DOI: 10.1534/genetics.110.116772

A Genetic Model for the Female Sterility Barrier Between Asian and African Cultivated Rice Species

Andrea Garavito,*,† Romain Guyot,* Jaime Lozano,† Frédérick Gavory,‡ Sylvie Samain,‡ Olivier Panaud,§ Joe Tohme,† Alain Ghesquière* and Mathias Lorieux*,†,[1]

**Plant Genome and Development Laboratory, Institut de Recherche pour le Développement (IRD), 34394 Montpellier Cedex 5, France,* †*Agrobiodiversity and Biotechnology Project, International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia,* ‡*Génoscope, Institut de Génomique, Commissariat à l'Énergie Atomique (CEA), 91057 Evry, France and* §*Plant Genome and Development Laboratory, Université de Perpignan, 66860 Perpignan, France*

Manuscript received March 18, 2010
Accepted for publication April 28, 2010

ABSTRACT

S$_1$ is the most important locus acting as a reproductive barrier between *Oryza sativa* and *O. glaberrima*. It is a complex locus, with factors that may affect male and female fertility separately. Recently, the component causing the allelic elimination of pollen was fine mapped. However, the position and nature of the component causing female sterility remains unknown. To fine map the factor of the *S*$_1$ locus affecting female fertility, we developed a mapping approach based on the evaluation of the degree of female transmission ratio distortion (fTRD) of markers. Through implementing this methodology in four *O. sativa* × *O. glaberrima* crosses, the female component of the *S*$_1$ locus was mapped into a 27.8-kb (*O. sativa*) and 50.3-kb (*O. glaberrima*) region included within the interval bearing the male component of the locus. Moreover, evidence of additional factors interacting with *S*$_1$ was also found. In light of the available data, a model where incompatibilities in epistatic interactions between *S*$_1$ and the additional factors are the cause of the female sterility barrier between *O. sativa* and *O. glaberrima* was developed to explain the female sterility and the TRD mediated by *S*$_1$. According to our model, the recombination ratio and allelic combinations between these factors would determine the final allelic frequencies observed for a given cross.

INTRINSIC postzygotic reproductive barriers are a very common and important phenomenon, driving the establishment and conservation of species (Widmer *et al.* 2009). They are observed as a reduction in the hybrids' fitness, evidenced by a reduced viability or fertility. So far, a limited number of genes acting as postzygotic barriers have been identified in Drosophila (Ting *et al.* 1998; Presgraves *et al.* 2003; Brideau *et al.* 2006; Masly *et al.* 2006; Phadnis and Orr 2009; Tang and Presgraves 2009), and in other animal systems (Wittbrodt *et al.* 1989; Lee *et al.* 2008; Mihola *et al.* 2009), but also in Arabidopsis (Josefsson *et al.* 2006; Bomblies *et al.* 2007; Bikard *et al.* 2009) and rice (Chen *et al.* 2008; Long *et al.* 2008).

In rice improvement, hybrid inviability and sterility are major obstacles for the common utilization of closely related species in breeding programs, impairing the exploitation of the rich genetic diversity found within the *Oryza sativa* complex (known as genome group AA), and the beneficial effect of the high level of heterosis observed in the F_1 plants. One of the most relevant examples of this strong limitation comes from the African cultivated rice species *O. glaberrima* Steud. This species represents an interesting source of drought tolerance (Sarla and Swamy 2005), weed competitiveness (Dingkuhn *et al.* 1998), and nematode and virus resistances (Ndjiondjop *et al.* 1999; Soriano *et al.* 1999). Several of these traits have been already mapped (Lorieux *et al.* 2003; Ndjiondjop *et al.* 2003); however, their introduction to *O. sativa* has been hampered by the strong sterility barrier between the two species (Sano *et al.* 1979). This barrier is the result of several loci that might interact or act separately to render the F_1 hybrids pollen sterile and partially female fertile. Among these, the *S*$_1$ locus has the strongest effect over the fertility of the hybrids (Sano 1990).

Despite the strong hybrid sterility, obtaining fertile plants derived from *O. sativa* × *O. glaberrima* crosses has been possible by performing successive backcrosses followed by selfing (Ghesquière *et al.* 1997). Nevertheless, this fertility recovery is associated with the presence of the homozygote *O. glaberrima S*$_1$ allele (S_1^g), and is lost again when recrossing with *O. sativa* (Heuer and Miezan 2003). Furthermore, a strong transmission ratio

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.110.116772/DC1.

Sequence data from this article have been deposited with the EMBL Library under accession number FP340543.

[1]*Corresponding author:* Institut de Recherche pour le Développement (IRD), 911 Av. Agropolis, 34394 Montpellier Cedex 5, France.
E-mail: mathias.lorieux@ird.fr

distortion (TRD) of markers linked with S_1, in favor of the *O. glaberrima* alleles, results as a consequence of the systematic elimination of the *O. sativa* alleles from the descendants (Ghesquière *et al.* 1997; Doi *et al.* 1998b; Lorieux *et al.* 2000; Aluko *et al.* 2004).

As an alternative intended to unlock the full potential of *O. glaberrima* to rice breeding, the "iBridges" project, supported by the Generation Challenge Program (http://www.generationcp.org) aims to develop a set of *O. sativa* × *O. glaberrima* hybrids that would be fertile when crossed with *O. sativa*. To develop these interspecific bridges, genetic factors affecting their fertility must be identified and characterized so as to better understand the nature of the sterility barrier. Recently, the complex nature of the S_1 locus was suggested, and the male component was mapped to an interval equivalent to 45 kb on the genome of *O. sativa* cv. Nipponbare (Koide *et al.* 2008c). However, the nature and the genomic positions of the components affecting female fertility are still unknown.

In this article we describe the fine genetic and physical mapping of the female component of the S_1 locus, using an approach based on the evaluation of the degree of TRD caused by the elimination of the S_1 *O. sativa* alleles. This methodology enabled us to identify a 27.8-kb interval on the short arm of the *O. sativa* chromosome 6 as the location of the female component of the S_1 locus. Furthermore, the presence of additional factors involved in the sterility barrier mechanisms was predicted, and their localization was deduced. Direct comparison of the S_1 interval in *O. glaberrima* with the *O. sativa* orthologous sequence revealed the presence of several significant differences and a possible candidate factor. On the basis of our data, a model involving the epistatic interaction of the additional factors with S_1 was developed to explain the female sterility and the TRD found in the interspecific hybrids.

MATERIALS AND METHODS

Plant materials: BC_1F_1 populations were obtained from crosses between the *O. sativa* accessions IR64 (*O. sativa* ssp. *indica*), Caiapo (CA) (tropical *japonica*), Curinga (CU) (tropical *japonica*) and Nipponbare (NIP) (temperate *japonica*), and the three *O. glaberrima* accessions, TOG5681, Ac103544 (MG12), and CG14. Backcrosses were made as follows: IR64/TOG5681//IR64 (referred as IR64xTOG5681), CA/MG12//CA (CAxMG12), CU/CG14//CU (CUxGG14), and NIP/MG12//NIP (NIPxMG12). A first set of 125 individuals from the cross IR64xTOG5681 was generated, sowed, and genotyped. Later on, 114 individuals from the cross CAxMG12, 101 from CUxCG14 and 59 from NIPxMG12 were generated, as well as a second set of 334 plants from the IR64xTOG5681 cross. Additional plants issued from crosses NIP/TOG5681//NIP (12) and CA/TOG5681//CA (34) were also genotyped. Seeds were first germinated *in vitro* in MS medium (Toshio and Folke 1962). Ten-day-old seedlings were transplanted in the greenhouse and afterward brought to the field at the International Center for Tropical Agriculture (CIAT, Cali, Colombia) headquarters.

Molecular marker analysis: To derive a new *O. sativa* × *O. glaberrima* genetic map, 140 markers belonging to version 1 of the Universal Core Genetic Map (UCGM) for rice (Orjuela *et al.* 2010) were evaluated in the first 125 BC_1F_1 individuals of the IR64xTOG5681 population. Marker saturation was carried out in the S_1 region on chromosome 6 (Sano 1990; Lorieux *et al.* 2000) (supporting information, Table S1). The rest of the chromosome was also saturated at a lower density. SSR markers were chosen on the basis of the physical location on the *O. sativa* cv. Nipponbare chromosome 6 (TIGR v. 6 data, (Ouyang *et al.* 2007). Primers were synthesized according to the SSR information available (International Rice Genome Sequencing Project 2005). Further saturation near S_1 was performed using new additional markers (Table S2) found with the MISA script (Thiel *et al.* 2003), and InDel markers were designed by comparing the Nipponbare sequence with the orthologous CG14 sequence (see below). Primers were designed with primer3 (Untergasser *et al.* 2007), confirmed to amplify a unique genomic region in Nipponbare using the Primer Blaster tool (Droc *et al.* 2009), and evaluated in the populations.

PCR reactions were carried out with an annealing temperature and magnesium concentration ranging from 1.5 to 2.5 mM optimized for each marker (Table S1 and Table S2), as previously described (Orjuela *et al.* 2010). Electrophoreses of PCR products were made in 4% agarose and revealed with ethidium bromide for polymorphisms greater than 12 bp, and in a Li-Cor sequencer (Li-Cor Biosciences) for smaller polymorphisms, using an M13 tag (IRD700 and IRD800).

Statistical analysis and mapping approach: Genetic map construction was made using the MapDisto v. 1.7 program (http://mapdisto.free.fr/). A minimal LOD score of 3.0 was retained to identify the linkage groups. The order of the markers was determined using the Order, Ripple, and Bootstrap commands. The Kosambi function (Kosambi 1944) was used to convert the recombination fractions to centimorgans (cM) and vice versa. The deviation from the expected Mendelian ratio (1:1) for each locus was determined by segregation χ^2 tests. Genotyping error candidates were detected using the tools provided in the Color Genotypes module, with a threshold of 0.005.

To fine map the female component of the S_1 locus, we used the estimation of the female TRD (fTRD) as an indicator of the closeness of a given marker to the factor causing the distortion. Indeed, it is expected that the maximum fTRD occurs at the location of the transmission ratio distortion locus (TRDL), the precision of the mapping depending on the population size, the marker saturation, and the level of TRD. We conducted a simulation study to estimate the precision of this approach (see File S1, Simulation.doc). Simulation and computations were performed using the MapDisto program (http://mapdisto.free.fr/). fTRD was measured using the k_f statistic, which represents the proportion of plants inheriting the overrepresented allele after female segregation of the heterozygote. Because the TRD on the S_1 locus in our BC_1F_1 populations depends solely on female gametogenesis in the F_1, the k_f and the k statistics are equivalent and were thus calculated as the number of heterozygotes over the total number of BC_1F_1 individuals.

Physical mapping of the S_1 locus: For the *O. glaberrima* cv. CG14, several BAC clones of the S_1 region were chosen on the basis of the available finger print contig (FPC) data from the OMAP project (Kim *et al.* 2008). Clones showing strong sequence identity in at least one of their BAC end sequences (BES) to the orthologous Nipponbare region were identified by the BLAST algorithm (Altschul *et al.* 1997). PCR confirmation was made with the same markers used for the genotyping of the S_1 region. A set of BAC clones was chosen,

and their overlapping was confirmed by PCR amplification with primers derived from their BES.

Sequence analysis, comparison and gene annotation method: BAC sequencing was performed by the Sanger method. The final BAC sequence was first analyzed using BLAST against different public and local plant nucleotide and protein databases. Coding regions were *ab initio* predicted using the FGENESH program (Salamov and Solovyev 2000) and then confirmed by comparative analysis with annotated gene models and proteins in *O. sativa* cv. Nipponbare, downloaded from the TIGR database (Ouyang *et al.* 2007). Predicted gene structures were manually evaluated by alignment with rice EST and full-length cDNA (FLcDNA) public sequences (Kikuchi *et al.* 2003). Detailed analysis was performed with the EMBOSS Analysis software (Rice *et al.* 2000).

Putative transposable elements (TEs) were first identified and annotated by RepeatMasker searches (http://www.repeatmasker.org) against local databases of rice TEs downloaded from the REPBASE (Jurka *et al.* 2005), and from the TIGR repeat database (Ouyang and Buell 2004). *De novo* prediction of TEs was performed according to the structure of the different class of TEs. The final annotation of the BAC sequences was performed using the Artemis tool (Rutherford *et al.* 2000), and comparison with the Nipponbare genome was accomplished using dot-plot alignments with the Dotter software (Sonnhammer and Durbin 1995).

Calculation of nonsynonymous and synonymous nucleotide substitution rates: Pairwise alignments of orthologous genes between *O. glaberrima* and *O. sativa* were conducted using the Needle tool (Rice *et al.* 2000) to estimate the degree of gene structure conservations. Rates of nonsynonymous and synonymous nucleotide substitutions (K_a/K_s) were calculated using the PAML package (Yang 1997) implemented in the PAL2NAL program (Suyama *et al.* 2006). Orthologous genes with clear distinct gene structures were removed from the analysis.

RESULTS

Genetic map of the IR64xTOG5681 cross, marker segregation and its association with sterility loci between *O. sativa* and *O. glaberrima:* Using a population of 125 individuals from the IR64xTOG5681 cross, a survey of the entire genome was completed to identify markers with fTRD and to compare their colocalization with previously described sterility loci between the two species. Using markers from the UCGM, an 1879.56-cM map length was obtained (Figure 1). Possible inversions were detected in centromeric regions (chromosomes 1, 5, 7, 8, 9, 11) and in the short arm of chromosome 10 (Figure S1). Several sites showing a significant deviation from the 1:1 expected segregation ($P < 0.05$) were also found (Figure 1). The *O. glaberrima* alleles were overrepresented on chromosomes 3, 6, and 11, while the *O. sativa* alleles were more common than expected on chromosomes 1, 2, and 7.

On the basis of marker information available from Gramene (Liang *et al.* 2008), markers surrounding the different sterility loci between the two species were localized on the map. All but two sites presenting fTRD (long arms of chromosomes 2 and 6) colocalized with the different loci described so far (Figure 1). The strongest fTRD ($P < 10^{-5}$) was found at the S_1 locus. Nonetheless, several sterility loci did not colocalize with markers showing fTRD. These results corroborate the relation between TRD and sterility loci.

Mapping of the female component of the S_1 locus by measuring the TRD: Recently, Koide and Colleagues (2008c) proved that the TRD of the *Wx* locus in pollen produced by heterozygote plants for the S_1 region was due to the abortion of gametes carrying the highly linked *O. sativa* S_1 allele (denoted in this article as S_1^s and equivalent to S_1^a (Sano 1990)). They also inferred that the S_1 locus has at least two components, one controlling male TRD (mTRD) and the other one fTRD. Additionally, the male component of the locus was mapped to a 45-kb interval in the chromosome 6 of Nipponbare.

To map the female component of the S_1 locus, we applied an approach based on the direct measurement of marker deviation from the expected Mendelian ratio, caused by the elimination of female gametes carrying the S_1^s allele. Due to meiotic recombination, the degree of TRD of a given marker would be inversely proportional to the genetic distance between it and the distorter factor. By constructing a highly saturated genetic map around S_1, and measuring precisely the fTRD observed for each marker, it would be possible to restrict the genomic location of the distorter factor, since the markers closest to S_1 would show the highest fTRD.

To examine the fTRD level, a saturated genetic map was obtained by two succeeding marker saturations in the S_1 region, the first one with approximately 1 SSR/100 kb, carried within the segment limited by the markers showing the highest fTRD in a previous work (RM190 and RM204, Lorieux *et al.* 2000), and the second one with 1 SSR/40 kb within the interval bounded by the two markers with the highest TRD (RM19357 and RM19369). Using the 125 IR64xTOG5681 BC_1F_1 population, an fTRD peak of $k_f = 0.970$ was found, limited by markers RM19357 and RM19367 (Figure 2A). On the basis of these data, the location for the female component of S_1 was attributed to an interval equivalent to 160 kb in the Nipponbare genome.

Mapping of the female component of the S_1 locus in the *O. sativa* ssp. *japonica* background: As the *O. sativa* genetic background affects the level of female sterility caused by S_1 (Sano 1990; Koide *et al.* 2008c), it is necessary to analyze crosses involving both *indica* and *japonica* to better understand the nature of the sterility barrier. For that purpose, three *O. sativa* ssp. *japonica* × *O. glaberrima* BC_1F_1 populations were developed, and the chromosome 6 genetic map was constructed and saturated.

A peak in the fTRD of $k_f = 0.912$ (limited by RM19353 and RM19367), $k_f = 0.911$ (RM190–RM5199), and $k_f = 0.881$ (RM19357–RM19367), was found for crosses CAxMG12 (Figure 2B), CUxCG14 (Figure 2C), and

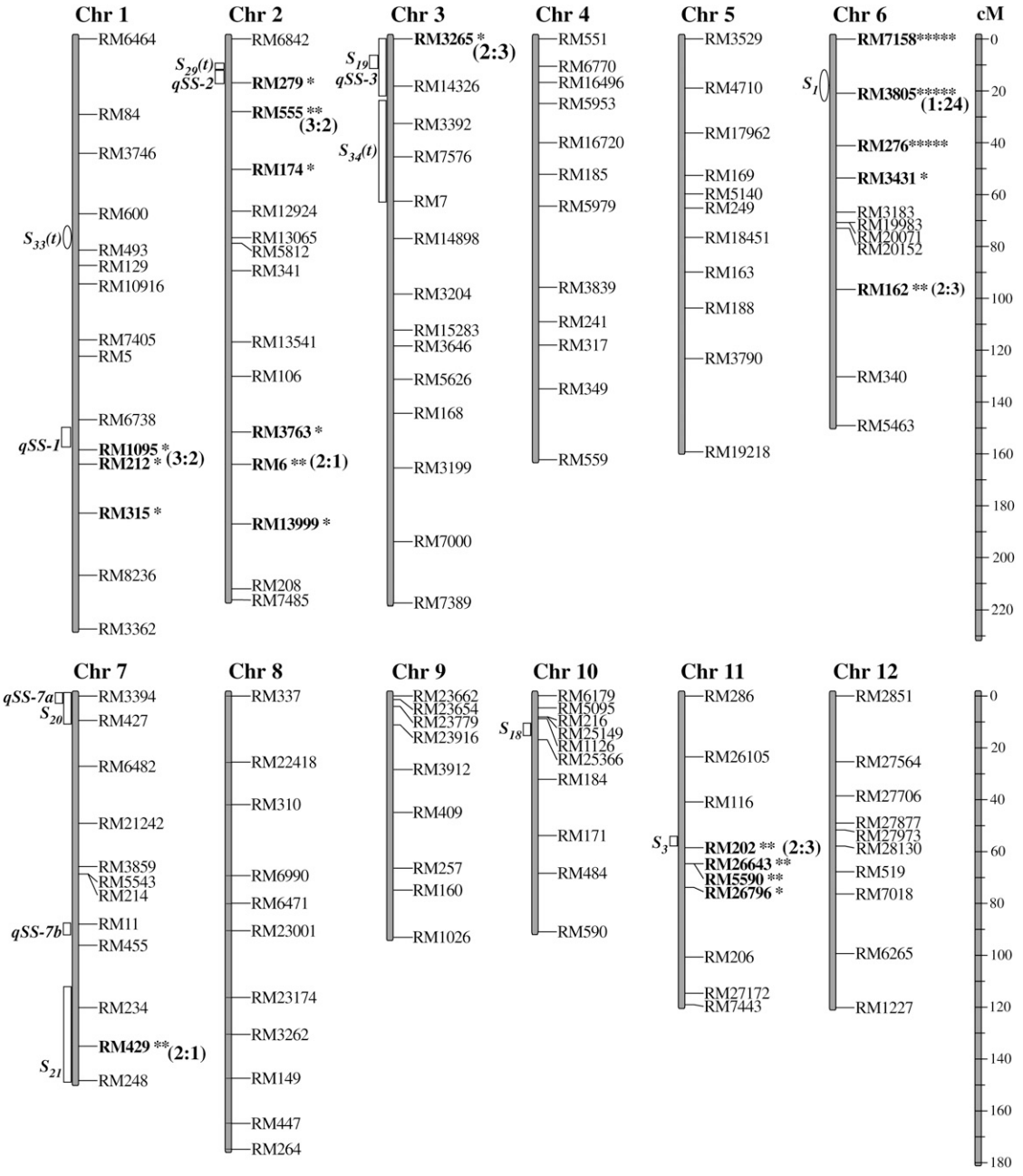


FIGURE 1.—Genetic map of the *O. sativa* cv. IR64 × *O. glaberrima* cv. TOG5681 cross, fTRD genomic sites, and previously described sterility loci. Genetic map constructed with 125 individuals of the IR64 × TOG5681 backcross, using 140 anchors from the UCGM. Markers presenting a significant TRD with (*) $P < 0.05$, (**) $P < 0.01$, and (*****) $P < 10^{-5}$ are indicated with their corresponding segregation ratio (*O. sativa*/*O. sativa*: *O. glaberrima*/*O. sativa*). The location of each sterility locus was inferred from the position of markers used to describe it on the Nipponbare chromosome 6 (S_1, SANO 1990; S_3, SANO 1983; S_{18}, DOI *et al.* 1998a; S_{19}, TAGUCHI *et al.* 1999; S_{20}, DOI *et al.* 1999; S_{21}, DOI *et al.* 1999; $S_{29}(t)$, HU *et al.* 2006; $S_{33}(t)$, REN *et al.* 2005; $S_{34}(t)$, ZHANG *et al.* 2005; *qSS-1, qSS-2, qSS-7a* and *qSS7b*, J. LI *et al.* 2008). Open bars are used to represent loci involved in pollen sterility, while ellipses represent loci affecting male and female fertility.

NIPxMG12 (Figure 2D), respectively. Also, different levels of significant TRD were found in the centromeric region according to the cross. This marker evaluation shows that the fTRD caused by S_1 is located in the same genomic region in different *O. glaberrima* and *O. sativa indica* or *japonica* accessions, as it involves the same molecular markers for all of the crosses. However, the degree of female gamete elimination varies between the different *japonica* populations and is lower than the one found for the *indica* cross.

Construction of a physical map around the S_1 locus in *O. glaberrima:* To compare the region between the two species, a physical map of the S_1 locus of *O. glaberrima* was established. Twenty-six BAC clones from the *O. glaberrima* cv. CG14 BAC library were chosen, on the basis of data available from the OMAP project (KIM

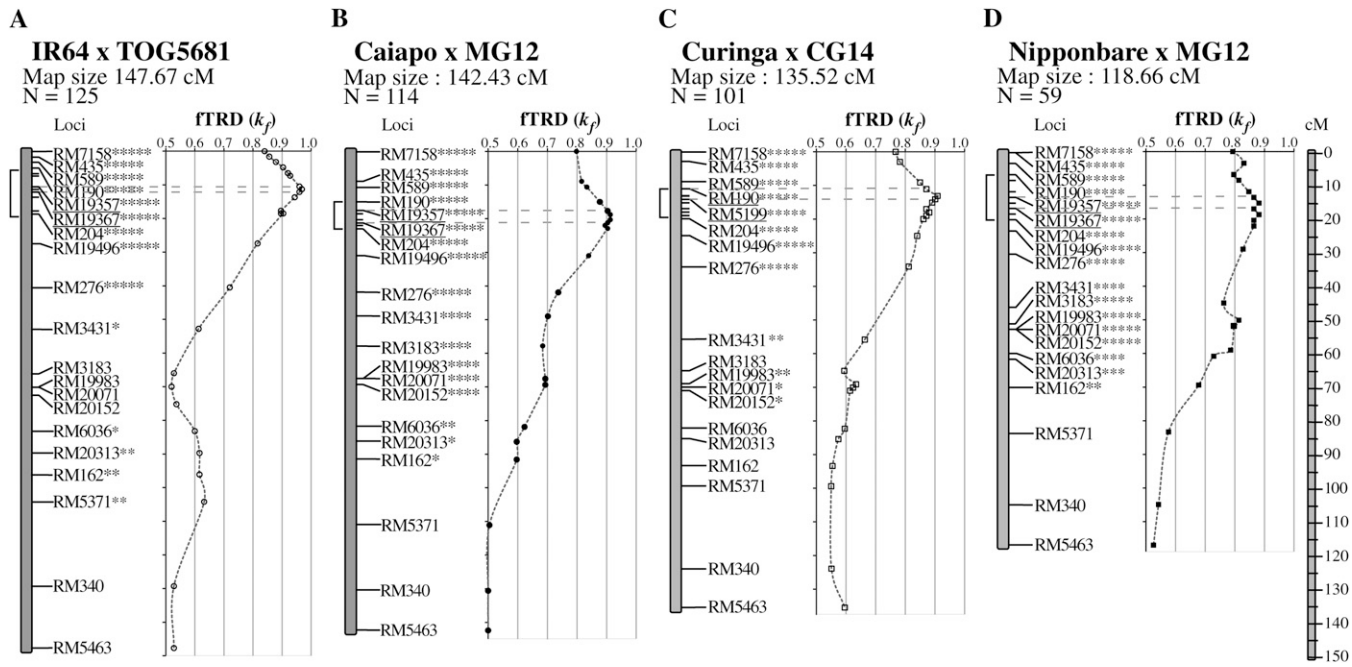


FIGURE 2.—Mapping of the female component of the S_1 locus using the fTRD measure. Chromosome 6 genetic map for (A) IR64xTOG5681, (B) CAxMG12, (C) CUxCG14, and (D) NIPxMG12 populations. The fTRD value (k_f) is indicated for each marker. Significant fTRD is shown (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 10^{-3}$; ****, $P < 10^{-4}$; and *****, $P < 10^{-5}$). Underlined markers limit the fTRD peak for each cross. For each cross, only relevant markers are shown; see Figure 3 for the complete list of markers.

et al. 2008). PCR amplifications with the same markers from the genetic map around S_1 confirmed that 14 belonged to the genomic region. A minimum tiling path (MTP) of 8 clones (covering 800 kb in the Nipponbare genome) was established after PCR confirmations of the overlaps between the clones (Figure 3A). On the basis of the amplification profile, clone OG-BBa0049I08 was selected for complete sequencing, as it contains the markers with the highest fTRD. The final 165-kb sequence for the BAC (EMBL accession number FP340543) was obtained with 12.5× coverage and an error rate below 1 base/100 kb.

Fine mapping of the female component of S_1: The factor causing fTRD was mapped in the same genomic region for the two *O. sativa* backgrounds. To map it more finely, new polymorphic SSR and InDel markers were designed by sequence comparison between the two species. High saturation with new markers (close to one marker per gene) was completed in the existing populations and in 334 new BC_1F_1 individuals from the IR64xTOG5681 cross.

An fTRD peak limited by markers RM19357 and RM19359 was found for both IR64xTOG5681 (k_f = 0.980) and CAxMG12 (k_f = 0.921) crosses (Figure 3, B and C). For the CUxCG14 and NIPxMG12 crosses, the values for k_f were 0.911 (RM190–RMC6_22639) and 0.881 (RM19357–RM19367), respectively (Figure 3, D and E). Consequently, the segment between markers RM19357 and RM19359 was defined as the location for the factor affecting female fertility of the hybrids.

To reduce the size of the candidate interval, the recombination events were examined around the fTRD peak using additional markers designed for that purpose. Surprisingly, recombination occurred within the same small segment for the three individuals limiting each side of the interval (Figure 3F). This approach allowed us to reduce the S_1 female component interval to 27.8 kb in *O. sativa* and to 50.3 kb in *O. glaberrima*. This segment is positioned within the interval described to contain the male component of the locus (KOIDE *et al.* 2008c). The fine mapping of the female component of S_1 indicates that the factor(s) responsible for the female and male semi-sterility of the interspecific hybrids are located in the same interval, in both *O. sativa* subspecies.

Chromosome 6 genotype of viable S_1^s female gametes produced by the F_1 hybrids: Overall, only a small number of female S_1^s gametes produced by the F_1 hybrid are able to survive the allelic elimination exerted by S_1. To characterize these gametes, the complete markers' genotype of the S_1^s/S_1^s plants (each originated from a viable S_1^s female gamete) was examined in detail all along chromosome 6. The 34 S_1^s/S_1^s BC_1F_1 plants from the four populations under study, as well as two individuals from two additional small populations, were analyzed in detail (Table 1). All of the plants shared an *O. sativa* chromosomal segment of at least 7.3 cM (r = 0.0725) around S_1^s, equivalent to 886 kb in the genome of Nipponbare (Table 1). The probability of observing no recombination events between the bounds of this segment in 36 lines is $p_0 = (1 - 0.0725)^{36} = 0.066$. This

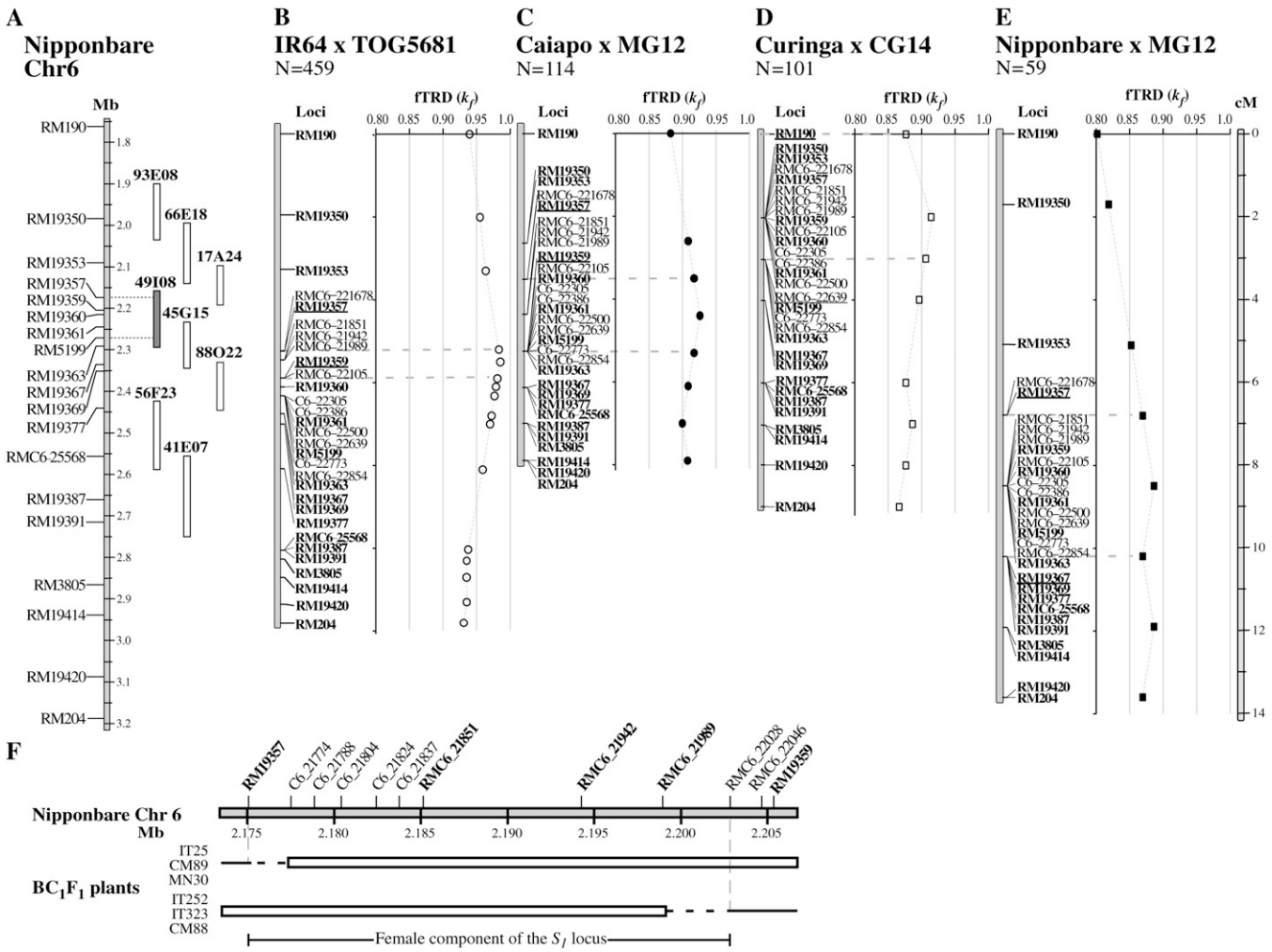


FIGURE 3.—Fine mapping of the female component of the S_1 locus. (A) Physical map of the S_1 region in the *O. glaberrima* cv. CG14. Positions of markers around the S_1 locus in the Nipponbare chromosome 6 and their amplification pattern in the CG14 BAC clones (OG_BBa0093E08, OG_BBa0066E18, OG_BBa0017A24, OG_BBa0049I08, OG_BBa0045G15, OG_BBa0088O22, OG_BBa0056F23. and OG_BBa0041E07) are shown. Saturation with new markers obtained with the OG_BBa0049I08 sequence, in crosses (B) IR64xTOG5681, (C) CAxMG12, (D) CUxCG14. and (E) NIPxMG12. Markers indicated in boldface type are common with Figure 2. Underlined markers limit the fTRD peak for each cross. (F) Recombination breakpoints limiting the fTRD peak. Marker positions in the Nipponbare genome are shown. Open bars represent the heterozygote genomic segments, while *O. sativa* homozygote segments are represented by straight lines. Dotted lines symbolize the recombination sites.

probability is low enough to suggest the necessary inheritance of other *O. sativa* factors in the vicinity of S_1 to ensure the viability of the S_1^s gametes. As S_1 is located 3.0 cM from the left bound (RM19350) and 4.3 cM from the right bound (RM3805) of the common segment, the existence of an additional factor on each side of it (from now on denoted as S_1A and S_1B, respectively) is suspected.

In the cross with the lowest S_1^s transmission (IR64xTOG5681), the 9 S_1^s/S_1^s plants share a much larger *O. sativa* chromosomal segment (with the exception of a heterozygote <7 cM singleton in one line, from RM19414 to RM204). This interval, limited by markers RM19350 and RM3183, starts 3.3 cM before and ends 48.2 cM after S_1. The probability of observing no recombination events between the bounds of this segment in 9 lines $p_0 = (1 - 0.387)^9 = 0.012$, strongly suggests that another factor (S_1C), located between markers RM204 and RM3183, plays an additional role in selecting the gametes, only in this cross. Considered together, these results indicate that two linked factors present in the *O. sativa* genome are necessary for the survival of female gametes carrying S_1^s and that an additional factor is also required in the IR64xTOG5681 cross.

Sequence annotation and comparison of the S_1 locus between *O. glaberrima* and *O. sativa:* The 164,664-bp sequence of the *O. glaberrima* BAC OG-BBa0049I08 was carefully annotated and analyzed. TEs and predicted genes are shown in Table S3 and Table S4 and in Figure S2, respectively.

To study the sequence divergence of the S_1 locus, fine comparative genomic approaches were carried out

TABLE 1

Frequency of S_1^s alleles in the *O. sativa* × *O. glaberrima* BC_1F_1 populations

Populations	Minimal *O. sativa* inherited region around S_1^s	N	$N(S_1^s)$	$f(S_1^s)$	$\sigma f(S_1^s)$	$f \pm \sigma f(S_1^s)$	$f \pm 2\sigma f(S_1^s)$
indica × *O. glaberrima*							
IR64 × TOG5681	RM19350–RM3183	459	9	0.020	0.006	0.013–0.026	0.007–0.033
japonica × *O. glaberrima*	RM190–RM3805	274	25	0.091	0.017	0.073–0.108	0.056–0.126
Caiapo × MG12		114	9	0.079	0.025	0.054–0.104	0.028–0.129
Curinga × CG14		101	9	0.089	0.028	0.061–0.117	0.032–0.146
Nipponbare × MG12		59	7	0.119	0.042	0.077–0.161	0.034–0.203
Additional individuals							
Caiapo × TOG5681		34	1	0.029	0.029	0.000–0.058	0.000–0.087
Nipponbare × TOG5681		12	1	0.083	0.080	0.004–0.163	0.000–0.243
O. sativa × *O. glaberrima*	RM19350–RM3805	779	36	0.046	0.008	0.039–0.054	0.031–0.061

$f(S_1^s)$, observed S_1^s frequency; σ, standard deviation of $f(S_1^s)$; confidence intervals of the S_1^s/S_1^s BC_1F_1 plants in *indica* and *japonica* crosses.

between CG14 and Nipponbare. Nucleotide sequences surrounding the male component of the locus mapped by Koide *et al.* (2008c) (between markers E0506 and E1920) were extracted from both genomes and analyzed in detail (Figure 4). Alignments of the 44.8- and 68.8-kb sequences showed an overall conservation of nucleotide sequences, with breaks that expanded the *O. glaberrima* segment (Figure 4). This 1.5× increase was mainly due to the insertion of 15.3 kb from 19 TEs, suggesting a higher local accumulation in comparison to the whole BAC sequence. Within the *O. glaberrima* 50.38-kb interval of the female component of the S_1 locus, seven genes and three pseudogenes were predicted (Table 2). Among the predicted genes, two are hypothetical proteins (49I08.7 and 49I08.75), three exhibited similarities with "early nodulins," one exhibited with ribosome biogenesis regulatory proteins (49I08.8), and one exhibited with F-box proteins (49I08.11). The three pseudogenes and the two hypothetical proteins were located between predicted genes 49I08.6 and 49I08.11; three of them (49I08.75, 49I08.9, and 49I08.10) corresponded to nonautonomous chimeric Mutator-like transposable elements (pack-MULEs) carrying fragments of cellular genes (Jiang *et al.* 2004).

The gene content in the segment was found to be variable between the two species. Neither the pseudogenes nor the hypothetical proteins were found at the collinear positions in *O. sativa* (Table 2 and Figure 4). In contrast, hypothetical gene Os06g04960 from *O. sativa* was not found to be collinear with *O. glaberrima*. Reannotation of the *O. sativa* segment showed that this gene overlapped different predicted TEs and that it was split away in *O. glaberrima* by the insertion of additional TEs and the predicted genes 49I08.7 and 49I08.75. All five collinear protein-coding genes showed a high conservation in gene structure, as introns–exons were strictly conserved. The only exception occurred in the first intron of gene 49I08.11, with the insertion of a 448-bp nonautononous transposon. All of the orthologous genes within the segment were found to be highly conserved at the nucleotide and protein levels, with sequence identities ranging from 96 to 100% and from 82.8 to 100%, respectively.

The K_a/K_s rates were calculated for 17 orthologous predicted genes present in the CG14 BAC and Nipponbare sequences (Table 2 and Table S4); the other genes were not analyzed due to the absence of one of the orthologous genes, due to deep gene structure differences, or because one of the two was annotated as a pseudogene. Most of the analyzed genes appeared to be under purifying selection, with the exception of four genes outside the S_1 interval (49I08.1, 49I08.19, 49I08.21, and 49I08.22) whose high K_a/K_s values seem to indicate a strong positive selection or a relaxed selective constraint. Despite an overall purifying evolution of genes in the sequenced BAC, four genes show high nonsynonymous changes. Two of them are located within the S_1 interval (49I08.8 and 49I08.11).

DISCUSSION

TRD and its association with interspecific sterility loci: Different processes such as the nonrandom segregation of chromosomes during meiosis (meiotic drive) (Fishman and Willis 2005), abortion of haploid gametes (as in hybrid sterility) (Orr and Irving 2005), or abortion of zygotes after fertilization (as in hybrid inviability) (Matsubara *et al.* 2003) could deviate the final allelic frequency in a population from the expected Mendelian ratio. The strong association between markers showing TRD and pollen sterility QTL has been already demonstrated in an interspecific cross of tomato (Moyle and Graham 2006). This association suggests that hybrid sterility may arise from incompatibilities

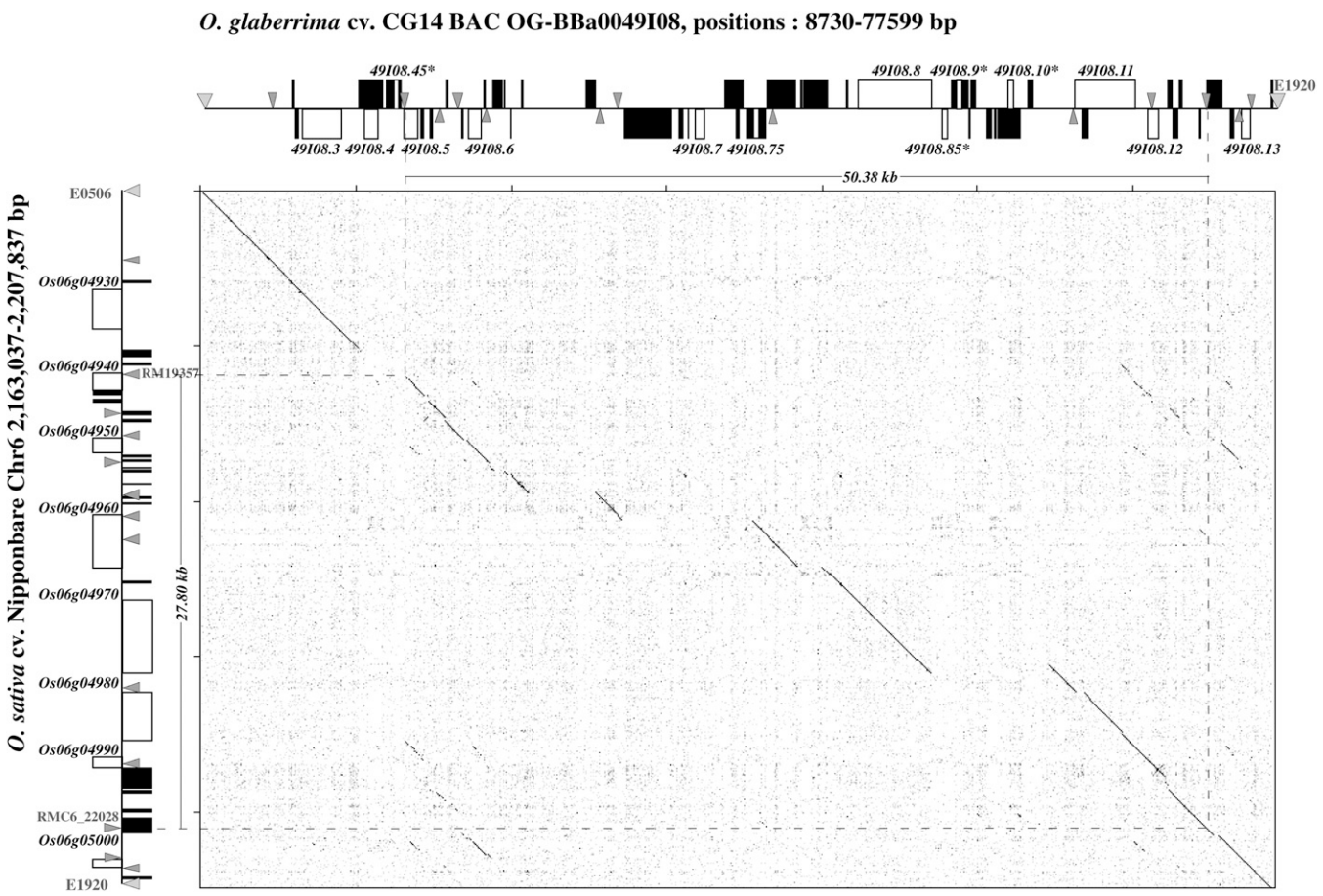


Figure 4.—Graphical comparison of the *O. sativa* and *O. glaberrima* sequences in the S_1 locus. Pairwise alignments of the respective 44.8- and 68.8-kb sequences of *O. sativa* and *O. glaberrima* in the S_1 locus, limited by markers E0506 and E1920 (Koide *et al.* 2008c). Predicted genes (open boxes), TEs (solid bars), and molecular markers (shaded triangles) are shown for each sequence. Dotted lines bound the localization of the factor affecting female fertility (between markers RM19357 and RMC6_22028).

between different factors that play a role in gamete development, resulting in the elimination of specific allelic combinations.

In our genetic map, the interpolated positioning of genes and QTL causing *O. sativa* × *O. glaberrima* hybrid sterility evidences their colocalization with markers showing fTRD (Figure 1). Differences in the intensity and the direction of the fTRD were found for chromosomes 1, 2, 3, 4, 7, and 12, in comparison with the genetic map obtained by J. Li *et al.* (2008), and for chromosomes 4, 5, and 11, in comparison with the genetic map obtained by Doi *et al.* (1998b). Some of those differences could be explained by a rapid evolution of the sterility barriers, generating different alleles of the incompatibility factors within species (Sweigart *et al.* 2007; Widmer *et al.* 2009). Colocalization of fTRD sites with pollen sterility loci could imply that these loci also have a role in female gamete development or in postpollination, since fTRD cannot be caused by loci affecting only male gametogenesis. Occurrence of fTRD in genomic regions where no sterility loci have been described (chromosomes 2 and 6) could be explained either by the existence of a new sterility locus segregating in this cross or by some mechanism other than gamete elimination biasing the segregation. The TRD in the long arm of chromosome 6, which has also been described for another *O. sativa* × *O. glaberrima* cross (J. Li *et al.* 2008), involves different markers in the two *O. sativa* subspecies (Figure 2), indicating that different factors may cause it. This observation would suggest that the regions near and below the chromosome 6 centromere might contain polymorphic factors within the species, possibly involved in the viability of gametes, zygotes, or endosperm in the *O. sativa* × *O. glaberrima* hybrids. Interestingly, the same genomic regions have been reported to bear loci for endosperm abortion, in hybrids between *O. rufipogon* and *O. sativa* (Matsubara *et al.* 2003; Koide *et al.* 2008b), and of female and male gametes (Koide *et al.* 2008a). It is plausible that a related mechanism, or different alleles of the same loci, work in hybrids of the two cultivated rice species, as factors involved in reproductive barriers between different species can be closely linked or can be allelic (Hu *et al.* 2006).

Use of the fTRD as a method to map sterility loci: Hybrid sterility is a common phenomenon in rice.

TABLE 2

Coding sequences, annotation, and comparison between orthologous *O. glaberrima* cv. CG14 and *O. sativa* cv. Nipponbare S_1 regions

O. glaberrima gene name	Best BLASTN homology	Best BLASTN EST homology	Best BLASTX homology (swiss prot)	Protein domain	Putative function	*Putative O. sativa* (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	K_a (d_N)	K_s (d_S)	K_a/K_s
OG-BBa0049I08.5	AK122162 *O. sativa* flcDNA (2*e*-148)	EL586675 *O. sativa* (2e-162)	Q02921 *Glycine max* Early nodulin 93 (8*e*-20)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04940	65	54.1	59.5	—	—	—	—
OG-BBa0049I08.6	AK121791 *O. sativa* flcDNA (1*e*-157)	CI240527 *O. sativa* (1e-158)	Q02921 *G. max* Early nodulin 93 (4*e*-14)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04950	89.1	76.8	78.4	376	—	—	—
OG-BBa0049I08.7	—	—	—	—	Hypothetical protein	No orthologous gene	—	—	—	—	—	—	—
OG-BBa0049I08.75	—	—	—	—	Hypothetical protein	No orthologous gene	—	—	—	—	—	—	—
OG-BBa0049I08.8	NM_001063299 *O. sativa* Os06g0142000 (0.0)	CB622551 *O. sativa* (0.0)	Q9SH88 *A. thaliana* Ribosome biogenesis regulatory protein (3*e*-48)	—	Putative protein	LOC_Os06g04970	98.6	97.7	98.4	2073	0.0112	0.0237	0.4728
OG-BBa0049I08.85[a]	AP002838 *O. sativa* BAC (2*e*-156)	CT858525 *O. sativa* (8e-67)	—	—	Hypothetical protein	No orthologous gene	—	—	—	—	—	—	—
OG-BBa0049I08.9[a]	AP004090 *O. sativa* BAC (8*e*-69)	CT844843 *O. sativa* (3e-68)	—	—	Hypothetical protein	No orthologous gene	—	—	—	—	—	—	—
OG-BBa0049I08.10[a]	AC121363 *O. sativa* BAC (3*e*-128)	CB649669 *O. sativa* (1*e*-57)	P47927 *A. thaliana* Floral homeotic protein APETALA 2 (4*e*-8)	—	Hypothetical protein	No orthologous gene	—	—	—	—	—	—	—
OG-BBa0049I08.11	AK073027 *O. sativa* flcDNA (0.0)	CB658549 *O. sativa* (0.0)	Q9FJT2 *A. thaliana* F-box/FBD/ LRR-repeat (5*e*-4)	—	Putative F-box protein	LOC_Os06g04980	98.1	96.5	97.7	1542	0.0173	0.0198	0.8715
OG-BBa0049I08.12	AK122162 *O. sativa* flcDNA (2*e*-175)	CI252819 *O. sativa* (2*e*-175)	Q02921 *G. max* Early nodulin 93 (1*e*-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04990	99.7	100	100	348	0.0000	0.0142	0.001

The *e*-value for the BLAST is presented in brackets. Predicted genes located in the S_1 locus, between markers RM19357 and RMC6_22028.

[a] Pseudogene.

However, only a few factors responsible for inter- or intraspecific sterility have been fine mapped (W. Li *et al.* 2006, 2008; D. Li *et al.* 2007; Zhao *et al.* 2007; Koide *et al.* 2008c), and only two have been recently identified (Chen *et al.* 2008; Long *et al.* 2008). Indeed, the mapping strategy that has been used requires isolating the genetic region into near isogenic lines and carrying out the production and phenotypic analysis of many recombinant plants.

To reduce the time and effort to map sterility loci, we developed a simpler approach that makes it possible to quickly fine map these genes on the basis of the effect that they have on their own segregation ratio (Phadnis and Orr 2009). The approach is based on the direct measurement of the TRD for each marker evaluated on a sequentially saturated map. Due to the hitchhiking effect with the distorter locus (Chevin and Hospital 2006), the variations on marker TRD are inversely proportional to the genetic distance between them and the factor causing the distortion. This leads to a maximal TRD in the genomic region containing the TRDL, which, in the case of S_1, is the factor causing the allelic elimination of gametes.

Different algorithms have been previously developed to predict the genetic positions of TRDL (Lorieux *et al.* 1995a,b; Vogl and Xu 2000; Zhu *et al.* 2007) and their utilization allowed to find the approximate location of several TRDL in a whole genome scale, showing a good colocalization of distorted markers with the predicted locations for the TRDL (Bouck *et al.* 2005; Hall and Willis 2005). However, they sometimes fail to detect the presence of TRDL in highly distorted locations (Fishman *et al.* 2001). Our method differs from these previous methodologies in that the use of sequential saturations achieving a locally highly saturated map (in our case close to one marker per gene) makes the estimation of the TRDL position unnecessary, since the highest TRD is a direct evidence of the cosegregation with the distorter.

Additionally, this methodology has several advantages in comparison to the traditional mapping strategies of sterility loci, which typically involve large, advanced backcross populations evaluated for pollen and spikelet fertility. First, it is faster, as it is based on the analysis of BC_1F_1 or F_2 populations. Second, it is accurate as the TRD could be caused by the gamete eliminator factor itself (Phadnis and Orr 2009). Third, the approach could be used to fine map not only hybrid sterility loci causing TRD, but also any TRDL located in a genomic region with a good recombination ratio, as long as enough markers can be designed in the species under study. However, its application is limited to cases where highly significant differences in the level of TRD are observed.

The complex nature of the S_1 locus: The sterility barrier due to the S_1 locus has been described to involve two separate components influencing male and female gametogenesis and unlinked modifiers that enhance fTRD on *O. sativa* ssp. *indica* (Koide *et al.* 2008c). Using our mapping methodology in four BC_1F_1 populations, a region of 27.8 kb in *O. sativa* cv. Nipponbare, and another one of 50.3 kb in *O. glaberrima* cv. CG14 were identified as the location of the female component of the S_1 locus. Surprisingly, in *indica* and *japonica* this component maps exactly within the interval where the male component was mapped (Koide *et al.* 2008c). This colocalization would indicate that a unique or two contiguous factors could be involved in the elimination of male and female S_1^s gametes in both *O. sativa* backgrounds. In the case of a single factor affecting both male and female gametogenesis, its expression or function may be different, or it may interact with additional gene products in female gametogenesis, to cause total elimination of S_1^s from pollen, and partial allelic elimination in megagametophytes.

On the basis of the genotype of the surviving S_1^s female gametes, the presence of two linked elements affecting their viability (S_1A and S_1B) was detected, and evidence for an additional factor (S_1C) was also found in one cross. Because S_1 alone is capable of inducing mTRD but not fTRD (Koide *et al.* 2008c), it is therefore possible that an epistatic interaction between the conspecific alleles of S_1, S_1A, and S_1B might be the main cause of fTRD and female sterility in *O. sativa* × *O. glaberrima* hybrids. In this scenario, the additional S_1C factor would have a supplementary deleterious effect over the S_1^s gametes, strengthening the interspecific sterility barrier in certain crosses by reducing the viability of recombinant gametes (Orr and Turelli 2001). The higher S_1^s elimination found in crosses involving TOG5681 with *indica* and *japonica* accessions (Table 1) may suggest that the elevated fTRD value does not depend exclusively on the *indica* or *japonica* background as previously suggested, but also on the presence of the S_1C incompatibility found between certain *O. glaberrima* and *O. sativa* accessions. Nevertheless, this concept needs confirmation because of the low number of plants issued from these additional crosses.

The hypothesis that a conspecific epistatic interaction between linked factors is necessary and responsible for the hybrids' female sterility and fTRD explains quite well previous observations describing how the loss of heterozygosity by recombination on the upstream, downstream, or both *O. glaberrima* segments near S_1 cause the recovery of female fertility and the elimination of fTRD on S_1^s/S_1^g plants (Sano 1990; Koide *et al.* 2008c). Furthermore, it has already been shown that a similar, complex conspecific epistasis causes sterility and TRD in *Drosophila pseudoobscura* intersubspecific hybrids, where sterility was observed only in males possessing an entire X-chromosome block derived from one of the parental subspecies (*bogotana*), and not in the ones carrying independently each one of the involved loci (Orr and Irving 2005).

On the basis of our data, the location for S_1A and S_1B can be predicted between markers RM19350 and S_1 (210 kb segment in Nipponbare) and between S_1 and RM3805 (654 kb), respectively. As for S_1C, it might be located within the 9.2-Mb segment between RM204 and RM3183. Interestingly, the intersubspecific sterility locus S_5 is located within this large region (Chen *et al.* 2008).

Predicted model for the reproductive sterility barrier between *O. sativa* and *O. glaberrima* mediated by S_1: The origin of postzygotic barriers has been explained by the Bateson–Dobzhansky–Muller (BDM) model as the accumulation of incompatibilities in epistatic interactions between genes (Bateson 1909; Dobzhansky 1936; Muller 1942). As previously mentioned, a conspecific epistasis between S_1A, S_1, and S_1B may be the cause of the female hybrid sterility between the two cultivated rice species. It is then plausible that BDM incompatibilities between S_1A^g and S_1B^g with S_1^s may play a part in the allelic elimination, as the F_1 plants produce viable S_1A^s–S_1^g or S_1^g–S_1B^s, but no S_1A^g–S_1^s or S_1^s–S_1B^g gametes.

In light of the available data, we considered several possible models where BDM incompatibilities during megasporogenesis and megagametogenesis are the cause of the female sterility barrier between *O. sativa* and *O. glaberrima* (see File S2, Genetic models.xlsx). The different models are all based on the following data and information: (1) the localization of S_1 found with the fTRD analysis, (2) the systematic transmission of an *O. sativa* gene block around S_1^s that suggests the presence of at least one additional locus on each side of S_1, (3) the maximal recombination rates within and around the S_1A–S_1B interval [S_1A to S_1 (r_1), S_1 to S_1B (r_2), and S_1B to the centromere (r_5)], (4) the loss of fTRD and of female sterility after recombination around S_1 (Sano 1990; Koide *et al.* 2008c), and finally, (5) the cytological observations during female gametogenesis of the hybrids, showing two types of aborted female gametes: the ones characterized by the absence of embryo sacs, probably due to an abortion during or just after meiosis, and the ones where embryo sacs have less than seven cells (Bouharmont *et al.* 1985; Koide *et al.* 2008c).

After consideration of the different models, we retained model 1 (see explanation of the different models in File S2, Genetic models.xlsx), as it is the only one that predicts well the three following parameters: (1) The S_1^s observed frequency [$f(S_1^s)$], (2) the total rate of gamete survival (D) observed as the final seed set of the hybrids, and (3) the observed recombination rates between S_1A–S_1 (r_1) and S_1–S_1B (r_2). The proposed model states that in the presence of S_1A^g–S_1^g–S_1B^g only sister cells ($1n$-$2c$ or $1n$-$1c$) that carry at least one copy of S_1^g would be viable after each meiotic division (Figure 5). When recombination exchanges S_1A^g or S_1B^g, the epistatic interaction acting against S_1^s would cease, allowing the development of the corresponding S_1^s daughter cells. In consequence, the only S_1^s meiocytes that could survive are the ones whose sister cell bears the S_1A^s–S_1^g–S_1B^g, S_1A^g–S_1^g–S_1B^s or S_1A^s–S_1^g–S_1B^s recombinant haplotypes. Nevertheless, at the end of megagametogenesis, only S_1^s megaspores carrying the S_1A^s–S_1^s–S_1B^s configuration would lead to the formation of a functional embryo sac.

The developed model enables calculation of the expected frequencies of $f(S_1^s)$, D, r_1, and r_2, on the basis of (1) the true recombination rates within the S_1A–S_1 (r_1) and S_1–S_1B (r_2) segments and (2) the genetic distance between S_1B and the centromere (r_5). After simplification (see File S3, Model 1 equations), the D, $f(S_1^s)$, r_1, and r_2 values predicted by the model can be expressed as

$$D = (1-2r_5)\left(\frac{r_2+1}{2}\right) + 2r_5\left[\frac{1}{2} + \frac{r_1}{2} - r_1 r_2\right],$$

$$f(S_1^s) = \left[\frac{r_2}{2} + r_5(r_1 - r_2 - 2r_1 r_2)\right] \Big/ D,$$

$$\hat{r_1} = \frac{r_1 \max}{2D}; \quad \hat{r_2} = \frac{r_2 \max}{2D}.$$

For the cross IR64xTOG5681, the presence of the additional factor S_1C must be taken into account when predicting the $f(S_1^s)$. When applying model 1 with the recombination rates observed in the cross ($r_{1\max} = 0.033$ and $r_{2\max} = 0.049$), and with an additional reduction due to the abortion of the S_1A^s–S_1^s–S_1B^s–S_1C^g gametes (with a $r_3 = 0.349$ representing the maximal recombination ratio between S_1B and S_1C), an estimated $f(S_1^s)$ of 2.23%, and D of 51.1% are predicted (File S2, Genetic models.xlsx, sheet 2). This calculated frequency fits very well within the significance interval for the cross (Table 1). Consequently, the formulas for D and $f(S_1^s)$ applicable to crosses where the effect of the S_1C locus is present convert to

$$D' = D/(1 - f(S_1^s))r_3,$$

$$f'(S_1^s) = (1 - r_3)f(S_1^s)/(1 - f(S_1^s)r_3),$$

and the estimates for r_1 and r_2 convert to

$$\hat{r_1} = \frac{r_1 \max}{2D'}; \quad \hat{r_2} = \frac{r_2 \max}{2D'}.$$

Although the predictions made by Model 1 are the closest to the observed parameters, the predicted values of $f(S_1^s)$ are slightly lower than the ones observed for the three *japonica* × *glaberrima* crosses. The estimated $f(S_1^s)$ (around 3%) lies at the inferior bound of the observed values. The S_1A^s–S_1^s–S_1B^s gametes seem to show some additional viability, meaning that the effect of the S_1 region could show some level of incomplete penetrance. This enhanced viability rate is still difficult to

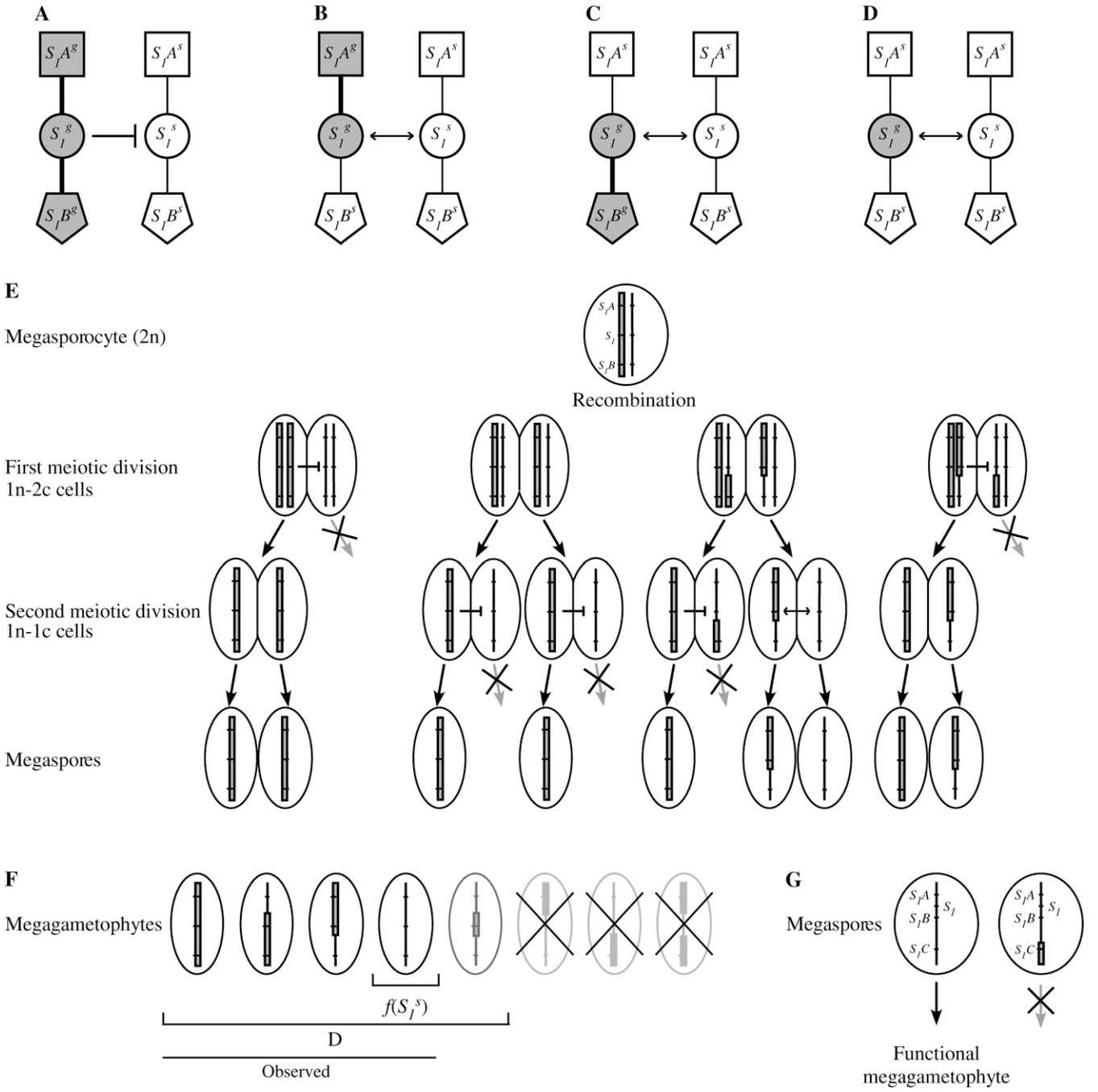


FIGURE 5.—Predicted model for the allelic gamete elimination involving S_1. Genetic model based on a BDM incompatibility that explains the allelic elimination of female gametes involving the S_1 locus. (A) In cells that carry S_1A^g, S_1^g, and S_1B^g (shaded) together with S_1A^s, S_1^s, and S_1B^s alleles (open), the epistatic interaction between the three *O. glaberrima* factors (thick lines) will cause the abortion (blunt-end arrow) of daughter cells carrying S_1^s. This allelic abortion will thus originate a fTRD. When recombination eliminates S_1B^g (B), S_1A^g (C), or both (D), the epistatic incompatible interaction acting against S_1^s would not be present (double-head arrow), allowing the development of S_1^s gametes and eliminating the fTRD. (E) Model explaining how recombination events between S_1A and S_1B, and the segregation of recombinant chromatids determine the fate of the cells produced during the two meiotic divisions of a heterozygote megasporocyte. In the presence of S_1A^g, S_1^g, and S_1B^g, only cells that possess at least one copy of S_1^g will continue their development after each meiotic division. When recombination between S_1A and S_1B occurs, the epistasis against S_1^s would not be present in all of the daughter cells produced during meiosis ($1n$-$2c$ or $1n$-$1c$), allowing the survival of some of the S_1^s bearing megaspores; however, only the nonrecombinant S_1A^s–S_1^s–S_1B^s are able to continue their development. As a result of the allelic incompatibilities, only the chalazal-most megaspores that originate from a viable configuration during meiosis would be capable of producing a functional embryonary sac. (F) Viable and aborted female gametes produced by heterozygotes for the S_1A–S_1B region, as predicted by the developed model. Configurations accounting for the gamete survival rate (D) and the frequency of *O. sativa* S_1 alleles ($f(S_1^s)$) are shown. All of the viable configurations, with the exception of the double recombinants, were observed in the genotypes of the 779 BC_1F_1 plants evaluated. (G) Female gamete elimination based on the presence of an additional factor (S_1C), which causes the abortion of the S_1A^s–S_1^s–S_1B^s–S_1C^g gametes. Shaded bars and straight lines represent the *O. glaberrima* and *O. sativa* chromatides, respectively.

explain considering the available data and could be due to: (1) unknown nonlinked factors (that we were not able to detect), which rescue some of the S_1^s gametes, and are segregating simultaneously with the three epistatic loci, or (2) intraspecific polymorphism on the epistatic factors, as seen in other postzygotic barriers (Sweigart *et al.* 2007), likely to cause differences in the levels of the BDM incompatibility. The existence of a sterility locus (S_{10}) observed in crosses between *indica* and *japonica* at the same location of S_1 (Sano *et al.* 1994; Zhu *et al.* 2005) favors the hypothesis of an allelic variation. The possible differences that might exist between the *indica* and *japonica* alleles for the S_1A, S_1 and S_1B loci cause a certain level of sterility between the two *O. sativa* subspecies and may alter the allelic elimination of female gametes produced by the *O. sativa* × *O. glaberrima* hybrids.

In summary, the predictions made by our model support the hypothesis that female sterility in the *O. sativa* × *O. glaberrima* hybrids is caused by the allelic elimination of gametes, due to a BDM incompatibility in an epistatic interaction between S_1 and the additional factors. Differences in the observed fTRD between *indica* and *japonica* could be due to polymorphisms within the two subspecies in S_1A, S_1, S_1B, or S_1C, or to the existence of other unlinked factors that rescue some of the gametes. Accordingly, the final allelic frequencies for a given cross would be determined by the alleles confronted in a given cross, by the recombination ratio between S_1A and S_1B, and by the presence of the S_1C-mediated incompatibility.

Genomic variations in the S_1 orthologous region and candidate genes: Divergent evolution between closely related species may be the cause of reproductive isolation. In Drosophila, a high level of gene divergence and strong positive selection appeared to be mechanisms influencing the evolution of speciation genes (Ting *et al.* 1998; Presgraves *et al.* 2003; Barbash *et al.* 2004; Brideau *et al.* 2006). Recent studies conducted with different species of the Oryza genus indicated an extensive conservation of the microcollinearity, with local small rearrangements, or structural variations due to insertions of transposable elements (Ammiraju *et al.* 2008; Lu *et al.* 2009). In contrast, the accumulation of structural genomic variations observed within the interval of the S_1 locus in *O. glaberrima* might imply a fast evolution of the segment, despite the recent divergence of the Asian and African rice species (approximately 0.7 MYA; Ma and Bennetzen 2004; Ammiraju *et al.* 2008).

We used orthologous comparative genomic approaches between *O. glaberrima* and *O. sativa* at the S_1 interval to explore structural changes and nucleotide divergence of orthologous coding regions as putative candidates for the locus. Ten genes are present in the 50.3-kb S_1 region of *O. glaberrima* and thus constitute the candidates for the S_1 locus. From these 10, 2 predicted genes attract our attention on the basis of their putative functions, differential presence on the species, or high sequence divergence. The first one, carried by a pack-MULE and present only in the *O. glaberrima* S_1 region, is gene 49I08.10. This gene shows similarities with an APETALA2 (AP2) transcription factor. Mutations in AP2 and other related proteins in Arabidopsis cause the arrest of mega-sporogenesis after the first meiotic division (Byzova *et al.* 1999). Pack-MULEs are known to have an important role in rice genome evolution, as they capture and relocate gene fragments in other genomic contexts (Jiang *et al.* 2004). A significant number of these elements are transcribed, being frequently associated with small RNAs that may modulate both pack-MULEs and paralogous gene expression (Hanada *et al.* 2009). If the complete paralogous AP2 gene has a function similar to that of the Arabidopsis genes, this pack-MULE could affect its expression on the heterozygotes, in a dose-dependent mechanism similar to hybrid dysgenesis (Michalak 2009), causing the abortion of female gametes.

The other interesting candidate is the one with the highest *K*a and K_a/K_s values in the S_1 interval (Table 2), gene 49I08.11, belonging to the super-family of F-box proteins. Members of this family constitute protein complexes known as SCF (Skp1–Cullin–F-box) involved in the control of a wide range of processes (Xu *et al.* 2009). Several F-box genes and their associated proteins have been related with the progression of the cell cycle, especially during sporogenesis and gametogenesis (Wang and Yang 2006; Pesin and Orr-Weaver 2008; Gusti *et al.* 2009). Intraspecific sequence divergence of members of such protein complexes can cause hybrid sterility, as demonstrated for the complex composed by an F-box and a SUMO E3 ligase-like protein, that controls the intersubspecific hybrid sterility mediated by the *Sa* locus in *O. sativa* (Long *et al.* 2008). Additionally, homologous F-box genes closely related to our putative candidate are involved in the self-incompatibility system of *Hordeum bulbosum* (Kakeda 2009), proving that such genes can actually act as reproductive barriers in cereals. From the two putative candidate genes described, this F-box protein is the one that fits best with the genetic model of a BDM incompatibility in an epistatic interaction developed in this article to explain the female gamete elimination caused by S_1. Taking into account their ability to tightly interact with other proteins, the rising of incompatibilities after divergences in the coding sequence of the interacting proteins seems to be a feasible mechanism for this postzygotic reproductive barrier. Moreover, considering that the strong hybrid sterility between *O. sativa* and *O. glaberrima* occurs independently of the direction of the interspecific cross (Sano *et al.* 1979; Bougerol and Pham 1989; Sano 1990), while hybrid dysgenesis, the possible mechanism explaining the role of gene 49I08.10 in the hybrid dysfunction, is a unidirectional phenomenon, the F-box protein appears as a more consistent candidate.

Perspectives: In this article, we have described and modeled the complex nature of the allelic female gamete elimination mediated by the S_1 locus. The knowledge generated by our results will contribute to a better understanding of plant interspecific sterility barriers and will also open the door to an efficient use of *O. glaberrima* in rice breeding programs, as is intended with the iBridges project. Hybrid plants inheriting the *O. sativa* chromosomal segment that contains the S_1, S_1A, and S_1B factors involved in the allelic elimination can be used to develop lines with all the potential genes of interest from *O. glaberrima* that would produce a more fertile progeny when crossed with *O. sativa*. More information is needed to characterize all the factors involved, as well as the molecular mechanism behind the allelic gamete elimination in hybrids from Asian and African rice species.

Our thanks go to R. Wing (AGI, Arizona) for providing the *O. glaberrima* BAC clones and J. L. Goicoechea for the *in silico* confirmation of the MTP; M. Morales for the *in vitro* germination of seeds and tissue harvest; and M. F. Alvarez and L. Melgarejo for their contribution in the construction of the IR64xTOG5681 genetic map. We also thank two anonymous reviewers for their constructive comments that helped us to improve significantly the manuscript.This research was supported by the Generation Challenge Program (Grant G4007.01). A grant from Département Soutien et Formation (DSF) Institut de Recherche pour le Développement (IRD) and International Center for Tropical Agriculture (CIAT) provided the Ph.D. scholarship for A. Garavito.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Aluko, G., C. Martinez, J. Tohme, C. Castano, C. Bergman *et al.*, 2004 QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* x *O. glaberrima*. Theor. Appl. Genet. **109:** 630–639.

Ammiraju, J. S. S., F. Lu, A. Sanyal, Y. Yu, X. Song *et al.*, 2008 Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell **20:** 3191–3209.

Barbash, D. A., P. Awadalla and A. M. Tarone, 2004 Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. PLoS Biol. **2:** e142.

Bateson, W., 1909 Heredity and variation in modern lights, pp. 81–101 in *Darwin and Modern Science*, edited by A. C. Seward. Cambridge University Press, Cambridge.

Bikard, D., D. Patel, C. Le Mette, V. Giorgi, C. Camilleri *et al.*, 2009 Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science **323:** 623–626.

Bomblies, K., J. Lempe, P. Epple, N. Warthmann, C. Lanz *et al.*, 2007 Autoimmune response as a mechanism for a Dobzhansky–Muller-type incompatibility syndrome in plants. PLoS Biol. **5:** e236.

Bouck, A., R. Peeler, M. L. Arnold and S. R. Wessler, 2005 Genetic mapping of species boundaries in Louisiana Irises using IRRE retrotransposon display markers. Genetics **171:** 1289–1303.

Bougerol, B., and J. L. Pham, 1989 Influence of the *Oryza sativa* genotype on the fertility and quantitative traits of F1 hybrids between the two cultivated Rice species *Oryza sativa* L and *Oryza glaberrima* Steud. Genome **32:** 810–815.

Bouharmont, J., M. Olivier and M. D. Chassart, 1985 Cytological observations in some hybrids between the rice species *Oryza sativa* L. and *O. glaberrima* Steud. Euphytica **34:** 75–81.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshwari, X. Wang *et al.*, 2006 Two Dobzhansky–Muller genes interact to cause hybrid lethality in *Drosophila*. Science **314:** 1292–1295.

Byzova, M. V., J. Franken, M. G. M. Aarts, J. de Almeida-Engler, G. Engler *et al.*, 1999 *Arabidopsis* STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. Genes Dev. **13:** 1002–1014.

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

Chevin, L. M., and F. Hospital, 2006 The hitchhiking effect of an autosomal meiotic drive gene. Genetics **173:** 1829–1832.

Dingkuhn, M., M. P. Jones, D. E. Johnson and A. Sow, 1998 Growth and yield potential of *Oryza sativa* and *O. glaberrima* upland rice cultivars and their interspecific progenies. Field Crops Res. **57:** 57–69.

Dobzhansky, T., 1936 Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. Genetics **21:** 113–135.

Doi, K., K. Taguchi and A. Yoshimura, 1998a A new locus affecting high F_1 pollen sterility found in backcross progenies of Japonica rice and African rice. Rice Genet. Newsl. **15:** 146–148.

Doi, K., A. Yoshimura and N. Iwata, 1998b RFLP mapping and QTL analysis of heading date and pollen sterility using backcross populations between *Oryza sativa* L. and *Oryza glaberrima* Steud. Breeding Sci. **48:** 395–399.

Doi, K., K. Taguchi and A. Yoshimura, 1999 RFLP mapping of *S20* and *S21* for F_1 pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. Rice Genet. Newsl. **16:** 65–68.

Droc, G., C. Perin, S. Fromentin and P. Larmande, 2009 OryGenesDB 2008 update: database interoperability for functional genomics of rice. Nucleic Acids Res. **37:** D992–D995.

Fishman, L., and J. H. Willis, 2005 A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (Monkeyflower) hybrids. Genetics **169:** 347–353.

Fishman, L., A. J. Kelly, E. Morgan and J. H. Willis, 2001 A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics **159:** 1701–1716.

Ghesquière, A., J. Sequier, G. Second and M. Lorieux, 1997 First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a 'contig line' concept. Euphytica **96:** 31–39.

Gusti, A., N. Baumberger, M. Nowack, S. Pusch, H. Eisler *et al.*, 2009 The *Arabidopsis thaliana* F-Box protein FBL17 is essential for progression through the second mitosis during pollen development. PLoS One **4:** e4780.

Hall, M. C., and J. H. Willis, 2005 Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence. Genetics **170:** 375–386.

Hanada, K., V. Vallejo, K. Nobuta, R. K. Slotkin, D. Lisch *et al.*, 2009 The functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. Plant Cell **21:** 25–38.

Heuer, S., and K. M. Miezan, 2003 Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. Theor. Appl. Genet. **107:** 902–909.

Hu, F. Y., P. Xu, X. N. Deng, J. W. Zhou, J. Li *et al.*, 2006 Molecular mapping of a pollen killer gene *S29(t)* in *Oryza glaberrima* and colinear analysis with *S22* in *O. glumaepatula*. Euphytica **151:** 273–278.

International Rice Genome Sequencing Project, 2005 The map-based sequence of the rice genome. Nature **436:** 793–800.

Jiang, N., Z. Bao, X. Zhang, S. R. Eddy and S. R. Wessler, 2004 Pack-MULE transposable elements mediate gene evolution in plants. Nature **431:** 569–573.

Josefsson, C., B. Dilkes and L. Comai, 2006 Parent-dependent loss of gene silencing during interspecies hybridization. Curr. Biol. **16:** 1322–1328.

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany *et al.*, 2005 Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. **110:** 462–467.

Kakeda, K., 2009 S locus-linked F-box genes expressed in anthers of *Hordeum bulbosum*. Plant Cell Reports **28:** 1453–1460.

Kikuchi, S., K. Satoh, T. Nagata, N. Kawagashira, K. Doi *et al.*, 2003 Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica*. Rice Sci. **301:** 376–379.

Kim, H., B. Hurwitz, Y. Yu, K. Collura, N. Gill *et al.*, 2008 Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*. Genome Biol. **9:** R45.

Koide, Y., M. Ikenaga, N. Sawamura, D. Nishimoto, K. Matsubara *et al.*, 2008a The evolution of sex-independent transmission ratio distortion involving multiple allelic interactions at a single locus in rice. Genetics **180:** 409–420.

Koide, Y., M. Ikenaga, Y. Shinya, K. Matsubara and Y. Sano, 2008b Two loosely linked genes controlling the female specificity for cross-incompatibility in rice. Euphytica **164:** 753–760.

Koide, Y., K. Onishi, D. Nishimoto, A. R. Baruah, A. Kanazawa *et al.*, 2008c Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species. New Phytol. **179:** 888–900.

Kosambi, D., 1944 The estimation of map distances from recombination values. Ann. Eugenics **12:** 172–175.

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

Li, D., L. Chen, L. Jiang, S. Zhu, Z. Zhao *et al.*, 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., P. Xu, X. Deng, J. Zhou, F. Hu *et al.*, 2008 Identification of four genes for stable hybrid sterility and an epistatic QTL from a cross between *Oryza sativa* and *Oryza glaberrima*. Euphytica **164:** 699–708.

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

Li, W., R. Zeng, Z. Zhang, X. Ding and G. Zhang, 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.

Liang, C., P. Jaiswal, C. Hebbard, S. Avraham, E. S. Buckler *et al.*, 2008 Gramene: a growing plant comparative genomics resource. Nucleic Acids Res.. **36:** D947–D953.

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

Lorieux, M., B. Goffinet, X. Perrier, D. G. León and C. Lanaud, 1995a Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. Theor. and Appl. Genet. **90:** 73–80.

Lorieux, M., X. Perrier, B. Goffinet, C. Lanaud and D. G. León, 1995b Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. Theor. Appl. Genet. **90:** 81–89.

Lorieux, M., M. N. Ndjiondjop and A. Ghesquiere, 2000 A first interspecific *Oryza sativa* x *Oryza glaberrima* microsatellite-based genetic linkage map. Theor. Appl. Genet. **100:** 593–601.

Lorieux, M., G. Reversat, S. X. Garcia Diaz, C. Denance, N. Jouvenet *et al.*, 2003 Linkage mapping of *Hsa-1(Og)*, a resistance gene of African rice to the cyst nematode, *Heterodera sacchari*. Theor. Appl. Genet. **107:** 691–696.

Lu, F., J. S. S. Ammiraju, A. Sanyal, S. Zhang, R. Song *et al.*, 2009 Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. Proc. Natl. Acad. Sci. USA **106:** 2071–2076.

Ma, J., and J. L. Bennetzen, 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA **101:** 12404–12410.

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

Matsubara, K., T. Khin and Y. Sano, 2003 A gene block causing cross-incompatibility hidden in wild and cultivated rice. Genetics **165:** 343–352.

Michalak, P., 2009 Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. Heredity **102:** 45–50.

Mihola, O., Z. Trachtulec, C. Vlcek, J. C. Schimenti and J. Forejt, 2009 A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science **323:** 373–375.

Moyle, L. C., and E. B. Graham, 2006 Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. Mol. Biol. Evol. **23:** 973–980.

Muller, H. J., 1942 Isolating mechanisms, evolution and temperature. Biol. Symp. **6:** 71–125.

Ndjiondjop, M. N., L. Albar, D. Fargette, C. Fauquet and A. Ghesquiere, 1999 The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. Plant Disease **83:** 931–935.

Ndjiondjop, M. N., L. Albar, D. Fargette, C. Brugidou, M. P. Jones *et al.*, 2003 Mapping a recessive gene conferring resistance to rice yellow mottle virus, pp. 257–259 in *Advances in Rice Genetics*, edited by G. S. Khush, D. S. Brar and B. Hardy. International Rice Research Institute, Manila, Philippines.

Orjuela, J., A. Garavito, M. Bouniol, J. Arbelaez, L. Moreno *et al.*, 2010 A universal core genetic map for rice. Theor. Appl. Genet. **120:** 563–572.

Orr, H. A., and S. Irving, 2005 Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. Genetics **169:** 671–682.

Orr, H. A., and M. Turelli, 2001 The evolution of postzygotic isolation: accumulating Dobzhansky–Muller incompatibilities. Evolution **55:** 1085–1094.

Ouyang, S., and C. R. Buell, 2004 The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res. **32:** D360–D363.

Ouyang, S., W. Zhu, J. Hamilton, H. Lin, M. Campbell *et al.*, 2007 The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. **35:** D883–D887.

Pesin, J. A., and T. L. Orr-Weaver, 2008 Regulation of APC/C activators in mitosis and meiosis. Ann. Rev. Cell Dev. Biol. **24:** 475–499.

Phadnis, N., and H. A. Orr, 2009 A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science **323:** 376–379.

Presgraves, D. C., L. Balagopalan, S. M. Abmayr and H. A. Orr, 2003 Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature **423:** 715–719.

Ren, G., P. Xu, X. Deng, J. Zhou, F. Hu *et al.*, 2005 A new gamete eliminator from *Oryza glaberrima*. Rice Genet. Newsl. **22:** 45–48.

Rice, P., I. Longden and A. Bleasby, 2000 EMBOSS: the European molecular biology open software suite. Trends Genet. **16:** 276–277.

Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice *et al.*, 2000 Artemis: sequence visualization and annotation. Bioinformatics **16:** 944–945.

Salamov, A. A., and V. V. Solovyev, 2000 *Ab initio* gene finding in *Drosophila* genomic DNA. Genome Res. **10:** 516–522.

Sano, Y., 1983 Genetic-studies of speciation in cultivated rice. 3. A new gene controlling sterility in F_1 hybrids of 2 cultivated rice species: its association with photoperiod sensitivity. J. Hered. **74:** 435–439.

Sano, Y., 1990 The genic nature of gamete eliminator in rice. Genetics **125:** 183–191.

Sano, Y., Y.-E. Chu and H.-I. Oka, 1979 Genetic studies of speciation in cultivated rice. 1. Genetic analysis for the F_1 sterility between *O. sativa* L. and *O. glaberrima* STEUD. Japan. J. Genet. **54:** 121–132.

Sano, Y., R. Sano, M. Eiguchi and H. Y. Hirano, 1994 Gamete eliminator adjacent to the *Wx* locus as revealed by pollen analysis in rice. J. Hered. **85:** 310–312.

Sarla, N., and B. P. M. Swamy, 2005 *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. Curr. Sci. **89:** 955–963.

Sonnhammer, E. L. L., and R. Durbin, 1995 A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:** GC1–GC10.

Soriano, I. R., V. Schmit, D. S. Brar, J.-C. Prot and G. Reversat, 1999 Resistance to rice root-knot nematode *Meloidogyne graminicola* identified in *Oryza longistaminata* and *O. glaberrima*. Nematology **1:** 395–398.

Suyama, M., D. Torrents and P. Bork, 2006 PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34:** W609–W612.

Sweigart, A. L., A. R. Mason and J. H. Willis, 2007 Natural variation for a hybrid incompatibility between two species of Mimulus. Evolution **61:** 141–151.

Taguchi, K., K. Doi and A. Yoshimura, 1999 RFLP mapping of *S19*, a gene for F_1 pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. Rice Genet. Newsl. **16:** 70–71.

Tang, S., and D. C. Presgraves, 2009 Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. Science **323:** 779–782.

Thiel, T., W. Michalek, R. Varshney and A. Graner, 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. **106:** 411–422.

Ting, C.-T., S.-C. Tsaur, M.-L. Wu and C.-I. Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504.

Toshio, M., and S. Folke, 1962 A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plantarum **15:** 473–497.

Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts *et al.*, 2007 Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. **35:** W71–W74.

Vogl, C., and S. Xu, 2000 Multipoint mapping of viability and segregation distorting loci using molecular markers. Genetics **155:** 1439–1447.

Wang, Y., and M. Yang, 2006 The ARABIDOPSIS SKP1-LIKE1 (ASK1) protein acts predominately from leptotene to pachytene and represses homologous recombination in male meiosis. Planta **223:** 613–617.

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

Wittbrodt, J., D. Adam, B. Malitschek, W. Maueler, F. Raulf *et al.*, 1989 Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. Nature **341:** 415–421.

Xu, G., H. Ma, M. Nei and H. Kong, 2009 Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. Proc. Natl. Acad. Sci. USA **106:** 835–840.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Zhang, Z., P. Xu, F. Hu, J. Zhou, J. Li *et al.*, 2005 A new sterile gene from *Oryza glaberrima* on chromosome 3. Rice Genet. Newsl. **22:** 26–29.

Zhao, Z. G., L. Jiang, W. W. Zhang, C. Y. Yu, S. S. Zhu *et al.*, 2007 Fine mapping of *S31*, a gene responsible for hybrid embryo-sac abortion in rice (*Oryza sativa* L.). Planta **226:** 1087–1096.

Zhu, C., C. Wang and Y. M. Zhang, 2007 Modeling segregation distortion for viability selection. I. Reconstruction of linkage maps with distorted markers. Theor. Appl. Genet. **114:** 295–305.

Zhu, S., L. Jiang, C. Wang, H. Zhai, D. Li *et al.*, 2005 The origin of weedy rice Ludao in China deduced by genome wide analysis of its hybrid sterility genes. Breeding Sci. **55:** 409–414.

Communicating editor: M. Kirst

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.110.116772/DC1

A Genetic Model for the Female Sterility Barrier Between Asian and African Cultivated Rice Species

Andrea Garavito, Romain Guyot, Jaime Lozano, Frédérick Gavory, Sylvie Samain, Olivier Panaud, Joe Tohme, Alain Ghesquière and Mathias Lorieux

DOI: 10.1534/genetics.110.116772

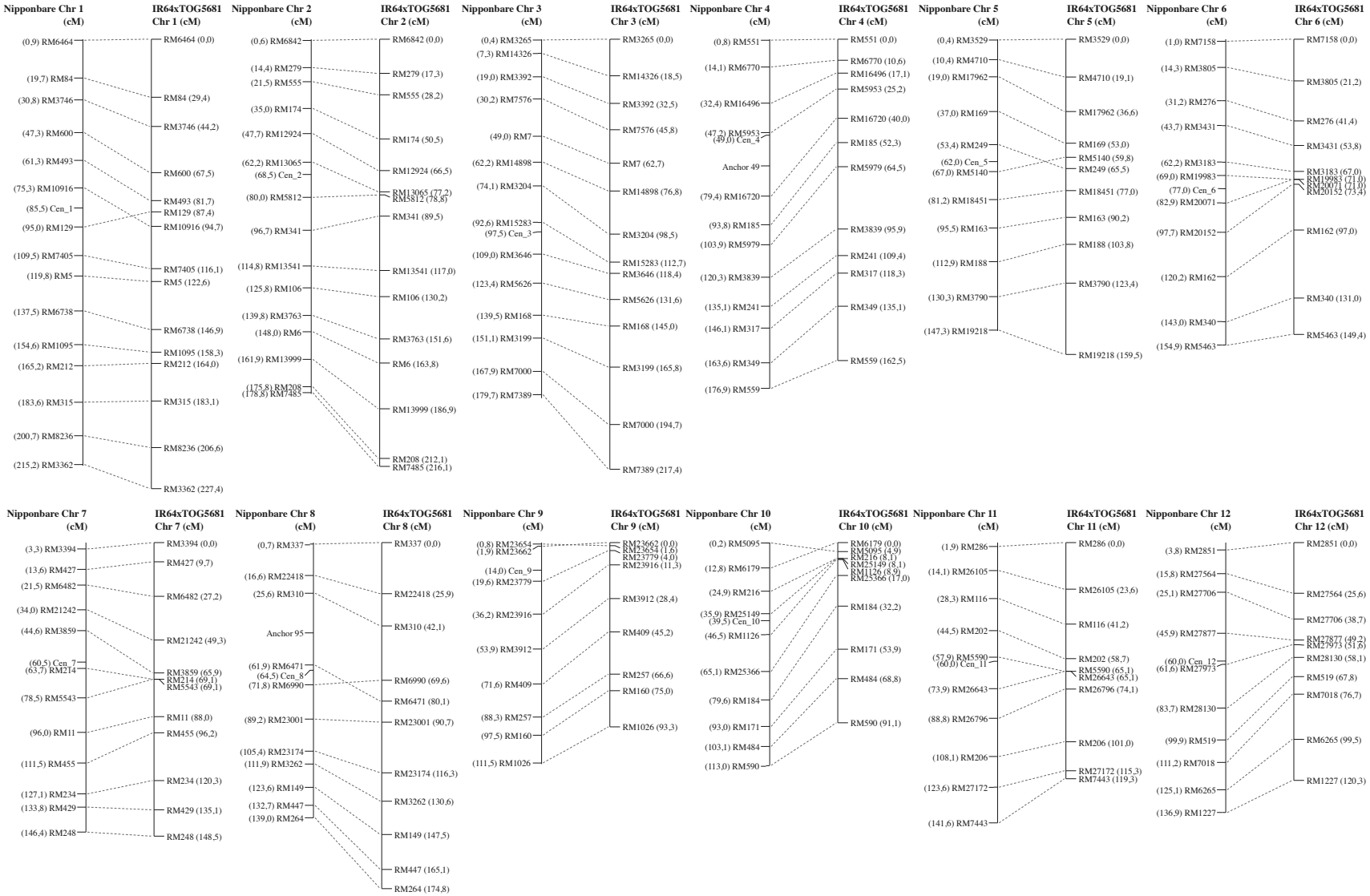


FIGURE S1.— Comparison of the *O. sativa* x *O. glaberrima* genetic map with the *O. sativa* cv. Nipponbare physical map. Comparison of the obtained genetic map for the IR64/TOG5681//IR64 cross, with the physical localization of SSR markers in the *O. sativa* cv. Nipponbare genome, based in the TIGR V.6 available data. **Cen:** Centromeric region**.**

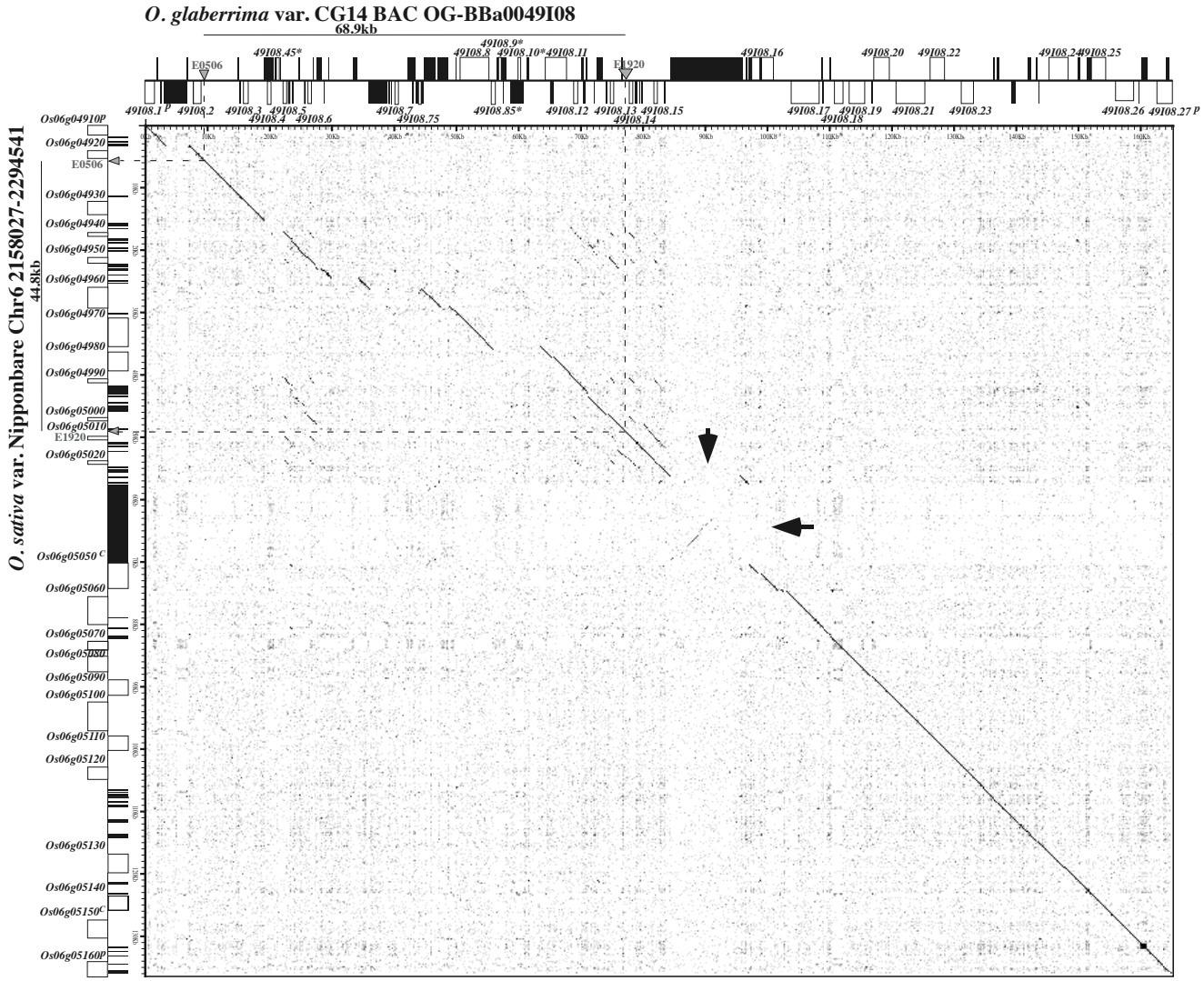


FIGURE S2.—Graphical comparison between *O. glaberrima* cv. CG14 BAC OG-BBa0049I08 and the orthologous *O. sativa* cv. Nipponbare genomic segment. Pairwise alignments of the *O. glaberrima* cv. CG14 BAC OG-BBa0049I08 and the respective *O. sativa* cv. Nipponbare orthologous region. Predicted genes (empty boxes) and TEs (dark bars) are shown for each of the sequences. Gray triangles mark the locus *S1*, as shown in Figure 4. **C:** Nipponbare genes manually corrected. **p:** Partial gene.
***:** Pseudogene. The insertion of two LTR retrotransposons are indicated with arrows.

FILE S1

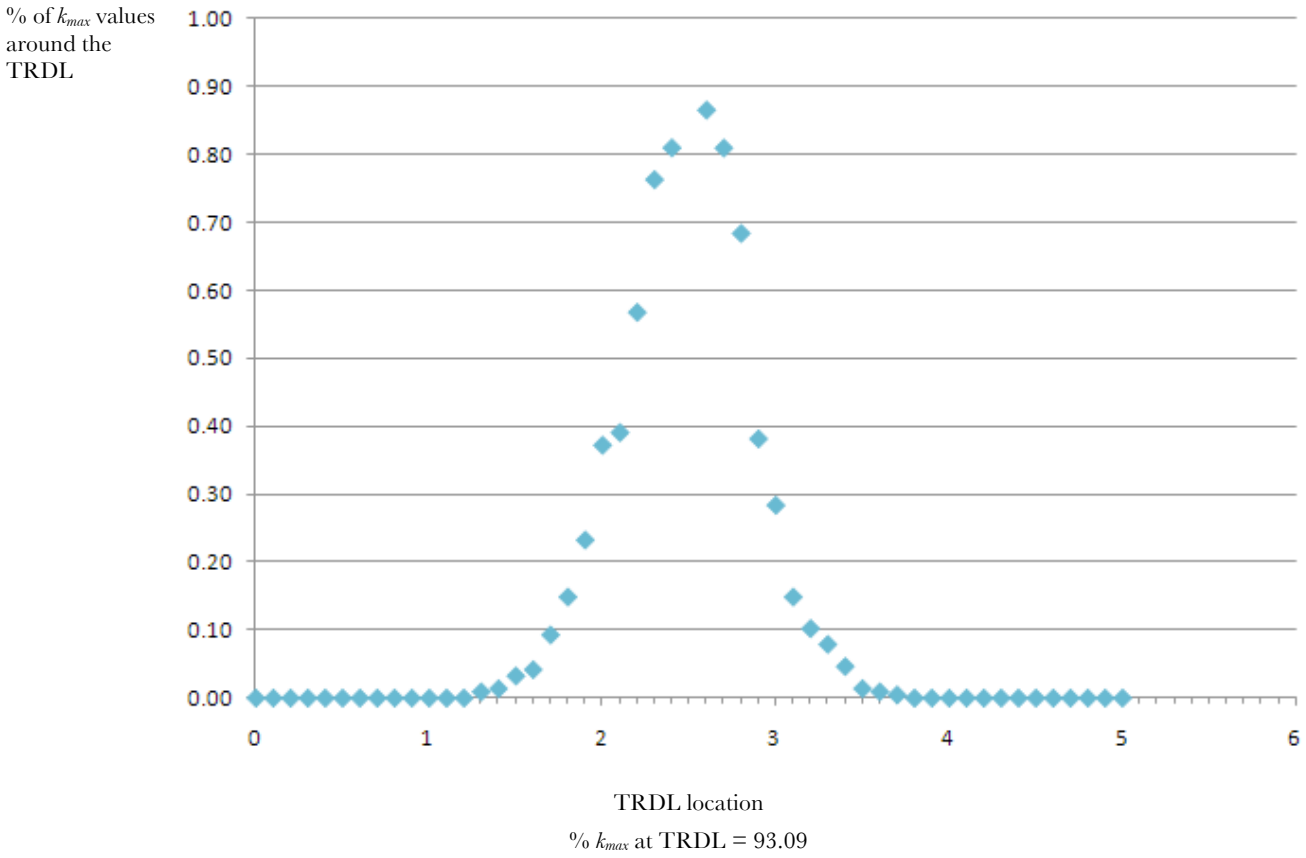
Genetic models for the allelic female gamete elimination caused by the *S1* locus

File S1 is available for download as an Excel file (.xlsx) at http://www.genetics.org/cgi/content/full/genetics.110.116772/DC1.

FILE S2

Estimation of the precision of the mapping of a transmission ratio distorter locus (TRDL) using the k_{max} measurement by simulation

In order to estimate the precision of the mapping of a transmission ratio distorter locus (TRDL) using the k_{max} measurement, we simulated 10,000 BC1F1 populations of 734 individuals each, bearing one chromosome segment of 5 cM long spanned by 50 markers (10 markers per cM). A TRDL was located in the middle of the segment at the position 2.5 cM, and induced a differential viability rate of 0.95. For each population generated, the k statistic was computed for each marker. We then determined the percentage of cases where the maximum value of k (k_{max}) was observed on each marker surrounding the TRDL and plotted this percentage value against the position on the chromosome segment.



The simulation results showed a percentage of k_{max} of 93.09 % at the TRDL position, and showed that 95.58 % of the k_{max} felt in an interval of 0.3 cM around the TRDL, meaning that there is a 95 % probability of observing a k_{max} in an interval of slightly less than 0.3 cM around the true TRDL position. Converting this genetic distance to a physical one using the local rate of 80 kbp/cM (derived from our data, see Figure 3) leads to an interval of slightly less than 24 kb.

FILE S3

Equations for calculating the expected gametic and allelic frequencies under Model 1

The expected frequencies F_c of the different 1n-2c configurations can be derived from the recombination frequencies between the three *S1* genes and between the *S1* complex locus and the centromere (see schematic representation 1 below), and are given in Table S3-1.

The expected frequencies of each gamete that belong to a specific 1n-2c configuration are equal to $F_c/4$, where F_c is the expected frequency of the specific 1n-2c configuration.

Let *D* be the sum of the frequencies of the viable gametes. Under model 1, *D* simplifies to:

$$D = (1 - 2r_5)\left(\frac{r_2 + 1}{2}\right) + 2r_5\left[\frac{1}{2} + \frac{r_1}{2} - r_1 r_2\right]$$

The expected frequencies of *viable* gametes are obtained by dividing their corresponding expected frequency by *D*.

Thus, the final frequency of gametes bearing the S_1^s allele, $f\left(S_1^s\right)$, is obtained by summing the individual frequencies of viable gamete that bear this allele. Under Model 1, this simplifies to:

$$f\left(S_1^s\right) = \left[\frac{r_2}{2} + r_5\left(r_1 - r_2 - 2r_1 r_2\right)\right] \Big/ D$$

Similar equations can be derived for Models 2 to 4, with numerical examples based on observed data given in File "S1 - Genetic models.xlsx".

Under all models, the estimated for recombination fractions r_1 and r_2 are naturally given by dividing the observed recombination fractions, $r_1$max and $r_2$max, by 2*D*.

$$\hat{r}_1 = \frac{r_1 \max}{2D};\ \hat{r}_2 = \frac{r_2 \max}{2D}.$$

In the case of the presence of the additional factor *S1C* (see schematic representation 2 below), the effect of *S1C* is applied after the initial selection due to the *S1A*- *S1*- *S1B* locus action.

As *S1C* only affects the $S1^s$ gametes when a recombination event occurs between *S1B* and *S1C*, the expected frequencies of these gametes are obtained by multiplying them by $1 - r_3$, where r_3 is the recombination fraction between *S1B* and *S1C*, and dividing them by the relative sum of all gamete frequencies, $1 - r_3 f\left(S_1^s\right)$. The frequencies of other viable gametes are obtained by dividing them by the relative sum of all gamete frequencies, $1 - r_3 f\left(S_1^s\right)$.

TABLE S3-1

Expected frequencies F_c of the different 1n-2c configuration expressed in function of the recombination frequencies between the three *S*$_1$ genes, and associated gametic frequencies under no selection and Model 1 selection

1n-2c configuration	*Expected configuration frequency without selection (F_c)*	*Expected gamete frequency without selection (F_g)*	*Survival under Model 1*	*Expected gamete frequency under selection (Model 1)*
S$_1$*A* *S*$_1$ *S*$_1$*B* s s s s s s	$F_c^{\,1} = (1-2r_5)\left[1-2\left(r_1+r_2-2r_1r_2\right)\right]$	$F_c^{\,1}/4$	Aborted	
		$F_c^{\,1}/4$	Aborted	
g g g g g g		$F_c^{\,1}/4$	Viable	F_g/D
		$F_c^{\,1}/4$	Viable	F_g/D
s s s g g g	$F_c^{\,2} = 2r_5\left[1-2\left(r_1+r_2-2r_1r_2\right)\right]$	$F_c^{\,2}/4$	Aborted	
		$F_c^{\,2}/4$	Viable	F_g/D
s s s g g g		$F_c^{\,2}/4$	Aborted	
		$F_c^{\,2}/4$	Viable	F_g/D
s s s g g s X	$F_c^{\,3} = 2r_2(1-2r_1)(1-2r_5)$	$F_c^{\,3}/4$	Viable	F_g/D
		$F_c^{\,3}/4$	Viable	F_g/D
s s g g g g		$F_c^{\,3}/4$	Aborted	0
		$F_c^{\,3}/4$	Viable	F_g/D
s s s g s s X	$F_c^{\,4} = 2r_1(1-2r_2)(1-2r_5)$	$F_c^{\,4}/4$	Aborted	0
		$F_c^{\,4}/4$	Aborted	0
s g g g g g		$F_c^{\,4}/4$	Viable	F_g/D
		$F_c^{\,4}/4$	Viable	F_g/D
s s s s g s X X	$F_c^{\,5} = 4r_1r_2(1-2r_5)$	$F_c^{\,5}/4$	Viable	F_g/D
		$F_c^{\,5}/4$	Viable	F_g/D
g s g g g g		$F_c^{\,5}/4$	Aborted	0
		$F_c^{\,5}/4$	Viable	F_g/D

s s s / s s g × g g s / g g g	$F_c^{\,6} = 4r_2r_5(1-2r_1)$	$F_c^{\,6}/4$	Aborted 0
		$F_c^{\,6}/4$	Aborted 0
		$F_c^{\,6}/4$	Viable F_g/D
		$F_c^{\,6}/4$	Viable F_g/D
s s s / s g g × g s s / g g g	$F_c^{\,7} = 4r_1r_5(1-2r_2)$	$F_c^{\,7}/4$	Viable F_g/D
		$F_c^{\,7}/4$	Viable F_g/D
		$F_c^{\,7}/4$	Aborted 0
		$F_c^{\,7}/4$	Viable F_g/D
s s s / g s g × s g s / g g g	$F_c^{\,8} = 8r_1r_2r_5$	$F_c^{\,8}/4$	Aborted 0
		$F_c^{\,8}/4$	Aborted 0
		$F_c^{\,8}/4$	Viable F_g/D
		$F_c^{\,8}/4$	Viable F_g/D

The sum of the frequencies of the viable gametes thus simplifies to:

$$D' = D\Big/\left(1 - f\left(S_1^s\right)\right)r_3,$$

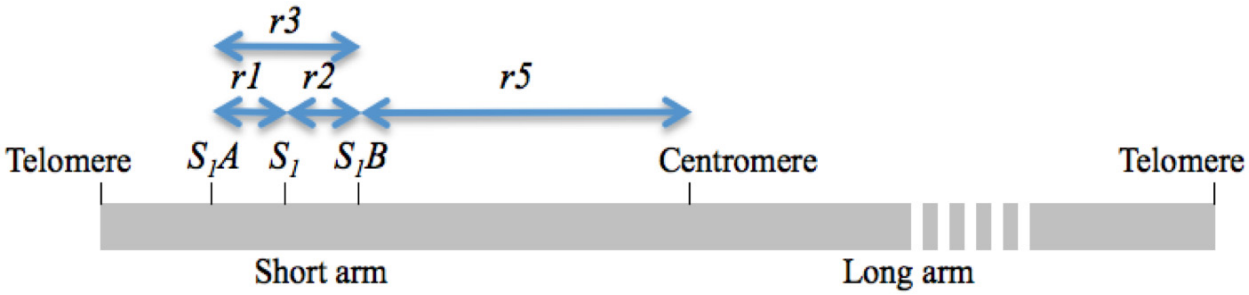
while the final frequency of gametes bearing the $S_1^{\,s}$ allele converts to:

$$f'\left(S_1^s\right) = \left(1 - r_3\right)f\left(S_1^s\right)\Big/\left(1 - f\left(S_1^s\right)r_3\right),$$

and the estimates for r_1 and r_2 convert to:

$$\hat{r}_1 = \frac{r_1 \max}{2D'};\ \hat{r}_2 = \frac{r_2 \max}{2D'}.$$

Schematic representation 1: positions of the three S_1 genes on rice chromosome 6, expressed as recombination franctions between each other and between the S_1 locus and the centromere.



Schematic representation 2: positions of the three S_1 genes and the S_1C gene on rice chromosome 6, expressed as recombination fractions between each other and between the S_1 locus and the centromere.

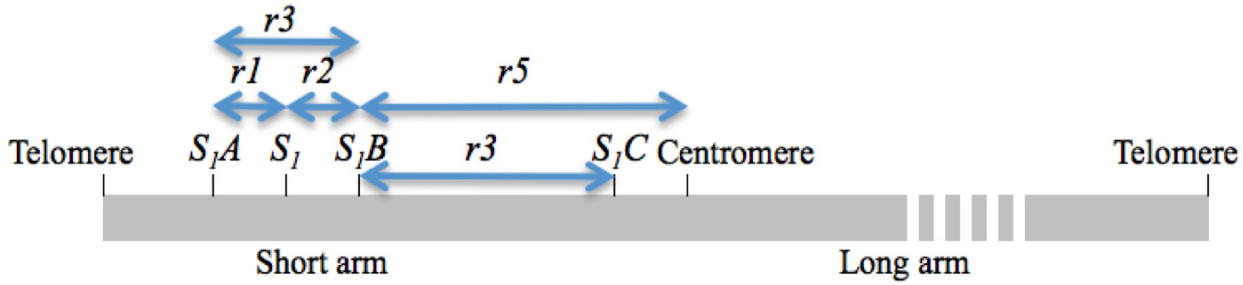


TABLE S1

SSR markers and PCR conditions used for saturation of chromosome 6

Molecular marker	Position in Nipponbare Chromosome 6 (TIGR V.6)	Annealing Temperature °C	$MgCl_2$ final concentration (mM)
RM7158	216 292	50	2.5
RM435	537 327	55	1.5
RM589	1 380 866	55	1.5
RM190	1 764 638	55	1.5
RM19350	1 967 120	55	1.8
RM19353	2 084 117	57	1.5
RM19357	2 175 040	55	1.8
RM19359	2 205 367	59	2.0
RM19360	2 215 120	55	2.0
RM19361	2 245 211	55	1.8
RM5199	2 270 164	55	1.8
RM19363	2 288 872	55	1.8
RM19367	2 335 233	55	1.5
RM19369	2 349 474	55	2.0
RM19377	2 440 456	55	2.0
RM19387	2 659 693	55	1.5
RM19391	2 714 833	55	1.5
RM3805	2 853 068	55	1.5
RM19414	2 941 180	59	1.5
RM19420	3 076 926	61	1.5
RM204	3 168 377	55	1.5
RM19496	4 351 871	55	1.8
RM276	6 230 046	55	1.8
RM3431	8 744 955	50	1.8
RM3183	12 447 059	50	1.5
RM19983	13 808 188	55	1.5
RM20071	16 542 271	55	1.8
RM20152	19 542 217	55	1.8
RM6036	21 548 402	59	1.5
RM20313	23 161 984	55	1.5
RM162	24 035 501	61	2.0
RM5371	25 825 382	55	1.5
RM340	28 599 182	55	1.8
RM5463	30 985 008	55	1.8

TABLE S2

New SSR and InDel markers in the S_1 locus

Molecular marker	Position in Nipponbare Chromosome 6 (TIGR V.6)	Forward Primer	Reverse Primer	Annealing Temperature °C	$MgCl_2$ concentration mM
RMC6_21678	2 167 869	AGGGAGCTCCAGTGTTGAGA	TTGTGCCCAAGCTTTAAGGTA	59	1.5
C6_21774	2 177 472	CAGTTTTGGGCATTTTTCG	ATAGGCGATCCCTAGAGCAA	61	1.5
C6_21788	2 178 838	GCTGCGATGATAAGGCAATA	CCCCTAGCTAGTTGGTGTGA	61	1.5
C6_21804	2 180 401	CCTTTACTTCTTTTGGAAGTTGG	GCACCACAGTTTCAGGCTCT	61	1.5
C6_21824	2 182 413	CTGCTACCATCCCTCTGCTC	GGAGCAATTTGAGAAGCTG	61	1.5
C6_21837	2 183 714	TGGATAGGGCTATAGGAGCA	CTTCCCTTCTCTGGATCGAC	61	1.5
RMC6_21851	2 185 106	AAGGGTCCTTTCCTCTAGCC	AGGGCGATTGACGCACAT	59	1.5
RMC6_21942	2 194 257	TCCAATCCCCCTAATTTCAG	TTTACCATGGGGAACTTGGA	59	1.5
RMC6_21989	2 198 927	CCTGATGAGCTCTGCTGATG	GTGCGTGCTTTGTGAATCTG	59	1.5
RMC6_22028	2 202 836	CACCATCCTCCTCCATGTTT	CCCTTTCATTTTGCCGTCTA	59	1.5
RMC6_22046	2 204 666	AAGAAGGTGAAGACATCAACGA	GCTGCTGAGTTCATGCGTAA	59	1.5
RMC6_22105	2 210 541	GATTTTGCCGTCAGTTTTGG	CCATACACGCAGCTCACAG	59	1.5
C6_22305	2 230 535	CAAGTGAAGCTGGTCATTGG	TCCGTTTGTTGGAATGGAAT	61	1.5
C6_22386	2 238 675	TGTATGAGAGGGTAGGATCAGTCA	TGCTCATGGAGAGCGATAAA	61	1.5
RMC6_22500	2 250 065	CTGCATCCAAAGTTCTGAAGG	ATCTGTTCATGGGGGCTGTA	59	1.5
RMC6_22639	2 263 917	TTGCCATTTCATGGAAATTTAG	GACTCTTAGCCACCAAGGAAA	59	1.5
C6_22773	2 277 347	GACCTTGATGTCAACCAGCA	TCGAAATTGGAGCACTCTGA	61	1.5
RMC6_22854	2 285 471	GGCTAGAGCGGTGATTTTCA	TGGTAGGCACAAGGATAGGG	59	1.5
RMC6_25568 *	2 556 896	ACGTCAGGGAGAGGTACTTT	CCCTAGGAAGGAAGAGAGAA	55	1.5

SSR and InDels based molecular markers, designed by comparison of the *O. glaberrima* cv. CG14 BAC OG-BBa0049I08 with the orthologous segment from *O. sativa* cv. Nipponbare. PCR conditions used for each marker are shown. ***:** SSR marker designed based only in the TIGR V6 data (putative_ssr_21715).

TABLE S3

List of identified putative TEs in the *O. glaberrima* BAC OG-BBA0049I08

Class	Repeat type	Group	Number	Size (bp)
I	LTR retrotransposon	**LTR retrotransposon**	7	18534
	Non-LTR retrotransposon	**LINE**	1	161
		SINE	2	401
II	Transposons	**DNA Transposon**	29	16887
		MITE	34	8279
	Unclassified	**Unclassified**	1	905
Total			74	45167

TABLE S4

Sequence annotation and comparison between orthologous coding sequences in the *O. glaberrima* cv. CG14 BAC OG-BBa0049I08 and *O. sativa* cv. Nipponbare

O. glaberrima Gene Name	Best BLASTN homology	Best BLASTN EST homology	Best BLASTX homology (swiss prot)	Protein domain	Putative Function	Putative *O. sativa* (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	Ka (dN)	Ks (dS)	Ka/Ks
OG-BBa0049I08.1P	NM_001063292 *O. sativa* Os06g0141100 (0.0)	CB618108 *O. sativa* (0.0)	Q94A82 *A. thaliana* NADH pyrophosphatase NUDT19 (8e-69)	pfam09296, NUDIX-like; cd03429, NADH pyrophosphatase	Putative NADH pyrophosphatase protein	LOC_Os06g04910 P	99	98.4	98.4	1224	0.0025	0.0000	99
OG-BBa0049I08.2	CT830197 *O. sativa* flcDNA (0.0)	CT850081 *O. sativa* (0.0)	O13801 *Schizosaccharomyces pombe* RNA-binding protein (2e-9)	pfam00641, Zn-finger	Putative Zinc finger protein	LOC_Os06g04920	99.3	99.3	100	438	0.0028	0.0266	0.1063
OG-BBa0049I08.3	AK121184 *O. sativa* flcDNA (2e-131)	EE592322 *O. sativa* (5e-169)	Q02921 *G. max* Early nodulin 93 (2e-11)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04930	61.3	60.9	62.2	/	/	/	/
OG-BBa0049I08.4	AC136219 *O. sativa* BAC (0.0)	CT843513 *O. sativa* (1e-92)	P73714 Synechocystis Zinc metalloprotease (1e-9)	cd06163 zinc metalloproteases	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.45*	/	/	/	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.5	AK122162 *O. sativa* flcDNA (2e-148)	EL586675 *O. sativa* (2e-162)	Q02921 *G. max* Early nodulin 93 (8e-20)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04940	65	54.1	59.5	/	/	/	/
OG-BBa0049I08.6	AK121791 *O. sativa* flcDNA (1e-157)	CI240527 *O. sativa* (1e-158)	Q02921 *G. max* Early nodulin 93 (4e-14)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04950	89.1	76.8	78.4	/	/	/	/
OG-BBa0049I08.7	/	/	/	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.75	/	/	/	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.8	NM_001063299 *O. sativa* Os06g0142000 (0.0)	CB622551 *O. sativa* (0.0)	Q9SH88 *A. thaliana* Ribosome biogenesis regulatory protein (3e-48)	/	Putative protein	LOC_Os06g04970	98.6	97.7	98.4	2073	0.0112	0.0237	0.4728
OG-BBa0049I08.85*	AP002838 *O. sativa* BAC (2e-156)	CT858525 *O. sativa* (8e-67)	/	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.9*	AP004090 *O. sativa* BAC (8e-69)	CT844843 *O. sativa* (3e-68)	/	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049I08.10*	AC121363 *O. sativa* BAC (3e-128)	CB649669 *O. sativa* (1e-57)	P47927 *A. thaliana* Floral homeotic protein APETALA 2 (4e-8)	/	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/

OG-BBa0049I08.11	AK073027 *O. sativa* flcDNA (0.0)	CB658549 *O. sativa* (0.0)	Q9FJT2 *A. thaliana* F-box/FBD/LRR-repeat (5e-4)	/	Putative F-box protein	LOC_Os06g04980	98.1	96.5	97.7	1542	0.0173	0.0198	0.8715
OG-BBa0049I08.12	AK122162 *O. sativa* flcDNA (2e-175)	CI252819 *O. sativa* (2e-175)	Q02921 *G. max* Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g04990	99.7	100	100	348	0.0000	0.0142	0.001
OG-BBa0049I08.13	CT837557 *O. sativa* flcDNA (4e-177)	CT856080 *O. sativa* (4e-177)	Q02921 *G. max* Early nodulin 93 (1e-14)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g05000	100	100	100	348	0.0000	0.0000	99
OG-BBa0049I08.14	CT827995 *O. sativa* flcDNA (1e-158)	EE592204 *O. sativa* (5e-176)	Q02921 *G. max* Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g05010	99.4	100	100	351	0.000025	0.0260	0.001
OG-BBa0049I08.15	CT827995 *O. sativa* flcDNA (9e-173)	CI050172 *O. sativa* (1e-176)	Q02921 *G. max* Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein	LOC_Os06g05020	99.1	98.3	99.1	348	0.0074	0.0143	0.5175
OG-BBa0049I08.16	NM_001063304 *O. sativa* Os06g0142500 (0.0)	CB617832 *O. sativa* (0.0)	Q9LMN7 *A. thaliana* Wall-associated receptor kinase 5 (5e-130)	cd00180, Serine/Threonine protein kinases	Putative Serine/Threonine protein kinases	LOC_Os06g05050	92	87.8	92.1	2291	0.0447	0.0571	0.7828
OG-BBa0049I08.17	NM_001063305 *O. sativa* Os06g0142600 (0.0)	CB635675 *O. sativa* (0.0)	O82804 *A. thaliana* Protein Early flowering 3 (1e-55)	/	Putative Early flowering protein	LOC_Os06g05060	99.2	98.8	99.1	2283	0.0057	0.0155	0.3669
OG-BBa0049I08.18	AP000399 *O. sativa* BAC (0.0)	CT860562 *O. sativa* (0.0)	O23081 *A. thaliana* Cysteine-rich receptor-like protein kinase 41 (2e-57)	cd00180, Serine/Threonine protein kinases	Putative Serine/Threonine protein kinases	LOC_Os06g05070	97.6	97.6	98.5	1386	0.0081	0.1700	0.0465
OG-BBa0049I08.19	CT834467 *O. sativa* flcDNA (0.0)	CT848007 *O. sativa* (0.0)	P00428 *Bos taurus* Cytochrome c oxidase subunit 5B, mitochondrial (6e-14)	cd00924, Cytochrome c oxidase subunit Vb	Putative cytochrome oxidase	LOC_Os06g05080	99.8	99.4	99.4	465	0.0032	0.0000	99
OG-BBa0049I08.20	AK103971 *O. sativa* flcDNA (0.0)	CA998930 *O. sativa* (0.0)	Q9SNQ2 *O. sativa* protein arginine N-methyltransferase (0.0)	cd02440, S-adenosylmethionine-dependent methyltransferases	Putative methyltransferase protein	LOC_Os06g05090	98.8	98.4	99	1143	0.0071	0.0299	0.2366
OG-BBa0049I08.21	AK121920 *O. sativa* flcDNA (0.0)	EC366141 *O. sativa* (0.0)	Q38854 *A. thaliana* 1-deoxy-D-xylulose-5-phosphate synthase, chloroplastic (0.0)	cd02007, Thiamine pyrophosphate (TPP) family, DXS subfamily, TPP-binding module	Putative 1-deoxy-D-xylulose-5-phosphate synthase	LOC_Os06g05100	99.9	99.6	99.7	2169	0.0018	0.0000	99
OG-BBa0049I08.22	AK071301 *O. sativa* flcDNA (0.0)	CB672512 *O. sativa* (0.0)	P22302 *Nicotiana plumbaginifolia* Superoxide dismutase chloroplastic (5e-60)	PRK10543, superoxide dismutase; pfam02777, Iron/manganese superoxide dismutases	Putative Superoxide dismutase	LOC_Os06g05110	99.9	99.6	100	768	0.0016	0.0000	99
OG-BBa0049I08.23	AK061597 *O. sativa* flcDNA (9e-80)	CI020918 *O. sativa* (9e-80)	/	/	Putative protein	LOC_Os06g05120	/	/	/	/	/	/	/

OG-BBa0049I08.24	AK120946 *O. sativa* flcDNA (0.0)	CX101001 *O. sativa* (0.0)	Q9SQI3 *Gossypium hirsutum* Myristoyl-acyl carrier protein thioesterase, chloroplastic (6e-146)	cd00586, 4-hydroxybenzoyl-CoA thioesterase (4HBT)	Putative Myristoyl-acyl carrier protein thioesterase	LOC_Os06g05130	97.4	99	99	1284	0.0044	0.0300	0.1457
OG-BBa0049I08.25	AP003487 *O. sativa* BAC (0.0)	CA083999 *Saccharum officinarum* (1e-99)	O49287 *A. thaliana* Putative pentatricopeptide repeat-containing protein (2e-93)	pfam01535, PPR repeat	Putative PPR protein	LOC_Os06g05140	99.8	99.6	99.6	2292	0.0018	0.0037	0.4789
OG-BBa0049I08.26	AP003487 *O. sativa* BAC (0.0)	CF326731 *O. sativa* (0.0)	Q80SY5 *M. musculus* Pre-mRNA-splicing factor 38B (2e-40)	pfam03371, PRP38 family	Putative protein	LOC_Os06g05150 C	98	97.2	97.5	1296	0.0070	0.0048	1.471
OG-BBa0049I08.27P	AK067270 *O. sativa* flcDNA (0.0)	FL869738 *Panicum virgatum* (0.0)	Q9LW86 *A. thaliana* Probable sulfate transporter (1e-141)	pfam00916,Sulfate transporter family	Putative sulfate transporter	LOC_Os06g05160 P	/	/	/	/	/	/	/

p: Partial gene; ***:** Pseudogene; **C:** Nipponbare genes manually corrected. Genes in grey are located in the S_1 locus.