species over time often leads to reproductive isolation. Reproductive isolation between diverging populations is re-enforced by genetic mating barriers that occur either pre- or post-zygotically. Hybrid sterility is the most common form of post-zygotic isolation in plants. Four post-zygotic sterility loci have been recently cloned in *Oryza sativa, Sa, s5, DPL1* and *DPL2*; these loci explain, in part, the limited hybridization that occurs between the domesticated cultivated rice varieties, *O. sativa* spp. *japonica* and *O. sativa* spp. *indica*. In the United States, cultivated fields of *japonica* rice are often invaded by conspecific weeds that have been shown to be of *indica* origin. Crop-weed hybrids have been identified in crop fields, but at low frequencies. This study aimed to examine the role of these known hybrid incompatibility loci in the interaction between cultivated and weedy rice, and further assess the evolutionary processes acting at these loci. We identified a novel neutral allele at *Sa*, designated *SaFX,* that seemingly acts to restore fertility. Additionally, we found a strikingly high frequency of neutral-type alleles in weedy populations at multiple loci, suggesting that weedy individuals, particularly those of the SH and BRH groups, should be able to freely hybridize with the local *japonica* crop.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Population divergence is a critical component of speciation that requires the participation of both pre- and post-zygotic mating barriers. Pre-zygotic barriers prevent formation of a zygote and include geographic, temporal, ecological and mechanical isolation, while post-zygotic barriers occur after the zygote has formed and include zygotic mortality, hybrid inviability and hybrid sterility. The complex interactions between these barriers, both intrinsically and extrinsically, complicate the ability to understand the genetic mechanisms of reproductive isolation (Widmer et al., 2009). Post-zygotic barriers provide an efficient venue to examine the molecular dynamics underlying the evolution of reproductive isolation in the lab because they occur, in large part, at the genetic level, as opposed to the environmental dependence of most pre-zygotic barriers. In plants, hybrid sterility is the most common form of post-zygotic isolation (Ouyang et al., 2010). Hybrid sterility is thought to evolve according to the Bateson-Dobzhansky-Muller (BDM) theory of speciation, which posits that deleterious interactions of alleles between genes of two species or populations is responsible for incompatibility (Orr, 1996) (Figure 1). Under this model, independent mutations occur in diverging populations, become fixed and then interact negatively in the background of the hybrid (Figure 1).

The sterility observed between crosses of two subspecies of Asian cultivated rice (*Oryza sativa* L.*)*, *indica* and *japonica*, is one of the most extensively studied hybrid incompatibilities (Oka, 1957; Harushima et al., 2002; Yang et al., 2004; Li et al., 2006, 2007, 2008; Chen et al., 2008, 2012; Kubo et al., 2008; Long et al., 2008; Wang et al., 2009, 2010; YiDan et al., 2009; Yamagata et al., 2010; Zhang et al., 2011; Zhao et al., 2011, 2012; Du et

al., 2011; You-Xin et al., 2012; Ji et al., 2012a). *Indica* and *japonica* cultivars were domesticated over 10,000 years ago and differ in their adaptability to different environments, various morphological characteristics and in their responses to a multitude of biotic and abiotic stresses (Reflinur et al., 2012). Gene exchange between cultivars would be highly beneficial to rice breeding practices, however, full exploitation of hybrid rice is limited by the tendency of these crosses to exhibit sterility (Reflinur et al., 2012). The identification of sterility-causing loci between these cultivars and subsequent determination of their contributions to sterility has been a topic of much research. Despite the numerous hybrid-incompatibility loci identified thus far in rice (YiDan et al., 2009), primarily through quantitative-trait analysis (QTL) studies, only a few have been cloned and subjected to experimental testing. Four such loci, encompassing three hybrid sterility systems, discovered in rice are *Sa*, *s5*, and *DOPPELGANGER1* (*DPL1*) and *DOPPELGANGER2* (*DPL2*). *Sa* and *DPL1/2* affect pollen viability and *s5* sterility results in embryo-sac abortion (Chen et al., 2008; Long et al., 2008; Mizuta et al., 2010).

The *Sa* locus comprises two adjacent genes on chromosome one, *SaM* and *SaF*. *SaM* encodes a small ubiquitin-like modifier E3 ligase-like protein and *SaF* encodes an F-box protein (Long et al., 2008). The genomic divergence between *indica* and *japonica* after domestication led to the deleterious interaction between *SaM* and *SaF*, causing male semi-sterility (usually about 50% from Long et al., 2008). *Indica* cultivars typically possess a *SaM*+/*SaF*+ haplotype, while *japonica* cultivars have a derived *SaM*-/*SaF*- haplotypes. A *SaM* heterozygote and a *SaF*+ allele are required to cause male semi-sterility and lack of any one of these alleles does not result in sterility (Long et al., 2008). SaM+ (257 aa) and SaM- (217 aa) are differentiated by a G-to-T polymorphism in the fifth intron of *SaM*+, resulting in

a truncated protein and removing of the sixth exon (Long et al., 2008). SaF+ and SaF- both contain 476 aa, but are differentiated by a C-to-T transition at position 287 that leads to aa change from Phe to Ser. Mechanistically, during the uninucleate stage of gamete development, SaM+ and SaF+ proteins travel to the SaM- bearing microspore and the direct interaction of SaF+ with SaM- causes selective abortion of SaM- bearing microspores (Long et al., 2008).

S5 (chromosome six) encodes an aspartic protease that is highly expressed in ovule tissue and consists of three alleles; an *indica*-type (*s5-i*), a *japonica*-type (*s5-j*) and a wide compatibility allele (*s5*-n) (Chen et al., 2008). SNPs C282A and C877T differentiate *s5-i* and *s5-j* alleles, while a 136 bp deletion in the N-terminus of *s5*-n renders it non-functional (Chen et al., 2008). The dimerization of *s5-i* and *s5-j* causes embryo sac abortion and usually reduces spikelet fertility by 46% (Chen et al., 2008). *s5* is of particular evolutionary interest because while this tri-allelic system seemingly acts to promote divergence between the two sub-species, s*5-n* functions, either directly or indirectly, to allow gene flow and hybridization by restoring complete fertility in combination with either allele.

*DPL*s encode small plant proteins and are highly expressed in mature anthers. *DPL1* (chromosome one) and *DPL2* (chromosome six) are paralogous hybrid incompatibility genes that resulted from independent mutations in *indica* and *japonica*, respectively (Mizuta et al., 2010). Nipponbare (*japonica*) contains a functional copy of *DPL1* (*DPL1*-N+) and a non-functional allele of *DPL2* (*DPL2-N-*). Kasalath (*aus*, a cultivated variety closely related to *indica*) carries a non-functional allele of *DPL1* (*DPL1*-K-) and a functional copy at *DPL2* (*DPL2*-K+). *DPL1* alleles are differentiated by a 517 bp insertion 204 bp downstream of the start codon. *DPL2* is 560 bp long encompassing a SNP at A434G (Mizuta et al., 2010).

Pollen carrying two defective alleles does not germinate (Mizuta et al., 2010). This dual-allele system is one of two "set of duplicate gametic lethal interaction model" genes known in rice (Mizuta et al., 2010).

All three of these known hybrid sterility loci act to promote divergence between the *indica* and *japonica* subspecies. The primary focus of cultivated rice reproductive isolation studies has been to understand the dynamics governing hybrid sterility in order to promote out-breeding between the *indica* and *japonica* lineages to improve crop-breeding strategies. However, a new concern has arisen that further highlights the urgency of understanding post-zygotic isolation: weeds.

Since crops were first domesticated, many weeds have evolved to take advantage of and invade the highly modified agricultural environment. In many cases, these weeds are closely related to the domesticated plants they invade. A weedy form of rice called weedy red rice invades cultivated rice fields worldwide and shares traits common in both cultivated and wild rice including dormancy, high shattering, high fecundity and rapid growth. Weedy rice infestations lead to reduction of rice yield and high financial loss (Fischer and Ramirez, 1993). The potential of valuable crop traits introgressing into native red rice populations and unfavorable weedy traits crossing into uniform cultivated fields poses a major threat to the global rice agroecosystem.

Weedy red rice is found in two main forms in the US: straw hull (SH), and black hull awned (BHA). This distinction is based on morphological and genetic differences (Reagon et al., 2010). Population structure is observed in the BHA group, which can be partitioned into BHA1 and BHA2. Other weedy groups include BRH and MX. BRH is a likely hybrid between the SH and BHA groups and MX individuals are hybrids between the local US

japonica crop and SH or BHA weeds. Weed evolution theories often include introgression via hybridization of wild and cultivated relatives or a de-domestication event that led to the spread of weedy species through the strong selection conditions of a monoculture, cultivated species (Lawton-Rauh and Burgos, 2010). Research in the Caicedo lab has shown that US weedy rice arose from cultivated rice groups from Asia and that colonization of the US led to a drastic reduction in population size (Reagon et al., 2010; Thurber et al., 2010). Studies of polymorphism have confirmed that the BHA and SH weedy groups are most closely related to and likely descendants from the *aus* and *indica* cultivated varieties, respectively (Londo and Schaal, 2007; Reagon et al., 2010).

Although weedy rice is classified as the same species as cultivated rice, and is partially inter-fertile with it, weedy rice in the US shows limited hybridization with the local crop, *japonica* (Shivrain et al., 2009); this may be due to self-pollination tendency of rice (Gealy and Gressel, 2005). However, since *indica* and *japonica* cultivars have been shown to experience limited hybrid compatibility, due to various deleterious genetic interactions, it is possible that similar hybrid barriers limit the amount of outcrossing between weedy rice and the local *japonica* cultivar. This study aims to investigate how the mating barriers between the *indica* and *japonica* subspecies have impacted the evolutionary dynamics of US weedy red rice. Analysis of the population structure of these genes (*Sa, s5*, and *DPL*s) will help predict the extent of weed-crop hybridization in the field and further enhance our knowledge of the processes responsible for the evolution of divergence-promoting substitutions within species.

CHAPTER 2

MATERIALS AND METHODS

2.1 Plant Material

Diverse *Oryza* seeds were obtained from the United States Department of Agriculture (USDA), the International Rice Research Institute (IRRI), collections contributed by Dr. David Gealy of the Dale Bumpers National Rice Research Center, and Susan McCouch of Cornell University (Table 1). Our working panel consists of 106 individuals from multiple *Oryza* species, including weedy rice (52), *O. rufipogon* (25), *O. nivara* (2) and various *O. sativa* cultivars including *aus* (6), *indica* (8), *japonica* (9): other AA genome *Oryza* species including *O. meridionalis* (1), *O. barthii* (2), *O. glaberrima* (1) were included as outgroups (Table 1). Plants were grown in a walk-in Conviron PGW36 growth chamber at the University of Massachusetts Amherst under 11 hour days at 25°C. Fifty *aus* and *indica* individuals obtained from USDA Genetic Stocks – *Oryza* Collection (GSOR) and IRRI were added to our survey for further sequencing needs for *s5* and *SaF* later in the project and is further explained within each section (Table 2). The five weedy groups used in our panel were previously defined by Reagon *et. al* 2010 based on 48 sequence tagged site markers (STS) and population structure analyses using InStruct (Gao et al., 2007).

2.2 Sequencing and Genotyping

DNA was extracted from leaf material of all accessions using a CTAB method. Primers were designed using Primer3 (Rozen and Skaletsky, 2000) to amplify portions of each hybrid sterility locus, taking into account previously described polymorphisms at each

locus (Chen et al., 2008; Long et al., 2008; Mizuta et al., 2010) (Table 3). Alleles were genotyped for all individuals at each locus using either DNA sequencing or differential gel migration. Sequencing was used when alleles were differentiated by a SNP, as was the case for *SaM+/SaM-*, *SaF+/SaF-*, *s5i/s5-j*, and *DPL2-K+/DPL2-N-*. Gel electrophoresis was used when alleles could be visualized by a size difference, as was done for *s5-n* and *DPL1-N+/DPL1-K-*. DNA sequences were aligned and edited using BioLign Version 2.09.1 (Tom Hall, NC State University). DNA sequences obtained were deposited into GenBank under accession numbers XXX. All PCR reactions were carried out on an Eppendorf epgradientS mastercycler with the following profile: 5 min at 95°, followed by 40 cycles of 30 sec at 94°, 30 sec at 52[1]°, 2 min at 68° with a final extension time of 5 min at 68°.

2.3 Genetic Diversity and Phylogenetic Analysis

Summary statistics for each sequenced locus were obtained with DnaSP version 5.0 (Librado and Rozas, 2009). Statistics included Watterson's estimator of nucleotide variation (θ_W), the average pairwise nucleotide diversity (π) (Nei and Li, 1979) and Tajima's D (TD) (Tajima, 1989). Summary statistics for each locus were compared against 48 genome-representative sequence tagged sites (STS) loci (Reagon et al., 2010) for outlier behavior. Heterozygotes were phased accordingly using the haplotype subtraction method (Clark, 1990). Genealogical relationships among haplotypes at each locus were determined with Neighbor-Joining analyses using a Kimura-2-parameter model in MEGA5 (Tamura et al., 2011). Weedy accessions were examined for possession of possible novel alleles, clade membership expectations based on known ancestry, and the likelihood of successful

[1] Tms varied for each primer pair, see Table 3.

introgression with US cultivars based on genotypes at each locus. Please note, at some loci, there are discrepancies between the number of genotyped alleles at a locus and those that were assigned a haplotype and used to create a phylogeny. Functional genotypes (i.e. classification as either *japonica*-type or *indica*-type) only required knowledge of a sequence at one (or two) SNPS, but haplotype designation needed polymorphism information across the entire sequenced region, which we could not provide for some individuals; alleles that were assigned a haplotype at each locus are asterisked in tables 4,5, and 6.

2.4 Crosses and Quantifying Pollen Viability

Crosses were performed between *Oryza* accessions carrying various *Sa* genotypes to observe the phenotypic consequences on pollen production (Table 7). Parents were planted in January 2012. Plants were grown in a walk-in Conviron PGW36 growth chamber at the University of Massachusetts Amherst under 11 hour days at 25°C. Panicles newly emerged from the boot, but not dehisced were chosen as the female parent to be pollinated. The top of each floret was cut off and the anthers removed with forceps, leaving the stigma intact; 20-30 florets were cut per panicle. A panicle on the cusp of dehiscence was chosen of the male parent. Both male and female panicles were then placed in a single glycine bag and secured with a paperclip. Bags were collected after one month to check for hybrid seed. F1 seeds were heat-treated overnight at 34°C and then either plated on a petri dish or planted; both methods led to successful germination. Once the F1 seeds grew enough to harvest tissue, DNA was extracted and PCR amplified to ensure heterozygosity.

Both homozygous parents of various *Sa* genotypes, as well as hybrid offspring were examined for pollen quality. For each sample, pollen grains of six anthers, from six different

florets were suspended in 100 µl of Lugol's Solution (LS), and then serially diluted in 90µl of LS. Pollen viability was quantified by obtaining a nonviable-to-viable ratio of pollen observed under a Leica MZ 16FA microscope using LS, a potassium iodide stain that reacts with starch in viable cells dying them black; non-viable cells do not stain and appear clear. Three ratios were calculated, using ImageJ (available at http://rsb.info.nih.gov/ij), within a fixed area from three fields of view and the average was used as the final percentage.

Quantity of pollen produced was measured using a 0.5 mm deep Nageotte Bright Line Hemacytometer, resulting in a pollen grain / µl concentration. We required the use of an extra-deep hemacytometer because the pollen grains were too large to employ a standard hemacytomer. Three measurements were obtained per individuals for both parameters. The average of those three raw measurements is reported in the Tables 9 and 10.

CHAPTER 3

RESULTS

3.1 The Genealogy of *s5*

A total of 1897 bp were sequenced for *s5* encompassing 949 bp downstream of the start codon in the second exon to 1016 bp after the end of *s5* (coordinates 5,759,638 – 5,761,535 on chromosome 6 according to MSU 6.0). 36 SNPs and 6 InDels were identified within this region. Of these, 7 were non-synonymous substitutions (NSS), 4 were silent-substitutions (SS) and 4 were InDels within coding regions, and 25 were SNPs and 3 were InDels in non-coding sequence. Twenty-nine haplotypes (alleles) were detected in our samples (Figure 2). Persistent amplification failures left us with a panel of 94 sequenced individuals at this locus including accessions from BHA1 (14), BHA2 (8), SH (14), BRH (4), MX (6), *japonica* (8), *indica* (6), *aus* (3) and wild (31) groups (Table 4). We classified alleles as either *indica*- or *japonica*-type based on the differentiating SNPs reported by Chen *et. al* (2008). *Indica*-type alleles were further grouped based on whether they contained the wide-compatibility deletion (Table 4). *Japonica*-type alleles with the wide-compatibility deletion have not been previously reported and were not detected (Table 4).

Previous analysis of the *s5* locus in wild groups found the majority of wild accessions to carry *indica*-like alleles (70%) (Du et al., 2011). Likewise, 97% of wild alleles in our study (30/31) are *indica*-type (Table 4). Previous surveys also found that *indica* and *japonica*-type alleles are fixed within their respective groups (Du et al., 2011). We found that all *indica* and *aus* individuals carry solely *indica*-type alleles, while only 62.5 % (5/8) of *japonica* individuals carry their respective *japonica*-type allele types (Table 4). The wide-

compatibility deletion was detected in 50% of cultivars, one of which is grown in the US (sus02), with *indica*-type alleles and in 13% of wild individuals with *indica*-type alleles (Figure 2, Table 4). A previous study found the compatibility deletion to occur in cultivated and wild populations at 27% and 18%, respectively (Du et al., 2011). No *indica* individuals possessed the compatibility deletion, but all *aus* (3) individuals did (Figure 2, Table 4).

The *s5* locus has not been previously examined in US weedy groups. As expected based on their putative ancestry, all BHA1, SH, and BRH individuals possess *indica*-type alleles (Table 4). Interestingly, all MX weeds also possess *indica*-type alleles, despite their mixed weed x *japonica* ancestry. The only weedy group not fixed for *indica*-type alleles was BHA2; only 63% of BHA2 (5/8) individuals carry *indica*-type alleles.

While the occurrence of *indica*-type alleles was expected in weedy groups due to shared ancestry, not all weedy haplotypes were identical to those detected in cultivars. Thirteen weedy individuals shared haplotypes 9 and 13 with their cultivated progenitors (Figure 2). However, nine novel haplotypes were also detected in the weedy populations, (Haplotypes 2-8, 14 and 15), seven of which possess the wide-compatibility deletion (Figure 2). These haplotypes are not shared with wild nor cultivated individuals in our panel, and have not been previously reported in the literature, except for haplotype 2 (Du et al., 2011). Haplotypes 4-9 and 12-15 are characterized by a novel SS (Cys → Cys) and a 10 bp deletion that seem to be in complete linkage disequilibrium. Haplotype 3 contains a novel NSS (Ser → Thr) that is also seen in haplotype 17.

Most weeds, regardless of population, contain the wide compatibility deletion, and the deletion is present in seven of the novel weed haplotypes. Surprisingly, we detected this deletion in only six of our cultivated individuals, all from the *aus* and *japonica* cultivar

groups (Table 4, *japonica* individuals not assigned a haplotype due to missing data). Because SH weeds are putative descendants from *indica* cultivars, we further explored the possible origins of the *s5* wide compatibility deletion in US weedy rice, by genotyping an additional 20 *aus* and 30 *indica* accessions from Bangladesh, India, Iran, Iraq, Myanmar, Nepal, Pakistan, and Sri Lanka (Table 2). The deletion was found to be common in *aus,* present in 19 individuals, mostly from Bangladesh (53%). The deletion was present, but rare in *indica,* detected in only 6 individuals, mostly from Nepal. Du et. al (2011) found 33% (4/12) of their cultivars with the wide-compatibility deletion to be from Bangladesh. The widespread presence of the wide-compatibility allele in US weedy rice groups suggests that this locus poses no barriers to hybridization with US crops, despite contrasting haplotypes among weedy and cultivated groups at this locus.

3.2 Genealogy and Allelic Distribution at the *DPL* loci

3.2.1 Allelic Distribution of *DPL1*

DPL1 was genotyped for functionality based on a 517 bp insertion as described by Mizuta *et. al* 2010. Individuals without the insertion were categorized as functional (*N*+) and non-functional if the insertion was present (*K*-) (Table 5). Mizuta *et. al* (2010) reported two variations of the functional allele that differ by a SNP (*N*+, *K*+), but we did not concern ourselves with determining which of the two functional alleles each individual had because that would have required sequencing. Thus, all of the functional alleles discovered at *DPL1* are labeled as *DPL1-N*+, but could be the alternate variant, *DPL1*-K+. Our sample set included 98 accessions from BHA1 (13), BHA2 (7), BRH (4), SH (11), MX (5), *aus* (5),

indica (6), *japonica* (14), and wild (33) groups (Table 5). We increased our *japonica* panel due to multiple amplification failures in *japonicas* specific to the US (Table 5).

A previous study found functional *DPL1* alleles (*DPL1-N+*) were more prevalent than non-functional ones (*DPL1-K-*) in cultivated and wild populations (75 and 71.5%, respectively) (Mizuta et al., 2010). Similarly, we found *DPL1-N+* alleles were present at higher frequencies in cultivated and wild populations (65% and 58%, respectively) (Table 5). Within cultivated populations, a previous study found that the *indica* group carried both alleles at equal frequencies (5/10), but the *japonica* group solely possessed functional *DPL1-N+* alleles (10/10) (Mizuta et al., 2010). While both alleles were present, we found the *DPL1-K-* allele at higher frequency in *indica* than the DPL1-*N+* allele, and fixed in all *aus* individuals surveyed (Table 5). Unexpectedly, we found 21% of our *japonica* individuals to carry the nonfunctional *DPL1-K-* allele, though surveyed US cultivars (sus02, sus04, sus06-09, sus11, and sus12) almost uniformly carried *DPL1-N+* alleles (Table 5).

As expected based on their *aus* ancestry, all members of BHA2 possess the non-functional *indica*-type allele (*DPL1-K-*). However, three individuals within BHA1, a group that also has *aus* ancestry, carry *DPL1-N+*, which could be suggestive of hybridization with the local *japonica* crop, as that allele is not found in any *aus* individual. MX and SH populations carry solely DPL1-*N+* alleles, even though we found this allele at lower frequency within *indica*, the domesticated progenitor of SH. In the case of MX weeds, this could be indicative of weed-crop hybridization in the US as 8/9 US *japonicas* (sus01, sus02, sus04, sus06-sus09, sus11, sus12) possess the *DPL1-N+* allele. However, given the existence of the *DPL1-N+* allele in *indica*, the contribution of ancestry to SH DPL1-*N+* alleles cannot be ruled out.

3.2.2 The Genealogy of *DPL2*

We sequenced 499 bp of *DPL2*, encompassing from 16 bp into the 1st exon through 19 bp into the second exon of the gene (coordinates 4,200,179-4,200,678 on chromosome six according to MSU 6.0). 95 samples were genotyped including accessions from BHA1 (13), BHA2 (9), SH (13), BRH (3), MX (5), *indica* (4), *japonica* (13), *aus* (4), and wild (31) groups. We increased our *japonica* panel to complement the added US cultivars at *DPL1* (Table 5). 4 InDels and 26 SNPs were found in our sample, and a total of 9 haplotypes (Figure 3). Of these, 2 NSS, 5 SS and 2 InDels were located within coding regions and 15 SNPs and 2 Indels were found in non-coding sequence. *DPL2-K+* has previously been found to be more prominent in wild populations (Mizuta et al., 2010). Accordingly, the *DPL2-K+* allele was fixed within our wild accessions (Table 5). Within cultivated populations, the non-functional (*DPL2-N-*) and functional (*DPL2-K+*) alleles were previously reported to be fixed within the *japonica* and *indica* populations, respectively. Comparably, the *DPL2-K+* allele was fixed in our *indica* population, but, strikingly, the *DPL2-N-* allele was rare in our *japonica* individuals, with only 15% (2/13) carrying *japonica*-like alleles (Table 5). Moreover, only one individual of US *japonica* individuals carried *DPL2-N-* (Table 5).

Similar to wild and domesticated groups, all weedy groups predominantly carried *indica*-type alleles by a wide margin (98%, 42/43). The only *japonica*-type allele was carried by one MX weed, rr18, thought to be a putative hybrid between an SH weed and a *japonica* cultivar (Reagon et al., 2010) (Table 5). The haplotype carried by this weed, haplotype 9, is the only *japonica*-type haplotype we found (Figure 3). Interestingly, along with the *japonica*–type defining SNP, this haplotype is also characterized by a novel NSS (Ser → Tyr) at pos 143, which could contribute to possible non-functionality. The remaining weedy

groups fell within only three haplotypes, 1, 4, and 8, with weedy individuals grouping with their domesticated progenitors in haplotypes 1 and 4 (Figure 3). A novel allele in the weedy populations was found in haplotype 8, encompassed by two BHA1 individuals, that is characterized by a 1 bp InDel in the coding region of *DPL2*, which could lead to a non-functional *DPL2* gene.

3.2.3 Genealogical Relationships across *DPL* loci

Previous study found four genotypes within wild populations, *DPL1-N+*/ *DPL2-N-* (17%), *DPL1-N+* / *DPL2-K+* (17%), *DPL1-K-* / *DPL2-K+* (33%), *DPL1-K+* / *DPL2-K+* (33%) (Mizuta et al., 2010). Only two genotypes were detected in our wild populations, *DPL1-K-* / *DPL2-K+* (42%) and *DPL1-N+*/ *DPL2-K+* (58%), but any of our *DPL1-N+* alleles could also be *DPL1-K+* (Table 5). Typical *japonica* cultivars are reported to carry *DPL1-N+/DPL2- N-* (Mizuta et al., 2010). Un-expectedly, the predominant genotype within our *japonica* panel was *DPL1-N+*/ *DPL2-K+*, a genotype previously reported to occur in 50% of *indica* samples (Mizuta et al., 2010), and only two of our *japonica* individuals retained the typical *DPL1-N+* / *DPL2-N-* genotype (Table 5). The other half of the *indica* population was previously found to carry *DPL1*-K- / *DPL2-K+* alleles, which we also found in 3 *japonica* individuals within our panel. Within our *indica* panel, we also found the two previously reported genotypes at equal frequencies, but only the *DPL1-K-* / *DPL2-K+* genotype was found among our *aus* cultivars (Table 5).

Among weedy populations, *DPL1-N+* / *DPL2-K+* was fixed in SH and nearly fixed in MX and BRH individuals, Outcrossing can be implicated from the BHA1 group that un-expectedly has three individuals with a *DPL1-N+ / DPL2-K+* genotype, given that *DPL1*-K-

/ *DPL2*-K+ is fixed within *aus*, its domesticated progenitor, and in BHA2. However, BHA1 could have obtained those alleles from other weedy groups, so hybridization with the local crop did not necessarily have to have occurred. The low frequency of *DPL2-N*- in US cultivars does not seem to present a barrier to crossing with weedy rice, because we would not expect to see the sterility-causing genotype, *DPL1-K*- / *DPL2-N*- very often.

3.3 Genealogical Relationships at the *Sa* locus

3.3.1 The Genealogy of *SaM*

We amplified a 634 bp portion of *SaM* starting 4 bp into the 4th exon through 44 bp downstream of 5th exon (22,379,680-22,380,314 on chromosome 1 according to MSU 6.0). Within this portion of the gene, we found 5 InDels and 20 SNPs, distributed among 20 haplotypes (Figure 4). 115 individuals were genotyped at this locus, including BHA1 (15), BHA2 (9), SH (16), BRH (5), MX (5), *aus* (5), *indica* (10), *japonica* (10), and wild (40) groups (Table 6). Alleles were further classified as either *indica* (*SaM*+) or *japonica* (*SaM*-)-like based on differentiating polymorphisms assigned by Long et al., (2008) (Table 6). Two previous studies found the *SaM*+ allele to be common in wild populations (Long et al., 2008; Wang et al., 2010). Our results were consistent with this, with 83% of wild individuals in our panel carrying *SaM*+ alleles (Table 6). The *SaM*+ allele has also been documented as the most common in *indica* cultivars, with previous panels carrying 92% (Wang et al., 2010) and 100% (Long et al., 2008) *SaM*+ alleles. Likewise, 90% (9/10) of our *indica* individuals carried the *SaM*+ allele (Figure 4, Table 6). The *SaM*- allele has been fixed in all *japonica* populations reported to date (Long et al., 2008; Wang et al., 2010), except in one individual

from China (Dianrui 406, Wang et al., 2010). Consistent with Wang *et. al*, we found one *japonica* individual (str02) with a *SaM+* allele in our sample, the rest possessed *SaM-* alleles. We expected *aus* to have predominantly *SaM+* alleles, due to its close relationship with *indica*, and, in successful amplifications, we found that 1 (of 5) *aus* individuals carried a *SaM-* allele (Table 6).

Consistent with our expectations based on US weed ancestry, most of our weeds (91% or 43/47), from all weedy populations possessed *SaM+* alleles. Even though only four individuals possessed the *SaM-* allele, it was seen in various groups (BHA2, SH, and MX) (Table 6). Additionally, two novel alleles were observed in the MX groups, MXSH (hap 19, *SaM+*) and MXBH (hap 20, *SaM-*) (Figure 4). *SaM-* alleles in weeds were largely identical (the same as haplotypes found in their progenitor groups) (Figure 4). Interestingly, hap 20 is characterized by a NSS (ala → pro) at pos 529, shared only with haps 2 and 5. The lack of allelic variation in the weedy groups suggests little introgression from crops to weeds at this locus.

3.3.2 The Genealogy of *SaF*

We amplified a 1.3 kb portion of *SaF* starting 674 bp into the first exon through 357 bp into the 3rd exon (22,375,204 -22,376,242 on chromosome 1 according to MSU 6.0). After amplification failures (mostly seen in SH: 15/16 and BRH: 5/5 groups), we genotyped a total of 92 individuals at this locus including accessions from BHA1 (15), BHA2 (8), MX (6), SH (1), *indica* (9), *japonica* (8), *aus* (7) and wild (38) groups (Table 6). Within this region, we found one InDel and 24 SNPs and 18 different haplotypes (Figure 5). Within coding regions we found 14 NSS and 6 SS, and in non-coding sequence, we identified 4

SNPs and the only InDel. Alleles were further classified as either *indica* (*SaF+*) or *japonica* (*SaF-*)-like based on differentiating polymorphisms assigned by Long et al., (2008) (Table 6). Previous reports found wild populations to carry both allele types at relatively equal frequencies, which was consistent with our findings (50% *SaF+*, 50% *SaF-*) among individuals that amplified. 22% (2/9) of our *indica* individuals carried a *SaF-* allele while previous reports found 10% (11/106) of *indica* individuals to carry *japonica*-type alleles (Long et al., 2008). The *SaF-* allele is consistently fixed within the *japonica* population in multiple studies (Long et al., 2008; Wang et al., 2010), except for 1 individual found in China that carries a *SaF+* allele (Wang et al., 2010). Likewise, *SaF-* is nearly fixed in our *japonica* samples. *SaF+* is also nearly fixed within our *aus* panel, as expected due to its close relationship with *indica* (Figure 5, Table 6).

Among samples that could amplify, we found that 90% of weedy individuals (27/30) carry *SaF+* alleles (Table 8). Consistent with their *aus* ancestor, *SaF+* is fixed in BHA1, and accounts for 87.5% (7/8) of alleles in BHA2 (Table 6). Similarly, the majority of MX weeds carry *SaF+* alleles (Table 6). The three weedy individuals detected with *japonica*-type alleles, rr56 (BHA2), rr22 (MX), and rr06 (SH), fall into haplotype 1, the most frequent *japonica*-type haplotype, but also the only *japonica*-type haplotype we detected in *indica* cultivars (Figure 5). There are no novel weedy alleles at this locus as there is strict grouping of the weedy populations with their domesticated ancestors in haplotypes 1 (*SaF-*), 3 (*SaF+*) and 4 (*SaF+*) (Figure 5). The lack of diversity in the weedy group is also due to failure of this gene to successfully amplify in a majority of SH and all BRH individuals.

3.3.3 *SaFX*

Early attempts to sequence and genotype the *SaF* gene were complicated by the amplification failures for the majority of samples belonging to the SH and BRH weedy groups and a few other *Oryza* samples (Table 8). We obtained whole-genome sequence of a single SH individual (Young and Caicedo, unpublished information), and noted a large deletion spanning the entire *SaF* gene in the weed sample. To determine the exact genomic boundaries of the *SaF* deletion, we designed primers on both sides of the inferred deletion breakpoints (Table 3) and sequenced four SH samples with amplification profiles consistent with the presence of a deletion. We found that the deletion spans 8,628 bp (from coordinates 22,371,187- 22,379,815 of chromosome 1 in the MSU 6.0 rice genome) and starts 2,902 bp in the gene upstream of *SaF*, *Os01g39660,* a putative transposon protein, through 1,240 bp into *SaM*, the gene located immediately downstream of *SaF*, knocking out the first four exons of this gene (Figure 6).

Since we found no evidence of this deletion having been reported before, we named the resulting *SaF* allele *SaFX*. To genotype for the *SaFX* allele, a complementary set of primers was used. The primers used to isolate and sequence the deletion (Table 3) amplify a 1500 bp band in individuals possessing the deletion, and fail to amplify in individuals where the deletion is not present. In contrast, primers designed to amplify only the *SaF* gene (Table 3) fail in individuals possessing the deletion, but amplify a product of 1200 bp in individuals lacking the deletion. We found that 61% of individuals that failed amplification for *SaF* (20/33) amplified with our *SaFX* primers (Table 8).

Given the prominence of this deletion in the SH groups, we expected to find this deletion in both cultivated and wild groups, particularly in the *indica* group that is believed to

be ancestral to the SH weed group. While the deletion was found in 6 wild individuals (Table 6), we found no individuals with the deletion among the *indicas* included within our initial survey. We thus expanded our sample set to genotype 50 more individuals belonging to the *aus* and *indica* cultivar groups (Table 2). The *SaFX* deletion was detected only in four *indica* individuals from Nepal (3) and India (1), suggesting a low frequency of this allele in wild and cultivated *Oryza* groups.

The boundary of the *SaFX* deletion extends four exons into *SaM* , but prior to the *SaM+/SaM-* differentiating SNP in the fifth intron (Figure 6). Taking this into account, we re-assigned *SaM* alleles in individuals containing *SaFX* as either *SaM+X* or *SaM-X*, to indicate which type of allele it carries (*indica* (+) or *japonica* (-)) and that a large portion of the gene is missing (Table 6).

3.3.4 Allelic Relationships across the *Sa* locus

Wild populations were previously reported to carry *SaM+* / *SaF+* (78%) and *SaM-* / *SaF-* (22%) genotypes (Wang et al., 2010). Of the 36 individuals we were able to genotype at both loci, 47% carried *SaM+* / *SaF+* and 14% carried *SaM- / SaF-* (Table 6). However, we also detected novel allelic combinations of *SaM-* / *SaF+* (5%), *SaM+* / *SaF-* (17%) and *SaM+X* / *SaFX* (17%) (Table 6). We believe *SaFX* arose in a *SaM+* background since it groups solely with *SaM+* alleles in wild individuals (Table 6). *Japonica* cultivars are known to commonly exhibit a *SaM-* / *SaF-* genotype. Consistently, this genotype was nearly fixed in our *japonica* individuals, including the US cultivars surveyed (Table 6). Likewise, the typical *indica* genotype, *SaM+* / *SaF+,* was also found to be nearly fixed in our *indica* and *aus* panels as well.

Consistent with *aus* ancestry, *SaM+* / *SaF+* is fixed and nearly fixed in BHA1 and BHA2, respectively. However, given that *SaFX* is fixed in BRH and nearly fixed in SH, hybridization between the weed and the crop seems very probable. We are not able to predict how *SaFX* affects hybrid sterility levels in F1 hybrids since this is its first report. However, we speculate that *SaFX* is a non-functional allele since it knocks out the entire *SaF* gene and 4 exons of *SaM* (Figure 6). In order to test this, we designed crosses that replaced *SaF+*'s role in the genotype known to cause male sterility between *indica* and *japonica* cultivars at *Sa*, a *SaM* heterozygote and a *SaF+* allele (see below).

3.4 The Consequences of *SaFX* on Pollen Viability

Because the *SaFX* gene has not been reported in the literature, we attempted to assess the phenotypic consequences of carrying this allele on pollen production. Knowing that a *SaM* heterozygote and a *SaF+* allele were required to cause male semi-sterility, we designed crosses to test the phenotypic consequences of replacing *SaF* with *SaFX* (Table 7). We decided to look at both the quantity and quality of pollen produced in individuals carrying and not carrying the *SaFX* alleles and in crosses performed between individuals that carry and do not carry this allele. Pollen quantity was assessed by using a hemacytometer to measure the amount of pollen grain produced by six anthers. Quality was measured by obtaining a non-viable to viable ratio from three fields of view under a microscope.

There were five main parental genotypes: *SaM+/SaF+, SaM+/SaF-*, *SaM-/SaF-*, *SaM+X/SaFX*, and *SaM-X/SaFX*. Parental genotypes were limited by the genotypes available within our panel. Ideally, our goal was to have each genotype behave as both the male and female parent, but *Oryza* crosses are difficult and varied heading dates prevented us

from making all our intended crosses (Table 7). Once we obtained hybrids and grew up the seeds, we validated heterozygosity, and there were some invalid progeny as well. The resulting, verified hybrid genotypes we were able to obtain were: *SaM+ / SaM-X // SaF- / SaFX* (2), *SaM+X / SaM- // SaF+ / SaFX* (1) and *SaM+* / *SaM-* // *SaF+* / *SaF-* (1). However, our *SaM+* / *SaM-* // *SaF+* / *SaF-* individual has not yet flowered and was not included in our pollen analyses.

3.4.1 Pollen Quality

Pollen quality was obtained by calculating non-viable ratios in ImageJ, (available at http://rsb.info.nih.gov/ij), of pollen grains from six anthers, to see if *SaFX* had any consequences on viability. Three raw non-viable ratios were obtained for each individual, and then averaged to generate one representative average / individual (Table 9). The individual with the highest non-viable ratio possessed a *SaM- / SaFX* genotype with an average of 44% non-viable pollen (Table 9). The genotype that produced the most viable pollen was *SaM+ / SaF-* with an average of 3.2% non-viability. We then performed a t-test comparing the parental individuals with a *SaFX* allele, against those without *SaFX*. There was no effect of *SaFX* on pollen quality ($p = 0.87$). Interestingly, the *SaM+ / SaFX* genotype produced an average of 8.45% non-viable pollen, less than other individuals that did not carry *SaFX* (*SaM- / SaF-* :12.23%) (Table 9). There was also no significant difference between averages of the two *SaFX* bearing genotypes (*SaM+* v. *SaM-* : $p = 0.0934$), although much higher levels of non-viable pollen were observed in individuals with a *SaM-* background. Note, however, that our measures are based on a few individuals for each *Sa* genotype, and we thus lack statistical power.

SaFX does not seem to affect pollen quality in the resulting F1 hybrids. The typical sterility-causing genotype at *Sa*, a *SaM* heterozygote and at least one *SaF+* allele, usually leads to a 50% decrease in total pollen viability (Long et al., 2008). This decrease in pollen viability was not seen in the F1 (Table 9) as the non-viable ratios among the progeny were 13.15 % (*SaM + / SaM*-X // *SaF-* / *SaFX*) and 5.3% (*SaM+X / SaM- // SaF+ / SaFX*) and do not differ significantly from parental averages (*SaM+* / *SaM-X* // *SaF-* / *SaFX* (2): p=0.6, p = 0.8 and *SaM+X* / *SaM-* // *SaF+* / *SaFX* : p= 0.8). Our hypothesis of *SaM* being non-functional out due to the *SaFX* deletion is strengthened by the observation of no decrease in pollen quality in our *SaM+X* / *SaM-* // *SaF+* / *SaFX* hybrid; a functional *SaM* heterozygote and *SaF+* allele should have resulted in a 50% decrease in total pollen viability.

3.4.2 Pollen Quantity

We also looked to see if the amount of pollen produced was affected by *SaFX*. Pollen count was measured using a 0.5 mm deep hemacytometer in each of the parental genotypes and in each of the progeny (Table 10). Three raw measurements were obtained for each individual, and then averaged to generate one representative average / individual (Table 10). There was much variation in the amount of pollen produced among parents, regardless of the *SaFX* allele (Table 10). The highest pollen-producing individual had a *SaM- / SaF-* genotype (97.5 pollen grains / µl) was and the lowest was *SaM-* / *SaFX* (14.1 pollen grains / µl) (Table 10). We then performed a t-test using parental averages, from Table 10, of *SaFX* carrying individuals against those that do not carry *SaFX* and found no statistical significance between groups (P=0.54).

There was no significant effect on pollen production in two of the hybrid progeny (SaM+*X* / *SaM-* // *SaF-* / *SaFX* , p = 0.3321 and *SaM+* / *SaM-X* // *SaF+* / *SaFX*, p= 0.59. *SaFX* does not seem to have any negative effects on the amount of pollen produced, however, given the few individuals we have to use and the extreme variation found among averages, this finding lacks statistical power.

3.5 Summary Statistics

Genetic diversity statistics were calculated for all members of the main weedy populations: BHA1, BHA2, and SH, as well as *indica*, *japonica, aus* and wild populations at each locus (Table 11). Amplification complications varied the number of individuals available for analysis in each group at each locus. We attempted to look for any common patterns in polymorphism trends across the three hybrid incompatibility loci.

A previous study looked at 48 genome-representative STS loci among *Oryza* and weedy groups and found wild populations to harbor the most genetic diversity, followed by intermediate levels in the cultivars with the lowest levels observed in US weedy groups, due to a genetic bottleneck upon US colonization (Reagon et al., 2010). In our study, the wild populations followed the expected trend, except at *s5* and *DPL2,* where BHA2 and *japonica* harbors the highest levels of nucleotide diversity (π), respectively (Table 11). Levels of diversity for weed groups varied among loci, but were sometimes an order of magnitude larger than the genome-wide averages. Particularly noticeable were the high levels of diversity observed in the BHA groups at the *s5* locus. At this locus, all three weed groups surpassed the levels of diversity seen in their putative cultivated ancestors (Table 11); moreover, all weed groups possessed novel haplotypes not seen in cultivated or weed groups.

TD values for most groups across all loci tended to be more negative than genomic averages, indicating a tendency towards an excess of rare mutations at many of these loci and many of these groups. In this respect, the *DPL2* locus stood out for the wild rice group, and is consistent with the many haplotypes observed in the wild group (Figure 3). Excess of rare alleles in populations can be due to population expansion, selection for diversity, or relaxed selection. Also, noticeably, BHA2 at *s5* is associated with a very positive TD (1.85), implicating an excess of common (intermediate frequency) alleles, and consistent with the almost equal frequencies of *indica* and *japonica*-type alleles (Figure 2, Table 5). Patterns of polymorphism are usually indicative of balancing selection or admixture.

CHAPTER 4

DISCUSSION

The genetic mechanisms underlying post-zygotic isolation have not yet been fully elucidated. In plants, hybrid sterility is the most common form of post-zygotic isolation and has been the subject of much research in order to improve hybrids within various crops including corn (Kohls et al., 2011), rice (Chen et al., 2008; Long et al., 2008; Mizuta et al., 2010), and wheat (Vinod et al., 2011). This study sought to investigate how three post-zygotic mating barriers between the *indica* and *japonica* subspecies of cultivated rice, *s5*, *DPL,* and *Sa* loci, have impacted the potential of crop-weed hybridization in cultivated US rice fields. The ability to predict levels of hybridization will be vital for effective weed management strategies in the rice agroecosysytem.

4.1 Evolution of Hybrid Sterility Loci in US Weedy Groups

Allelic distributions and summary statistics were obtained at each hybrid sterility locus in order to evaluate how these loci evolved in US weedy populations. Previous study of neutrally evolving loci among *Oryza* and weedy groups found weedy populations to harbor the lowest levels of diversity, due to a genetic bottleneck upon US colonization. Levels of diversity among our weedy groups varied across loci, but were sometimes an order of magnitude higher than the genome-wide averages (Table 11). Particularly striking were the high levels of variation observed at *s5* (Table 11). At this locus, all three weed groups surpassed the levels of diversity seen in their putative cultivated ancestors and possessed novel haplotypes not seen in cultivated or wild groups. This could imply that having the

wide-compatibility deletion, which makes the allele non-functional, removes selection at this locus (Table 5). The high levels of diversity observed in BHA2 are most likely due to hybridization given the presence of several *japonica*-type alleles (Tables 4,5, and 6). Furthermore, the relatively equal allele frequency of *indica* and *japonica*-type alleles in BHA2 at *s5* is supported by a very positive TD (1.85), which can be indicative of admixture. Novel alleles are not prevalent at *Sa* and *DPL*, as at least one weedy group is depauperate of diversity at each locus (Table 11). Accordingly, the negative TDs associated with the weedy groups at these loci, except for BHA2 at *DPL2*, could indicate selection against new mutations. There is lack of an overall evolutionary trend observed among hybrid sterility loci in the weedy groups, except that there are generally higher levels of diversity at the locus that confers female sterility, *s5*, as opposed to male sterility, *Sa* and *DPL*. It is noteworthy to point out that more non-coding sequence was collected for s5 than the other loci and non-coding sequence tends to be more variable than coding sequence.

Expectedly, the *indica* vs *japonica* type genotypes found within cultivated populations are also found within their respective descendent weedy group, except in *aus* and the BHAs at *Sa*, albeit at varying frequencies (Table 6). However, there are genotypes found within the main weedy populations (BHA1, BHA2, and SH) that are not found in their putative ancestors, suggesting hybridization. For example, several members of BHA1 at *DPL* carry *DPL1-N+* / *DPL2-K+* alleles, a genotype not found in any of their putative *aus* ancestors. This does not directly indicate hybridization with the local *japonica* crop, as the *DPL1-N+* allele could have also come from a cross with an SH, BRH or MX individual. BRH, a weedy hybrid of SH and BHA, carries alleles common in both parental groups at *s5, DPL,* and *Sa.* Interestingly, BRH seems to carry the genotype primarily found in the SH

population at *DPL2* and *Sa*. The MX population, a supposed *japonica* and SH or BHA hybrid, carry alleles found in all three parental groups at *s5, DPL2*, and *Sa*. However, at *DPL2,* MX most likely seems to be a cross between *japonica* and SH as the predominant genotype in MX, *DPL1-N+* / *DPL2- K+,* is fixed within the SH population. Conversely, at *Sa*, *SaFX* is nearly fixed in SH and found in no MX individuals. This indicates that both types of hybrids, *japonica* / SH and *japonica* / BHA, exist in US cultivated fields.

4.2 Origins and Implications of *SaFX*

This is the first report of the *SaFX* allele. Our analyses suggest that *SaFX* has no significant effects on pollen quality or quantity. This deletion could have arisen to counteract the sterility-causing interaction between a *SaM* heterozygote and *SaF+* allele. *SaFX* is comparable to the *s5-n* allele at the *s5* locus, enabling gene flow between the *japonica* and *indica* populations. The prominence of *SaFX* in weedy groups lead us to believe that the selection promoting the divergence between the *indica* and *japonica* populations was relaxed in US cultivated fields since *indica* cultivars are not grown here. However, the high occurrence of *SaFX* due to founder effects cannot be ruled out. The evolution of *SaFX* creates a new version of a tri-allelic system involved in the evolution of speciation genes. Another neutral-type allele has been reported at *Sa*, characterized by two polymorphisms, a 6 bp insertion in *SaM* and a SNP in *SaF* (Wang et al., 2010). The rise of neutral alleles in hybrid sterility systems is not uncommon, as neutral-type alleles were also reported at other hybrid sterility loci (LeiGang et al., 2009; Wang et al., 2009; Li et al., 2012).

All neutral genes reported act to restore fertility. Given that these hybrid sterility loci are in place to promote divergence between two populations, from an evolutionary

standpoint, it seems counter-productive for alleles that encourage gene flow to have evolved in the first place. However, as is seen in the killer-protective system at *s5* (Yang et al., 2012), an aspartic protease, some sterility genes do not function directly in pollen or seed production and deleterious interactions between alleles at these loci, while promoting sterility, also cause other problems within the organism (endoplasmic reticulum stress in the case of *s5*,Yang et al., 2012) unrelated to gamete development. We know *s5* is not required for gamete development, despite expression in the ovary, as the growth and fertility of *s5* RNAi knockdowns were not impaired (Ji et al., 2012b). Likewise, among our crosses where *SaF* (F-box protein) was deleted and *SaM* (E3-ligase-like protein) was most likely non-functional, plants also grew normally. In that respect, if the gene(s) contributing to hybrid sterility function directly in pollen or seed development, then a neutral allele would not be expected to evolve, as is the case for *DPL*, whose role is implicated in pollen development (Mizuta et al., 2010).

The complications caused within the organism cannot be lethal, as *indica* and *japonica* varieties have been propagated successfully with such incompatible genic combinations (Li et al., 1997; Reflinur et al., 2012), but should have some advantage over individuals without these deleterious interactions. This assumes two conditions: 1) that the pressure promoting divergence between populations is stronger than the selection over having an individual without deleterious interactions and 2) that there is a tangible fitness advantage in individuals that do not have these deleterious interactions that is able to be selected upon. The latter condition could be further complicated if the advantage was environmentally-dependent, which could, in part, explain the lack of an overall evolutionary

trend observed across hybrid sterility loci (Table 11). Intra-cellular complications caused by incompatible interactions at *Sa,* besides hybrid semi-sterility, have not been reported.

Further studies need to be done to evaluate how *SaFX* behaves between inter-subspecific crosses. We believe *SaFX* arose before domestication as it was found in three individuals in the wild population originating from Laos (or22), Cambodia (or25) and Papa New Guinea (or02) (Table 7). The four *indica* individuals with the deletion are from Nepal (3) and India (1) (Table 2). The low frequency of this allele in both wild and cultivated populations suggests that some sort of selection pressure has not enabled the rise in frequency of the *SaFX* allele.

4.3 Potential Impacts of Crop-Weed Hybridization

Our data implies that the evolution of weedy red rice in the US reduced the strength of the genetic mating barriers observed between *indica* and *japonica* cultivated rice species. Weeds should be able to freely cross with the crop if they carry compatible genotypes with *japonica* at each of the sterility loci. Given the prominence of the wide compatibility deletion at the *s5* locus in weedy groups, there do not seem to be many barriers acting to prevent hybridization with the local *japonica* crop. Interestingly, we found a US *japonica* (sus02) carrying an *indica*-type allele with the wide-compatibility deletion (Table 5). The sterility causing genotype at *DPL*, *DPL1-K-/DPL2-N-*, is not seen within our weedy panel and is not expected to be a problem due to the rarity of the *N-* allele among cultivars and weedy groups in the US (Table 6). Our data already suggests out-crossing at *DPL*, as the predominant genotype observed in US cultivars, *DPL1-N+/DPL2-K+*, is fixed within the SH population, and nearly fixed in the BRH and MX populations (Table 5). Crop-weed

hybridization is also possible at *Sa*, as *SaFX*, a seemingly neutral allele, is fixed in BRH and nearly fixed in SH. *SaFX* is absent in both BHA groups, indicating these groups are unlikely to out-cross. Interestingly, a study measuring gene flow rate from rice cultivars in Arkansas found SH and BRH to have the highest outcrossing rates of all weedy groups in experimental crosses (Shivrain et al., 2009). Accordingly, neutral alleles, *s5-n* and *SaFX* are fixed in BRH, along with all members of SH, except for 2 individuals, rr19 and rr17, at *s5* and except for one individual at *SaFX*, rr06 (Tables 4 and 6).

Lack of barriers to successful hybridization between crop and weeds only occur within the BRH and SH groups at all three loci (Tables 4, 5, and 6). Two (out of X) members of BRH, rr33 and rr36, carry *SaFX*, *s5-n* and functional alleles at *DPL1* and *DPL2*; this combination should enable them to cross freely (Tables 4, 5, and 6). Additionally, eight (out of X) individuals in SH (rr07, rr11, rr12, rr31, rr34, rr35, rr45, rr53) also carry those same alleles at all four loci. Individuals that possess *SaFX* almost uniformly carry *DPL1-N+* and *DPL2-K+* and *s5-n*, except for rr48 (Tables 4, 5, and 6).

Another study that looked at the *Sa* locus in ten weedy rice individuals in Yunnan, China found only two genotypes, *SaM+ / SaF+* and *SaM- / SaF-*, at equal frequencies (Wang et al., 2010),. Since the *SaFX* allele was identified in wild and cultivated populations, it should have been able to rise in frequency in both Asian and US weedy populations. However, the striking rise in frequency of *SaFX* in the US, and apparent lack of *SaFX* in Asian weedy populations, could be indicative of different selective pressures operating in Asian and US cultivated fields.

50% of cultivated rice grown in the US is grown in Arkansas, of which, 27% is herbicide-resistant (HR) (Shivrain et al., 2009). If the hybrid sterility loci used in this study

adequately represent how other hybrid sterility loci evolved within weedy groups, then the capacity of weedy groups to freely hybridize within these HR populations lends itself to the creation of HR red rice, undermining many weed prevention strategies. Moreover, experimental hybrids within these fields exhibit higher fitness than the cultivated parent, causing them to be selectively favored and will ensure their success in the field (Shivrain et al., 2009). In this respect, HR rice changes the selective environment for weeds, such that an HR weed will be more beneficial than a pure weed. Alternatively, gene flow of unfavorable weedy traits, including shattering and dormancy, could be passed into native cultivated fields that could interfere with uniform harvesting conditions. Likewise, the escape of an HR gene from red rice could also get loose in fields dedicated to un-genetically modified (GM) rice (Shivrain et al., 2009).

Genes involved in post-zygotic isolation are expected to be more numerous and have less effect than pre-zygotic genetic mating barriers (Widmer et al., 2009), with many polygenic effects on fertility. Accordingly, there have been ~57 quantitative trait loci (QTL) mapped that affect hybrid fertility in rice, of which 16 affect embryo-sac sterility, 30 are involved in male semi-sterility, and 11 affect spikelet sterility (YiDan et al., 2009). A recent study identified main-effect quantitative trait loci (QTL) for male and female sterility between *Oryza sativa* cv. Ilpumbyeo (Korean *japonica* cultivar) and Dasanbyeo (Korean *indica* variety) (Reflinur et al., 2012). They found 15 main-effect QTLs, one of which was *s5*, *qSF6.1* (*Sa* and *DPL* were not identified). In a hybrid reciprocal backcross, they allotted *sSF6.1* to account for 76% of phenotypic variance in one population (of 8) (Reflinur et al., 2012). Given the complexity of interactions between hybrid sterility loci, both intrinsically

and extrinsically, more studies need to adopt Reflinur's (2012) whole-genome approach to identifying QTLs.

This study identified the potential of crop-weed hybridization in US cultivated rice fields. The next three vital steps for effective weed management strategies are 1) to enhance our knowledge of the evolutionary processes acting within red rice populations in the US, 2) determining how much each of these loci is contributing to and responsible for semi-sterility between *japonica* and *indica* hybrids in consortium with 3) further mapping of candidate QTLs.

Table 1: Accessions Used

Seed Source ID ^	Lab ID	Sub-Pop	Accession	Origin
1096-01	rr05	BHA1	Redrice	AR
18A	rr16	BHA1	Redrice	AR
1996-9 (BLKH '89)	rr20	BHA1	Redrice	AR
LA3	rr21	BHA1	Redrice	LA
StgB	rr23	BHA1	Redrice	AR
StgS	rr24	BHA1	Redrice	AR
1005-02	rr28	BHA1	Redrice	AR
Prairie Co. Short - 8	rr38	BHA1	Redrice	AR
Prairie Co. Tall - 10	rr39	BHA1	Redrice	AR
Prairie Co. Tall - 11	rr40	BHA1	Redrice	AR
Prairie Co. Tall - 17	rr41	BHA1	Redrice	AR
1995-13	rr43	BHA1	Redrice	LA
1995-14	rr44	BHA1	Redrice	LA
1166-02	rr49	BHA1	Redrice	MS
2002-2-pot 1	rr57	BHA1	Redrice	AR
1025-01	rr02	BHA2	Redrice	AR
1081-01	rr03	BHA2	Redrice	AR
1188-01	rr13	BHA2	Redrice	AR
10A	rr14	BHA2	Redrice	AR
TX4	rr25	BHA2	Redrice	TX
1042-01	rr29	BHA2	Redrice	AR
1107-01	rr50	BHA2	Redrice	AR
1214-02	rr52	BHA2	Redrice	LA
1202-02	rr56	BHA2	Redrice	AR
1092-02	rr32	BRH	Redrice	MS
1111-01	rr33	BRH	Redrice	AR
1300-02	rr36	BRH	Redrice	MO
1183-01	rr48	BRH	Redrice	AR
1120-02	rr54	BRH	Redrice	AR
1996-1	rr18	MX	Redrice	AR
MS4 (R78-8 '93)	rr22	MX	Redrice	MS
2002-51	rr46	MX	Redrice	AR
2004-1-A	rr47	MX	Redrice	AR
2002-2-pot 21	rr58	MX	Redrice	AR
1196-01	rr06	SH	Redrice	AR
1098-01	rr07	SH	Redrice	MO
1134-01	rr08	SH	Redrice	AR
1160-01	rr11	SH	Redrice	AR

1179-01	rr12	SH	Redrice	LA
16B	rr15	SH	Redrice	AR
1995-15	rr17	SH	Redrice	AR
1996-5 (SHA- '93)	rr19	SH	Redrice	MS
1002-02	rr27	SH	Redrice	AR
1047-01	rr30	SH	Redrice	LA
1073-02	rr31	SH	Redrice	MO
1190-01	rr34	SH	Redrice	LA
1199-01	rr35	SH	Redrice	MO
1995-12	rr42	SH	Redrice	LA
1996-8	rr45	SH	Redrice	AR
1210-02	rr51	SH	Redrice	AR
1163-01	rr53	SH	Redrice	LA
RA5339	sau01	*aus*	Kasalath	India
RA4979	sau02	*aus*	Jhona_349	India
RA5323	sau03	*aus*	DV85	Bangladesh
RA5345	sau04	*aus*	Bj1	India
RA5361	sau05	*aus*	Dhala_shaitta	Bangladesh
22739	sau06	*aus*	Bei Khe	Cambodia
29016	sau07	*aus*	Aus 196	Bangladesh
RA4878	sin01	*indica*	Khao Dawk Mali - 105	Thailand
RA4911	sin02	*indica*	Rathuwee	Sri Lanka
RA4956	sin03	*indica*	LalAman	India
RA4984	sin04	*indica*	Dholi	Bangladesh
RA4967	sin05	*indica*	Ai-chiao-hong	China
RA4974	sin06	*indica*	CHAU	Vietnam
RA4978	sin07	*indica*	CHHOTE_DHAN	Nepal
RA4987	sin08	*indica*	POPOT-165	Indonesia
RA5344	sin09	*indica*	Dee_geo_woo_gen	USA
GSOR 301033	sin10	*indica*	Bei Khe	Cambodia
RA4945	ste03	*Temp_japonica*	Nep_hoa_vang	Vietnam
RA4882	str02	*Trop_japonica*	Kotobuki Mochi	Japan
RA4970	str04	*Trop_japonica*	Mirti	Bangladesh
RA4955	str06	*Trop_japonica*	Cicih_beton	Indonesia
	str08	*Trop_japonica*	Asse_y_pung	Philippines
CIOR 1206	sus01	*Trop_japonica*	DELITUS	USA
RA4998 PI 475833	sus02	*Trop_japonica*	Lemont	USA
CIOR 1645	sus05	*Trop_japonica*	CAROLINA_GOLD	USA
CIOR 9463	sus11	*Trop_japonica*	PALMYRA	USA
CIOR 1779	sus12	*Trop_japonica*	REXORO	USA

	obr01	Wild	*O. barthii*	Nigeria
	obr02	Wild	*O. barthii*	Mali
86779	ogb01	Wild	*O. glabberima*	Liberia
101148	omd01	Wild	*O. meridionalis*	Austrailia
103821	onv01	Wild	*O.nivara*	Thailand
86662	onv02	Wild	*O.nivara*	China
100588	or01	Wild	*O.rufipogon*	Taiwan
106523	or02	Wild	*O.rufipogon*	Papau New Guinea
106122	or03	Wild	*O.rufipogon*	India
100916	or05	Wild	*O.rufipogon*	China
104501	or06	Wild	*O.rufipogon*	India
104599	or07	Wild	*O.rufipogon*	Sri Lanka
104624	or08	Wild	*O.rufipogon*	China
104714	or09	Wild	*O.rufipogon*	Thailand
104871	or11	Wild	*O.rufipogon*	Thailand
105388	or12	Wild	*O.rufipogon*	Thailand
105491	or13	Wild	*O.rufipogon*	Malaysia
105568	or14	Wild	*O.rufipogon*	Philippines
105720	or16	Wild	*O.rufipogon*	Cambodia
105855	or17	Wild	*O.rufipogon*	Thailand
105888	or18	Wild	*O.rufipogon*	Bangladesh
106103	or20	Wild	*O.rufipogon*	India
106134	or21	Wild	*O.rufipogon*	India
106150	or22	Wild	*O.rufipogon*	Laos
106169	or24	Wild	*O.rufipogon*	Vietnam
106321	or25	Wild	*O.rufipogon*	Cambodia
106346	or26	Wild	*O.rufipogon*	Myanmar
106453	or27	Wild	*O.rufipogon*	Indonesia
106518	or28	Wild	*O.rufipogon*	Vietnam
81990	or29	Wild	*O.rufipogon*	Myanmar
106163	or30	Wild	*O.rufipogon*	Laos

^ Entries with only numbers are IRGC identification numbers of seed we received from IRRI

RA: seed received from Susan McCouch

CIOR or PI refers to seed we obtained from USDA-NPGS

Table 2: Additional Individuals genotyped for *s5* and *SaF*

ID type	ID	name	species	subtype	country
GSOR	301251	Shim Balte	*O. sativa*	*aus*	Iraq
GSOR	301006	ARC 7229	*O. sativa*	*aus*	India
GSOR	301046	DZ78	*O. sativa*	*aus*	Bangladesh
GSOR	301097	Mehr	*O. sativa*	*aus*	Iran
GSOR	301144	T26	*O. sativa*	*aus*	India
GSOR	301191	P 737	*O. sativa*	*aus*	Pakistan
GSOR	301201	Thavalu	*O. sativa*	*aus*	Sri Lanka
GSOR	301252	Halwa Gose Red	*O. sativa*	*aus*	Iraq
GSOR	301302	Aswina 330	*O. sativa*	*aus*	Bangladesh
GSOR	301306	DD 62	*O. sativa*	*aus*	Bangladesh
GSOR	301307	DJ 123	*O. sativa*	*aus*	Bangladesh
GSOR	301310	DM 43	*O. sativa*	*aus*	Bangladesh
GSOR	301311	DM 56	*O. sativa*	*aus*	Bangladesh
GSOR	301314	DV 123	*O. sativa*	*aus*	Bangladesh
GSOR	301318	Jamir	*O. sativa*	*aus*	Bangladesh
GSOR	301319	Kachilon	*O. sativa*	*aus*	Bangladesh
GSOR	301326	Paung Malaung	*O. sativa*	*aus*	Myanmar
GSOR	301336	Karkati 87	*O. sativa*	*aus*	Bangladesh
GSOR	301363	Kalubala Vee	*O. sativa*	*aus*	Sri Lanka
GSOR	312017	Dular	*O. sativa*	*aus*	India
GSOR	301069	Jaya	*O. sativa*	*indica*	India
GSOR	301070	JC149	*O. sativa*	*indica*	India
GSOR	301123	Rathuwee	*O. sativa*	*indica*	Sri Lanka
GSOR	301199	SLO 17	*O. sativa*	*indica*	India
GSOR	301274	IR-44595	*O. sativa*	*indica*	Nepal
GSOR	301288	LD 24	*O. sativa*	*indica*	Sri Lanka
GSOR	301303	BR24	*O. sativa*	*indica*	Bangladesh
GSOR	301305	Dawebyan	*O. sativa*	*indica*	Myanmar
GSOR	301315	EMATA A 16-34	*O. sativa*	*indica*	Myanmar
GSOR	301327	Sabharaj	*O. sativa*	*indica*	Bangladesh
GSOR	301329	Yodanya	*O. sativa*	*indica*	Myanmar
GSOR	301342	BALA	*O. sativa*	*indica*	India
GSOR	301344	JC 117	*O. sativa*	*indica*	India
GSOR	312012	Swarna	*O. sativa*	*indica*	India
GSOR	312016	Aswina	*O. sativa*	*indica*	Bangladesh
IRGC	83345	CHABO GHAIYA	*O. sativa*	*indica*	Nepal
IRGC	16163	AATTE	*O. sativa*	*indica*	Nepal

IRGC	61895	AKELEY MASINO	*O. sativa*	*indica*	Nepal
IRGC	61898	ANAGA	*O. sativa*	*indica*	Nepal
IRGC	61899	ANDI KALO	*O. sativa*	*indica*	Nepal
IRGC	61902	ASAME	*O. sativa*	*indica*	Nepal
IRGC	61913	BATESURA	*O. sativa*	*indica*	Nepal
IRGC	61923	BIRUMPHOOT	*O. sativa*	*indica*	Nepal
IRGC	83301	BAKAI	*O. sativa*	*indica*	Nepal
IRGC	83344	BUWA DHAN	*O. sativa*	*indica*	Nepal
IRGC	110306	ANADI DHAN	*O. sativa*	*indica*	Nepal
IRGC	83311	BASMATI DHAN	*O. sativa*	*indica*	Nepal
IRGC	88476	ANJANA	*O. sativa*	*indica*	Nepal
IRGC	109364	BHUWA DHAN	*O. sativa*	*indica*	Nepal
IRGC	110311	BOOMBOLI GHAIYA	*O. sativa*	*indica*	Nepal

individuals possessing the s5-wide compatibility deletion

individuals that amplified for the *SaF* deletion

Table 3: Primers Used for Amplification

Locus	Primer	Primer Sequence	Approximate Amplicon Size
SaM	SaF_007_Forward	5'- CTACAATTTAAGGCTGCACTGG-3'	730 bp
Tm:55°C	SaF_001_Reverse	5'-CCCTGATGAAGTCAATGGCA-3'	
SaF	SaM_004_Forward	5'- GAGCTACGAGAGATCCAGATA-3'	1270 bp
Tm:55°C	SaM_005_Reverse	5'-GCCAAAACACATGAAAATGG-3'	
	SaM_001_Forward	5'-CCGCTGTGTTCTCCTGCTCA-3'	1230 bp
Tm:55°C	SaFdel01_Forward	5'-TGCCCAATCAAGTGAGTCTG-3'	1500 bp
	SaFdel01_Reverse	5'-GACAGTGGGCAATTGGTGTA-3'	
s5	s5_indel_Forward	5'-TCAACCCATTTCCTTTCCTACG-3'	450-575 bp
Tm:58°C	s5_indel_Reverse	5'-CGCTCGATCGGATTAACAAGC-3'	
	s5_sr_Forward	5'- ATCCCGGCAGATCCTACAC-3'	1000 bp
	s5_sr_Reverse	5'- CGTACCCGCAAGCTATAAGA-3'	
	s5_snp_1_Forward	5'-GTGCATGACCTTTGCTCAGA-3'	1200 bp
Tm:54°C	s5_snp_1_Reverse	5'- GATGCAAACCTTGCTAACCA-3'	
DPL1	DPL1_ins_5_Forward	5'- GTTCAGACATTCGCGGATTAAGG-3'	1580 bp
Tm:58°C	DPL1_ins_4_Reverse	5'- GAAATGGCTCATTCCGTAACACC-3'	
DPL2	DPL2_snp_Forward	5'- TGCACATAGGCATTTTCAGG-3'	560 bp
Tm:55°C	DPL2_snp_Reverse	5'- AGCAGCTGAACACAAGAGCA-3'	

Table 4: *s5* Allele-Type Classifications

Sub-Pop	Ind.	*s5- i / s5-j*	Hap #
aus	sau01	**s5-i*	9
aus	sau02	**s5-i*	13
aus	sau03	**s5-i*	9
BHA1	rr05	**s5-i*	4
BHA1	rr16	**s5-i*	9
BHA1	rr21	**s5-i*	9
BHA1	rr23	**s5-i*	9
BHA1	rr24	**s5-i*	9
BHA1	rr28	**s5-i*	9
BHA1	rr38	**s5-i*	9
BHA1	rr39	**s5-i*	13
BHA1	rr40	**s5-i*	13
BHA1	rr41	**s5-i*	14
BHA1	rr43	**s5-i*	15
BHA1	rr44	**s5-i*	13
BHA1	rr49	**s5-i*	9
BHA1	rr57	**s5-i*	10
BHA2	rr02	**s5-j*	2
BHA2	rr03	**s5-j*	3
BHA2	rr13	*s5-i*	
BHA2	rr14	**s5-i*	7
BHA2	rr25	**s5-i*	9
BHA2	rr50	**s5-j*	3
BHA2	rr52	**s5-i*	9
BHA2	rr56	**s5-i*	16
BRH	rr33	**s5-i*	12
BRH	rr36	**s5-i*	12
BRH	rr48	**s5-i*	12
BRH	rr54	**s5-i*	12
indica	sin01	**s5-i*	26
indica	sin04	**s5-i*	26
indica	sin05	**s5-i*	16
indica	sin06	*s5-i*	
indica	sin07	**s5-i*	16
indica	sin08	**s5-i*	16
japonica	ste03	**s5-j*	27
japonica	str02.1	*s5-i*	

s5-i	***indica* -type allele**
s5-j	***japonica* -type allele**

japonica	str02.2	*s5-i*	
japonica	str08	**s5-j*	1
japonica	sus01	**s5-j*	1
japonica	sus02	*s5-i*	
japonica	sus05	**s5-j*	17
japonica	sus11	**s5-j*	18
MX	rr18	**s5-i*	10
MX	rr22	**s5-i*	9
MX	rr46	**s5-i*	10
MX	rr47.1	**s5-i*	10
MX	rr47.2	*s5-i*	
MX	rr58	**s5-i*	10
SH	rr07	*s5-i*	
SH	rr11	**s5-i*	5
SH	rr12	**s5-i*	6
SH	rr15	**s5-i*	8
SH	rr17	*s5-i*	
SH	rr19	**s5-i*	11
SH	rr30	**s5-i*	12
SH	rr31	**s5-i*	12
SH	rr34	**s5-i*	12
SH	rr35	**s5-i*	12
SH	rr42	**s5-i*	12
SH	rr45	**s5-i*	12
SH	rr51	**s5-i*	12
SH	rr53	**s5-i*	12
wild	obr01	**s5-i*	29
wild	or08.1	**s5-i*	19
wild	or08.2	**s5-j*	20
wild	or11.1	*s5-i*	
wild	or11.2	*s5-i*	
wild	or12.1	*s5-i*	
wild	or12.2	*s5-i*	
wild	or13.1	*s5-i*	
wild	or13.2	*s5-i*	
wild	or17.1	**s5-i*	21
wild	or17.2	**s5-i*	21
wild	or18.1	*s5-i*	

wild	or18.2	*s5-i*	
wild	or20.1	*s5-i*	
wild	or20.2	*s5-i*	
wild	or21.1	*s5-i*	
wild	or21.2	*s5-i*	
wild	or22.1	**s5-i*	22
wild	or22.2	**s5-i*	22
wild	or24.1	**s5-i*	23
wild	or24.2	**s5-i*	23
wild	or25.1	*s5-i*	
wild	or25.2	*s5-i*	
wild	or26.1	**s5-i*	11
wild	or26.2	**s5-i*	24
wild	or27.1	**s5-i*	16
wild	or27.2	**s5-i*	16
wild	or28.1	**s5-i*	25
wild	or28.2	**s5-i*	25
wild	or30.1	**s5-i*	28
wild	or30.2	**s5-i*	28
* individuals included in haplotype analysis			
individuals with wide-compatibility deletion			

Table 5: *DPL1* and *DPL2* Allele-Type Classifications

Sub-Pop	Ind.	*DPL1*	*DPL2*	Hap #
aus	sau01	K-	*K+	**1**
aus	sau02	K-	*K+	**1**
aus	sau03	K-	*K+	**1**
aus	sau05	K-		
aus	sau07	K-	*K+	**1**
BHA1	rr05		*K+	**1**
BHA1	rr16	N+	*K+	**8**
BHA1	rr20	N+	*K+	**1**
BHA1	rr21	N+	*K+	**1**
BHA1	rr23	K-	*K+	**8**
BHA1	rr24	K-		
BHA1	rr28	K-	*K+	**1**
BHA1	rr38	K-		
BHA1	rr39	K-	*K+	**1**
BHA1	rr40	K-	*K+	**1**
BHA1	rr41	K-	*K+	**1**
BHA1	rr43		*K+	**1**
BHA1	rr44	K-	*K+	**1**
BHA1	rr49	K-	*K+	**1**
BHA1	rr57	K-	*K+	**4**
BHA2	rr02	K-	*K+	**4**
BHA2	rr03	K-	*K+	**4**
BHA2	rr13		*K+	**1**
BHA2	rr14	K-	*K+	**1**
BHA2	rr25	K-	*K+	**1**
BHA2	rr29	K-	*K+	**1**
BHA2	rr50	K-	*K+	**4**
BHA2	rr52		*K+	**1**
BHA2	rr56	K-	*K+	**1**
BRH	rr32	N+	*K+	**1**
BRH	rr33	N+	*K+	**1**
BRH	rr36	N+	*K+	**1**
BRH	rr48	K-		
indica	sin01	K-	*K+	**1**
indica	sin02	N+	*K+	**4**
indica	sin06	K-		

DPL1	
K-	***indica*** **-type allele**
N+	***japonica*** **-type allele**
DPL2	
K+	***indica*** **-type allele**
N-	***japonica*** **-type allele**
	not enough sequence information

indica	sin07	K-		
indica	sin09	N+	*K+	**1**
indica	sin10	K-	*K+	**1**
japonica	ste03	N+	*N-	**9**
japonica	str02	K-	K+	
japonica	str04	K-	*K+	**4**
japonica	str06	N+	*K+	**4**
japonica	str08	N+	*K+	**4**
japonica	sus01	K-	K+	
japonica	sus02	N+	K+	
japonica	sus04	N+		
japonica	sus06	N+	*K+	**4**
japonica	sus07	N+	*K+	**4**
japonica	sus08	N+	*K+	**4**
japonica	sus09	N+	*K+	**4**
japonica	sus11	N+	*N-	**9**
japonica	sus12	N+	*K+	**4**
MX	rr18	N+	*N-	**9**
MX	rr22	N+	*K+	**1**
MX	rr46	N+	*K+	**4**
MX	rr47	N+	*K+	**1**
MX	rr58	N+	*K+	**4**
SH	rr06	N+	*K+	**1**
SH	rr07	N+	*K+	**1**
SH	rr08	N+	*K+	**1**
SH	rr11	N+	*K+	**1**
SH	rr12	N+	*K+	**1**
SH	rr19	N+	*K+	**1**
SH	rr27		*K+	**1**
SH	rr30		*K+	**1**
SH	rr31	N+	*K+	**1**
SH	rr34	N+	*K+	**1**
SH	rr35	N+	*K+	**1**
SH	rr45	N+	*K+	**1**
SH	rr53	N+	*K+	**1**
wild	ogb01		*K+	**2**
wild	omd01	N+	*K+	**3**
wild	onv01	N+	*K+	**1**
wild	onv02	N+	*K+	**4**

wild	or01.1	K-		
wild	or01.2	K-		
wild	or02.1	K-		
wild	or02.2	K-		
wild	or03.1	N+	*K+	**5**
wild	or03.2	N+	*K+	**5**
wild	or05.1	N+	*K+	**1**
wild	or05.2	N+	*K+	**1**
wild	or06.1	K-	*K+	**1**
wild	or06.2	K-	*K+	**1**
wild	or07.1	K-		
wild	or07.2	K-		
wild	or08.1	N+	*K+	**1**
wild	or08.2	N+	*K+	**1**
wild	or09.1	N+	*K+	**6**
wild	or09.2	N+	*K+	**6**
wild	or12.1	N+	*K+	**1**
wild	or12.2	N+	*K+	**1**
wild	or13.1	K-	*K+	**1**
wild	or13.2	K-	*K+	**1**
wild	or14.1	N+	*K+	**1**
wild	or14.2	N+	*K+	**1**
wild	or16.1	K-	*K+	**4**
wild	or16.2	K-	*K+	**7**
wild	or18.1	K-	*K+	**1**
wild	or18.2	K-	*K+	**1**
wild	or22.1	N+	*K+	**1**
wild	or22.2	N+	*K+	**1**
wild	or28.1		*K+	**1**
wild	or28.2		*K+	**1**
wild	or29.1	N+	*K+	**1**
wild	or29.2	N+	*K+	**1**
* individuals included in haplotype analysis				

Table 6: *Sa* Allele-Type Classifications

Sub-Pop	Ind.	*SaM*	Hap #	*SaF*	Hap #
aus	sau01	**SaM+*	2	**SaF+*	3
aus	sau02	**SaM+*	2	**SaF+*	3
aus	sau03	*SaM-*		**SaF+*	3
aus	sau04			**SaF-*	1
aus	sau05	**SaM+*	2	**SaF+*	8
aus	sau07	*SaM+*		**SaF+*	5
aus	sau10			**SaF+*	3
BHA1	rr05	**SaM+*	2	**SaF+*	3
BHA1	rr16	**SaM+*	2	**SaF+*	3
BHA1	rr20	**SaM+*	2	**SaF+*	3
BHA1	rr21	**SaM+*	2	**SaF+*	3
BHA1	rr23	**SaM+*	2	**SaF+*	3
BHA1	rr24	**SaM+*	2	**SaF+*	3
BHA1	rr28	**SaM+*	2	**SaF+*	3
BHA1	rr38	**SaM+*	2	**SaF+*	3
BHA1	rr39	**SaM+*	2	**SaF+*	3
BHA1	rr40	**SaM+*	2	**SaF+*	3
BHA1	rr41	**SaM+*	2	**SaF+*	3
BHA1	rr43	**SaM+*	2	**SaF+*	3
BHA1	rr44	**SaM+*	2	**SaF+*	3
BHA1	rr49	**SaM+*	2	**SaF+*	3
BHA1	rr57	**SaM+*	2	**SaF+*	3
BHA2	rr02	**SaM+*	2		
BHA2	rr03	**SaM+*	2	**SaF+*	3
BHA2	rr13	**SaM+*	2	**SaF+*	3
BHA2	rr14	**SaM+*	2	**SaF+*	3
BHA2	rr25	**SaM+*	2	**SaF+*	3
BHA2	rr29	**SaM+*	2	**SaF+*	3
BHA2	rr50	**SaM+*	2	**SaF+*	3
BHA2	rr52	**SaM+*	2	**SaF+*	3
BHA2	rr56	**SaM-*	3	**SaF-*	1
BRH	rr32	**SaM+*	10		
BRH	rr33	*SaM-X*		*SaFX*	
BRH	rr36	**SaM+X*	10	*SaFX*	
BRH	rr48	**SaM+X*	10	*SaFX*	
BRH	rr54	**SaM+X*	10	*SaFX*	

SaM	
SaM+	***indica*** **-type allele**
SaM-	***japonica*** **-type allele**
SaM+X	***indica*** **-type allele, missing 4 exons**
SaM-X	***japonica*** **-type allele, missing 4 exons**
SaF	
SaF+	***indica*** **-type allele**
SaF-	***japonica*** **-type allele**
SaFX	***SaF*** **deleted**
	not enough sequence information

indica	sin01	**SaM+*	2	**SaF+*	3
indica	sin02	**SaM+*	2	**SaF+*	3
indica	sin03	**SaM+*	10		
indica	sin05	*SaM+*		**SaF+*	4
indica	sin06.1	**SaM+*	2	**SaF+*	3
indica	sin06.2	*SaM+*		**SaF+*	5
indica	sin07	**SaM+*	2	**SaF+*	3
indica	sin08	*SaM+*		**SaF-*	1
indica	sin09	**SaM+*	2	**SaF+*	7
indica	sin10	**SaM-*	3	**SaF-*	1
japonica	ste03	**SaM-*	3	**SaF-*	1
japonica	str02	**SaM+*	2		
japonica	str04	**SaM-*	3	**SaF-*	1
japonica	str06	**SaM-*	3	**SaF-*	1
japonica	str08	**SaM-*	3	**SaF-*	1
japonica	sus01	**SaM-*	3	**SaF-*	1
japonica	sus02	**SaM-*	3	*SaF-*	
japonica	sus05	**SaM-*	3	**SaF-*	1
japonica	sus11	**SaM-*	5		
japonica	sus12	**SaM-*	3	**SaF-*	1
MX	rr18			**SaF+*	4
MX	rr22	**SaM-*	3	**SaF-*	1
MX	rr46	**SaM+*	19	**SaF+*	4
MX	rr47.1	**SaM+*	20	**SaF+*	3
MX	rr47.2	**SaM+*	2	*SaF+*	
MX	rr58	**SaM+*	2	**SaF+*	3
SH	rr06	**SaM-*	3	**SaF-*	1
SH	rr07	**SaM+X*	10	*SaFX*	
SH	rr08	**SaM+X*	10	*SaFX*	
SH	rr11	**SaM+X*	10	*SaFX*	
SH	rr12	**SaM+X*	10	*SaFX*	
SH	rr15	*SaM-X*		*SaFX*	
SH	rr19	**SaM+*	10		
SH	rr27	**SaM+*	10		
SH	rr30	**SaM+X*	10	*SaFX*	
SH	rr31	**SaM+X*	10	*SaFX*	
SH	rr34	**SaM+X*	10	*SaFX*	
SH	rr35	**SaM+X*	10	*SaFX*	
SH	rr42	*SaM+X*		*SaFX*	

SH	rr45	*SaM+X	10	SaFX	
SH	rr51	*SaM+X	2	SaFX	
SH	rr53	*SaM+X	10	SaFX	
wild	ogb01	*SaM+	4	*SaF+	6
wild	omd01	*SaM+	6	*SaF+	2
wild	onv01	*SaM-	5	*SaF-	1
wild	onv02	*SaM+	2	*SaF+	5
wild	or02.1	*SaM+X	2	SaFX	
wild	or02.2	*SaM+X	2	SaFX	
wild	or03.1	*SaM+	2		
wild	or03.2	*SaM+	7		
wild	or05.1			*SaF-	15
wild	or05.2			*SaF-	15
wild	or06.1	*SaM-	8	*SaF+	3
wild	or06.2	*SaM-	9	*SaF+	18
wild	or08.1	*SaM+	10	*SaF+	9
wild	or08.2	*SaM+	10	*SaF+	9
wild	or09.1	*SaM+	11	*SaF+	4
wild	or09.2	*SaM+	11	*SaF+	4
wild	or11.1	*SaM+	12	*SaF+	10
wild	or11.2	*SaM+	12	*SaF+	10
wild	or12.1	*SaM+X	13	SaFX	
wild	or12.2	*SaM+X	13	SaFX	
wild	or13.1	*SaM+	2	*SaF+	3
wild	or13.2	*SaM+	2	*SaF+	3
wild	or14.1			*SaF-	13
wild	or14.2			*SaF-	13
wild	or16.1			*SaF-	12
wild	or16.2			*SaF-	12
wild	or17.1	*SaM+	14	SaF-	
wild	or17.2	*SaM+	13	SaF-	
wild	or18.1	SaM+		*SaF-	9
wild	or18.2	SaM+		*SaF-	9
wild	or20.1	*SaM+	2	*SaF+	3
wild	or20.2	*SaM+	2	*SaF+	3
wild	or21.1	*SaM+	15	*SaF-	11
wild	or21.2	*SaM+	15	*SaF-	11
wild	or22.1			SaFX	
wild	or22.2			SaFX	

wild	or24.1	**SaM-*	16	**SaF-*	1
wild	or24.2	**SaM-*	16	**SaF-*	1
wild	or25.1	**SaM+X*	10	*SaFX*	
wild	or25.2	**SaM+X*	10	*SaFX*	
wild	or26.1	**SaM+*	17	**SaF+*	16
wild	or26.2	**SaM+*	17	**SaF+*	16
wild	or27.1	**SaM+*	18	**SaF+*	17
wild	or27.2	**SaM+*	18	**SaF+*	17
wild	or28.1	**SaM+*	2		
wild	or28.2	**SaM+*	2		
wild	or29.1			**SaF-*	14
wild	or29.2			**SaF-*	14
wild	or30.1	**SaM-*	3	**SaF-*	1
wild	or30.2	**SaM-*	3	**SaF-*	1

* individuals included in haplotype analysis

Table 7: Original Predicted Crosses

Cross 1	***SaM+ / SaF+***	**x**	***SaM- / SaF-***	***SaM+/ SaM- // SaF+ / SaF-***
	sin02		rr06	**Suffer from sterility**
	rr05		rr22	
	rr39		rr06	
	rr03		rr56	
Cross 2	***SaM+X / SaFX***	**x**	***SaM- / SaF-***	***SaM+X / SaM- // SaFX / SaF-***
	rr07		rr06	**Functional consequences of *SaFX***
	rr08		rr56	
	rr11		rr22	
	rr36		rr56	
Cross 3	***SaM-X / SaFX***	**x**	***SaM+ / SaF-***	***SaM-X / SaM+ // SaFX / SaF-***
	rr15		or18	**Functional consequences of *SaFX***
	rr15		or08	
	rr33		or17	
	rr33		sin08	
Cross 4	***SaM-X / SaFX***	**x**	***SaM+ / SaF+***	***SaM-X / SaM+ // SaFX / SaF+***
	rr15		sin02	**Should be infertile;**
	rr15		rr39	**Functional consequences of *SaFX***
	rr33		rr03	
	rr33		rr05	

-Red squares indicate successful and validated hybrids.
-Cross 1 is still growing in the greenhouse and did not flower in time for pollen analyses.

Table 8: Individuals that failed for original *SaF* amplification

Lab ID	Sub-Pop	Population		Origin
rr20	BHA1	Redrice	1996-9_(BLKH_89)	AR
rr33	BRH	Redrice	1111-01	AR
rr36	BRH	Redrice	1300-02	MO
rr48	BRH	Redrice	1183-01	AR
rr54	BRH	Redrice	1120-02	AR
sin04	*indica*	Dholi Boro	27513	Bangladesh
rr07	SH	Redrice	1098-01	MO
rr08	SH	Redrice	1134-01	AR
rr11	SH	Redrice	1160-01	AR
rr12	SH	Redrice	1179-01	LA
rr15	SH	Redrice	16B	AR
rr17	SH	Redrice	1995-15	AR
rr19	SH	Redrice	1996-5_(SHA_'93)	MS
rr30	SH	Redrice	1047-01	LA
rr31	SH	Redrice	1073-02	MO
rr34	SH	Redrice	1190-01	LA
rr35	SH	Redrice	1199-01	MO
rr42	SH	Redrice	1995-12	LA
rr45	SH	Redrice	1996-8	AR
rr51	SH	Redrice	1210-02	AR
rr53	SH	Redrice	1163-01	LA
str02	*japonica*	Kotobuki Mochi	2545	Japan
or29	Wild	*O.rufipogon*	81990	Myanmar
or01	Wild	*O.rufipogon*	100588	Taiwan
or05	Wild	*O.rufipogon*	100916	China
or07	Wild	*O.rufipogon*	104599	Sri Lanka
or12	Wild	*O.rufipogon*	105388	Thailand
or14	Wild	*O.rufipogon*	105568	Philippines
or03	Wild	*O.rufipogon*	106122	India
or22	Wild	*O.rufipogon*	106150	Laos
or25	Wild	*O.rufipogon*	106321	Cambodia
or02	Wild	*O.rufipogon*	106523	Papau New Guinea
obr02	Wild	*O. barthii*	101226	Mali

SaF Deletion

Table 9: Pollen Non-Viable Averages

	Genotype	*N	°% Non-Viable Avg/ Ind.
P	*SaM+ / SaF+*	3	8.4
			6.1
			8.8
		^SD	6.5
	SaM- / SaF-	4	12.3
			19
			6
			29.5
		SD	14.2
	SaM+ / SaF-	2	8
			3.2
		SD	2.7
	SaM+X / SaFX	4	5.4
			6.6
			15
			6
		SD	5.6
	SaM-X / SaFX	2	6.8
			44
		SD	24.4
F1	*SaM+X / SaM- // SaF-/ SaFX*	2	13.1
			13.2
		SD	5.8
	SaM+ / SaM-X // SaF+ / SaFX	1	5.3
		SD	4.5

*N = number individuals / genotype

° for each individual, three measurements were obtained, listed in this table is the average of those three measurements / individual

^ standard deviations calculated from original raw measurements

Table 10: Pollen Quantity Averages

	Genotype	*N	°Avg/ Ind.
P	*SaM+ / SaF+*	2	21.2
			59.6
		^SD	22.2
	SaM- / SaF-	3	27.47
			46
			97.5
		SD	37.6.29
	SaM+ / SaF-	1	42.4
		SD	11.4
	SaM+X / SaFX	3	26.5
			85.6
			34.47
		SD	14.6
	SaM-X / SaFX	2	32.8
			14.1
		SD	11.2
F1	*SaM- / SaM+X // SaF-/ SaFX*	2	97.2
			27.6
		SD	41.2
	SaM+ / SaM-X // SaF+ / SaFX	1	51.7
		SD	1.2

*N = number individuals / genotype

° for each individual, three measurements were obtained, listed in this table is the average of those three measurements / individual

^ standard deviations calculated from original raw measurements

Table 11: Summary Statistics

	s5			*DPL2*			*SaM*			*SaF*			STS*		
	π	θ	TD	π	θ	TD	π	θ	TD	π	θ	TD	π	θ	TD
BHA1	0.0013	0.002	-1.396	0.00033	0.0007	-1.149	0	0	n/a	0	0	n/a	0.0007	0.0008	-0.177
BHA2	0.00397	0.00292	1.846	0.00108	0.0008	0.986	0.0014	0.00232	-1.609	0.0024	0.00037	-1.055	0.0005	0.0006	0.042
SH	0.00074	0.00114	-1.269	0	0	n/a	0.00193	0.00294	-1.183	0	0	n/a	0.0006	0.0004	-0.441
indica	0.00046	0.00047	-0.05	0.00108	0.00118	-0.612	0.00347	0.00435	-0.901	0.00118	0.00142	-0.689	0.0016	0.0017	-0.026
aus	0.00035	0.00035	n/a	0	0	n/a	0.00194	0.00232	-1.048	0.00083	0.0018	-1.358	0.0012	0.0011	0.092
japonica	0.00252	0.00216	0.826	0.00309	0.00307	0.022	0.00206	0.00278	-1.035	0	0	n/a	0.0011	0.0014	-0.773
wild	0.00213	0.00347	-1.385	0.00328	0.01043	-2.374	0.00716	0.00738	-0.096	0.00291	0.00422	-1.037	0.0044	0.0056	-0.729

* based on averages from Reagon et. al 2010

red text indicates significance

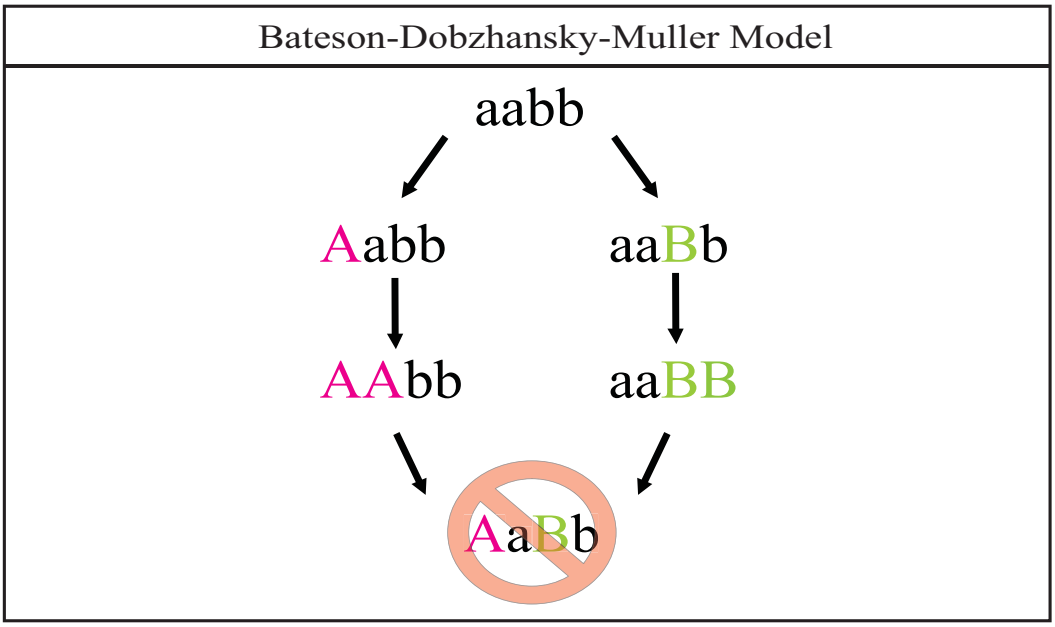

Figure 1: Genetic Model of Hybrid Incompatibility: The Bateson-Dobzhansky-Muller Model starts with an ancestral population that diverges. A mutated allele evolves (A and B) at different loci and become fixed within each population. Incompatibility results from the deleterious interaction of A and B in the F1

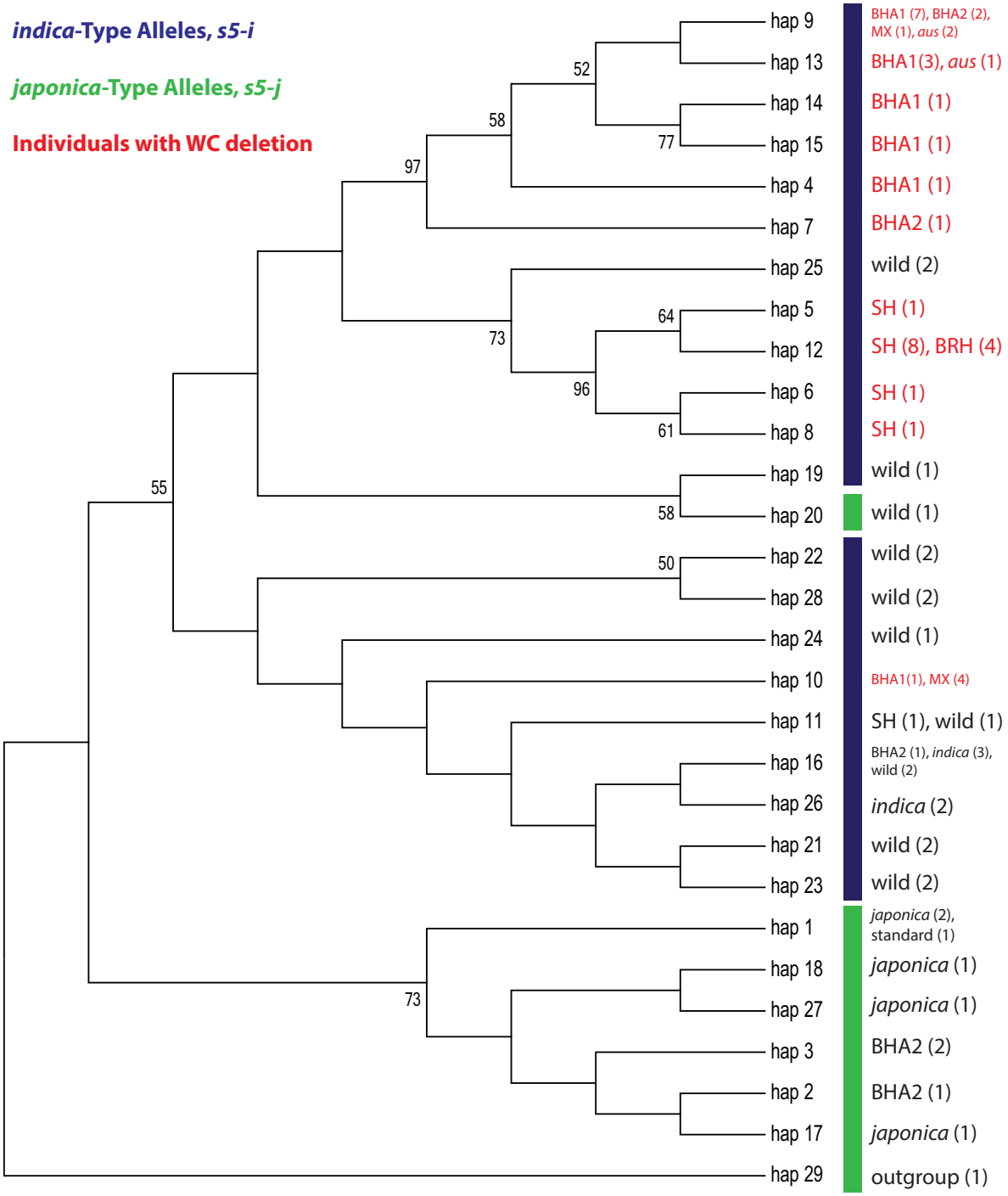

Figure 2. Haplotype Tree for *s5*: Neighbor-joining tree inferred from 500 replicates. Bootstraps below 50 not shown. Evolutionary distances were computed using a Kimura-2-parameter. WC = wide-compatibility. Outgroup = *O. barthii*.

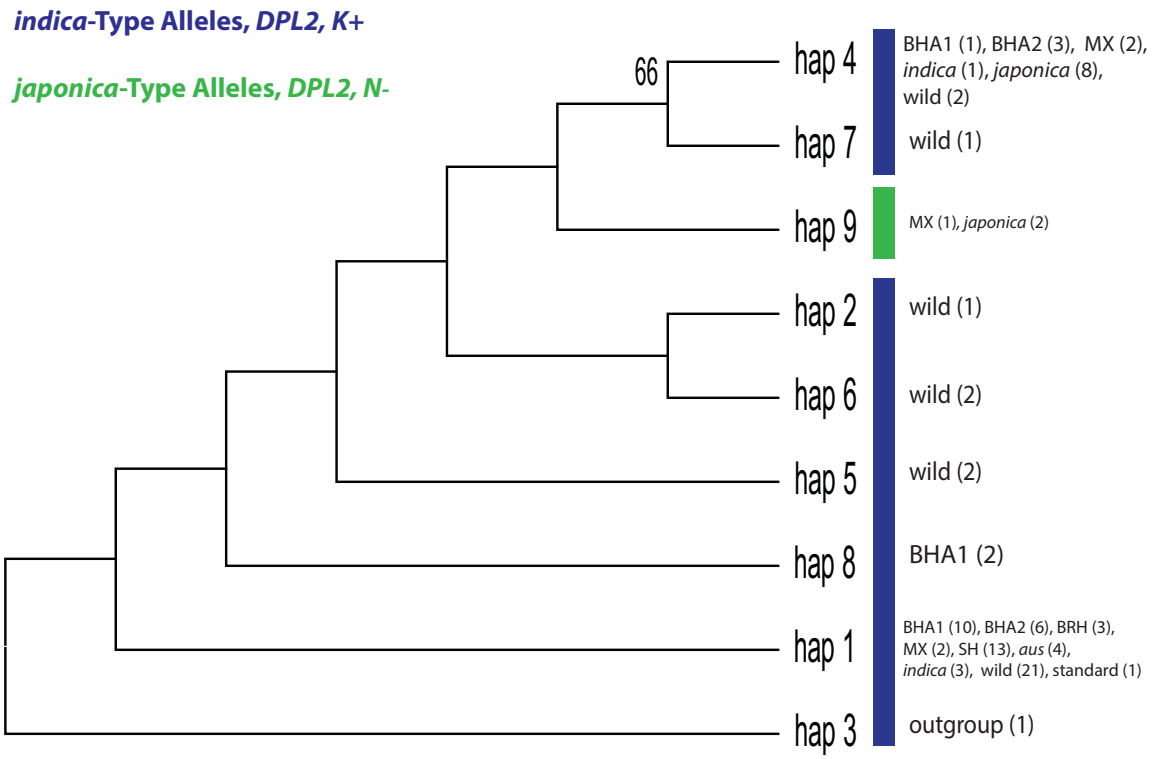

Figure 3. Haplotype Tree for *DPL2*: Neighbor-joining tree inferred from 500 replicates. Bootstraps below 50 not shown. Evolutionary distances were computed using a Kimura-2-parameter. Outgroup = *O. meridionalis*.

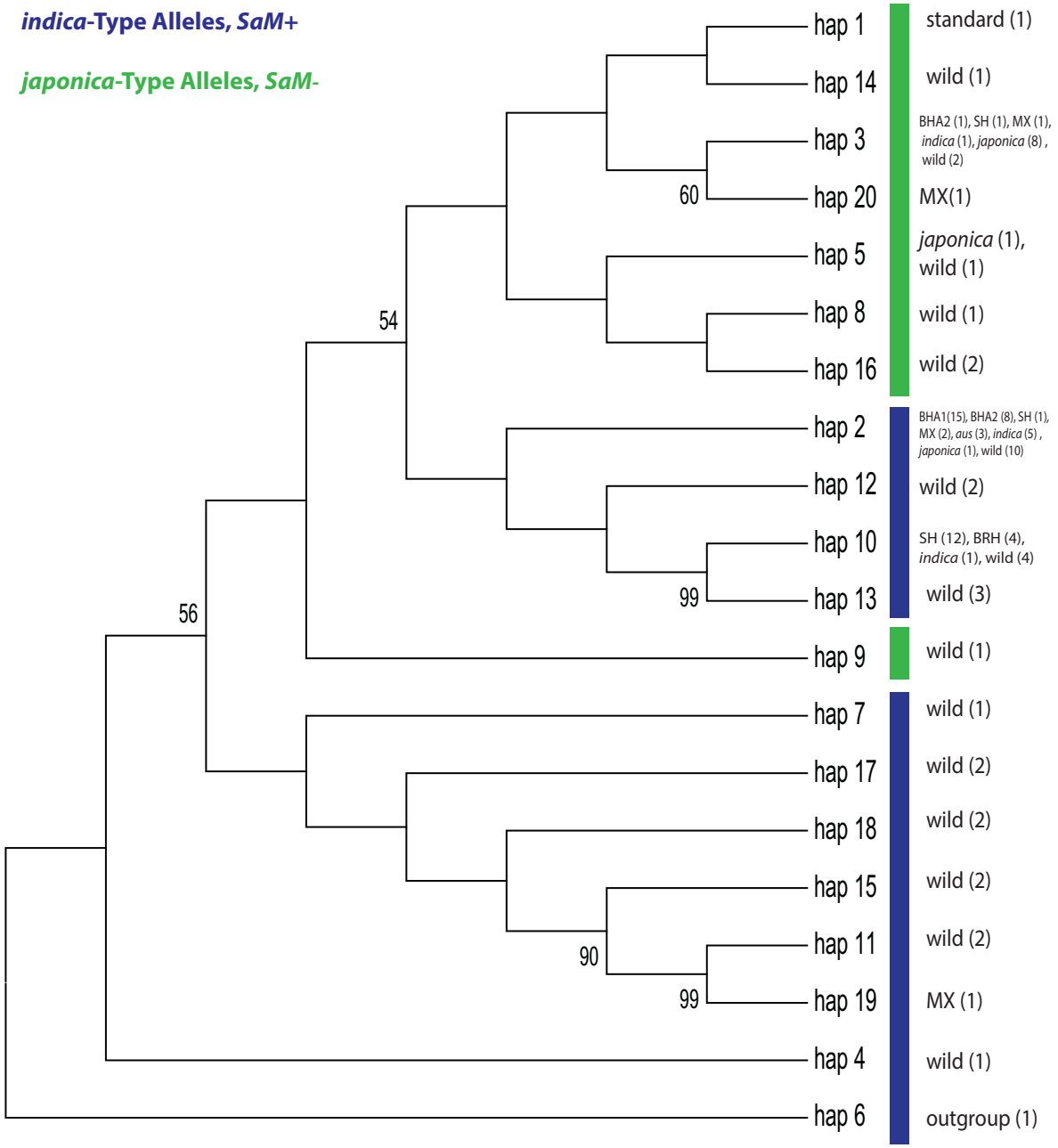

Figure 4. Haplotype Tree for *SaM*: Neighbor-joining tree inferred from 500 replicates. Bootstraps below 50 not shown. Evolutionary distances were computed using a Kimura-2-parameter. Outgroup = *O. meridionalis*.

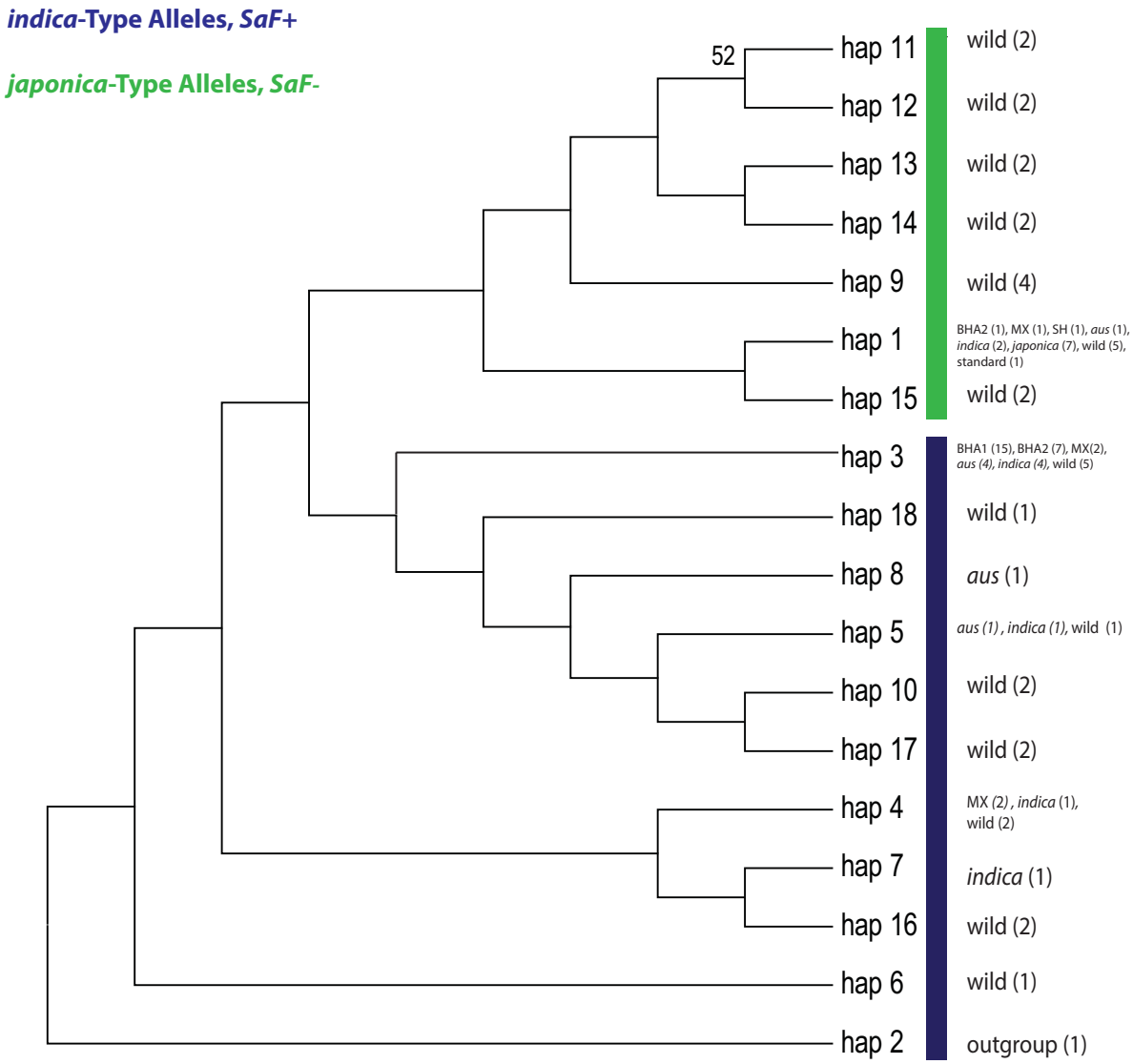

Figure 5. Haplotype Tree for *SaF*: Neighbor-joining tree inferred from 500 replicates. Bootstraps below 50 not shown. Evolutionary distances were computed using a Kimura-2-parameter. Outgroup = *O. meridionalis*.

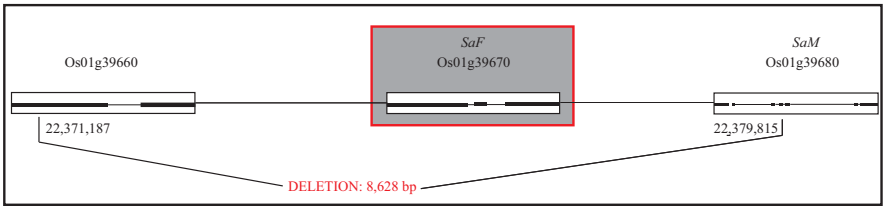

Figure 6. *SaFX* Deletion: *SaFX* is characterized by a 8,628 bp deletion starting in the gene upstream of *SaF*, Os01g39660 (a putative transposon protein), through the first four exons of *SaM*.

BIBLIOGRAPHY

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.

Chen, M., Zhao, Z., Jiang, L., and Wan, J. (2012). A new gene controlling hybrid sterility in rice (Oryza sativa L.). Euphytica *184*, 15–22.

Clark, A.G. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol *7*, 111–122.

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.

Fischer, A., and Ramirez, A. (1993). Red Rice (oryza-Sativa) - Competition Studies for Management Decisions. Int. J. Pest Manage. *39*, 133–138.

Gao, H., Williamson, S., and Bustamante, C.D. (2007). A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data. Genetics *176*, 1635–1651.

Gealy, D.R., and Gressel, J. (2005). Gene movement between rice (Oryza sativa) and weedy rice (Oryza sativa) - a U.S. temperate rice perspective. 323–354.

Harushima, Y., Nakagahra, M., Yano, M., Sasaki, T., and Kurata, N. (2002). Diverse variation of reproductive barriers in three intraspecific rice crosses. Genetics *160*, 313–322.

Ji, Q., Zhang, M., Lu, J., Wang, H., Lin, B., Liu, Q., Chao, Q., Zhang, Y., Liu, C., Gu, M., et al. (2012a). Molecular basis underlying the S5-dependent reproductive isolation and compatibility of indica/japonica rice hybrids. Plant Physiol. *158*, 1319–1328.

Ji, Q., Zhang, M., Lu, J., Wang, H., Lin, B., Liu, Q., Chao, Q., Zhang, Y., Liu, C., Gu, M., et al. (2012b). Molecular Basis Underlying the S5-Dependent Reproductive Isolation and Compatibility of indica/japonica Rice Hybrids. Plant Physiol. *158*, 1319–1328.

Kohls, S., Stamp, P., Knaak, C., and Messmer, R. (2011). QTL involved in the partial restoration of male fertility of C-type cytoplasmic male sterility in maize. Theor Appl Genet *123*, 327–338.

Kubo, T., Yamagata, Y., Eguchi, M., and Yoshimura, A. (2008). A novel epistatic interaction at two loci causing hybrid male sterility in an inter-subspecific cross of rice (Oryza sativa L.). Genes Genet. Syst. *83*, 443–453.

Lawton-Rauh, A., and Burgos, N. (2010). Cultivated and weedy rice interactions and the domestication process. Molecular Ecology *19*, 3243–3245.

LeiGang, S., XiangDong, L., Bo, L., XingJuan, Z., Lan, W., JinQuan, L., and YongGen, L. (2009). Identifying neutral allele Sb at pollen-sterility loci in cultivated rice with Oryza rufipogon origin. Chin. Sci. Bull. *54*, 3813–3821.

Li, D., Chen, L., Jiang, L., Zhu, S., Zhao, Z., Liu, S., Su, N., Zhai, H., Ikehashi, H., and Wan, J. (2007). Fine mapping of S32(t), a new gene causing hybrid embryo sac sterility in a Chinese landrace rice (Oryza sativa L.). Theor Appl Genet *114*, 515–524.

Li, J.Q., Shahid, M.Q., Feng, J.H., Liu, X.D., Zhao, X.J., and Lu, Y.G. (2012). Identification of neutral alleles at pollen sterility gene loci of cultivated rice (Oryza sativa L.) from wild rice (O. rufipogon Griff.). Plant Syst. Evol. *298*, 33–42.

Li, W., Zeng, R., Zhang, Z., Ding, X., and Zhang, G. (2008). Identification and fine mapping of S-d, a new locus conferring the partial pollen sterility of intersubspecific F-1 hybrids in rice (Oryza sativa L.). Theor. Appl. Genet. *116*, 915–922.

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, Z., Pinson, S.R.M., Paterson, A.H., Park, W.D., and Stansel, J.W. (1997). Genetics of Hybrid Sterility and Hybrid Breakdown in an Intersubspecific Rice (Oryza sativa L.) Population. Genetics *145*, 1139–1148.

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

Londo, J.P., and Schaal, B.A. (2007). Origins and population genetics of weedy red rice in the USA. Mol. Ecol. *16*, 4523–4535.

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.

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.

Nei, M., and Li, W. (1979). Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. Proc. Natl. Acad. Sci. U. S. A. *76*, 5269–5273.

Oka, H.-I. (1957). Genic analysis for the sterility of hybrids between distantly related varieties of cultivated rice. J Genet *55*, 397–409.

Orr, H.A. (1996). Dobzhansky, Bateson, and the genetics of speciation. Genetics *144*, 1331–1335.

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

Reagon, M., Thurber, C.S., Gross, B.L., Olsen, K.M., Jia, Y., and Caicedo, A.L. (2010). Genomic patterns of nucleotide diversity in divergent populations of US weedy rice. BMC Evol. Biol. *10*,.

Reflinur, Chin, J.H., Jang, S.M., Kim, B., Lee, J., and Koh, H.-J. (2012). QTLs for hybrid fertility and their association with female and male sterility in rice. Genes Genom. *34*, 355–365.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. *132*, 365–386.

Shivrain, V.K., Burgos, N.R., Gealy, D.R., Sales, M.A., and Smith, K.L. (2009). Gene flow from weedy red rice (Oryza sativa L.) to cultivated rice and fitness of hybrids. Pest Management Science *65*, 1124–1129.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. *28*, 2731–2739.

Thurber, C.S., Reagon, M., Gross, B.L., Olsen, K.M., Jia, Y., and Caicedo, A.L. (2010). Molecular evolution of shattering loci in US weedy rice. Mol. Ecol. *19*, 3271–3284.

Vinod, A.A., Tomer, S.M.S., and Chandl, S. (2011). Genetics of fertility restoration and test for allelism of restorer genes in wheat (Triticum aestivum L.). Indian J. Genet. Plant Breed. *71*, 223–230.

Wang, G.W., Cai, H.Y., Xu, Y., Yang, S.H., and He, Y.Q. (2009). Analysis on the spectra and level of wide-compatibility conferred by three neutral alleles in inter-subspecific hybrids of rice (Oryza sativa L.) using near isogenic lines. Plant Breed. *128*, 451–457.

Wang, Y., Zhong, Z.Z., Zhao, Z.G., Jiang, L., Bian, X.F., Zhang, W.W., Liu, L.L., Ikehashi, H., and Wan, J.M. (2010). Fine mapping of a gene causing hybrid pollen sterility between Yunnan weedy rice and cultivated rice (Oryza sativa L.) and phylogenetic analysis of Yunnan weedy rice. Planta *231*, 559–570.

Widmer, A., Lexer, C., and Cozzolino, S. (2009). Evolution of reproductive isolation in plants. Heredity *102*, 31–38.

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.

Yang, C.Y., Chen, Z.Z., Zhuang, C.X., Mei, M.T., and Liu, Y.G. (2004). Genetic and physical fine-mapping 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., 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.

YiDan, O., JiongJiong, C., JiHua, D., and QiFa, Z. (2009). Advances in the understanding of inter-subspecific hybrid sterility and wide-compatibility in rice. Chin. Sci. Bull. *54*, 2332–2341.

You-Xin, Y., Yan-Hong, L., Jing-Fei, T., Qasim, S.M., Zhi-Xiong, C., Lan, W., Jin-Quan, L., Xiang-Dong, L., and Yong-Gen, L. (2012). Wide-Compatibility Gene S-5(n) Exploited by Functional Molecular Markers and Its Effect on Fertility of Intersubspecific Rice Hybrids. Crop Sci. *52*, 669–675.

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.

Zhao, J., Li, J., Xu, P., Zhou, J., Hu, F., Deng, X., Deng, W., and Tao, D. (2012). A new gene controlling hybrid sterility between Oryza sativa and Oryza longistaminata. Euphytica *187*, 339–344.

Zhao, Z.G., Zhu, S.S., Zhang, Y.H., Bian, X.F., Wang, Y., Jiang, L., Liu, X., Chen, L.M., Liu, S.J., Zhang, W.W., et al. (2011). Molecular analysis of an additional case of hybrid sterility in rice (Oryza sativa L.). Planta *233*, 485–494.