

**UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC**

T H E S E

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

Discipline :

Biologie Intégrative des Plantes

Ecole Doctorale

**SIBAGHE - Systèmes intégrés en Biologie, Agronomie,
Géosciences, Hydrosciences et Environnement**

présentée et soutenue publiquement par

Andrea María GARAVITO-ESPEJO

le 15 décembre 2009

Titre :

**Inférences génétiques et génomiques pour la caractérisation du système majeur
de stérilité hybride entre les riz cultivés *Oryza sativa* et *O. glaberrima***

JURY

M. Jean-Loup NOTTEGHEM

Président

M. Mathias LORIEUX

Co-directeur de Thèse

M. Pascal TOUZET

Rapporteur

M. Jérôme SALSE

Rapporteur

Résumé : Les deux espèces de riz cultivé *O. sativa* et *O. glaberrima* possèdent des caractères complémentaires importants pour l'agriculture, comme un bon rendement et une meilleure qualité de grain pour *O. sativa*, et plusieurs gènes de résistance ou de tolérance à différents stress biotiques et abiotiques pour *O. glaberrima*. On souhaiterait donc pouvoir développer des lignées hybrides qui réunissent l'ensemble de ces caractéristiques. Cependant, les deux espèces sont séparées par une forte barrière reproductrice, qui est le résultat de l'action cumulée de plusieurs locus qui rendent les hybrides F₁ complètement mâle stériles et partiellement femelle stériles. Parmi ceux-ci, le locus *S₁* exerce l'effet le plus important sur la fertilité des hybrides.

L'objectif principal fixé pour ce travail de thèse est de contribuer à déterminer les bases génétiques de la barrière reproductrice entre *O. sativa* et *O. glaberrima*. A cette fin, nous avons utilisé différentes approches de génétique et de génomique comparative afin de mieux caractériser le locus *S₁*.

Nous avons ainsi pu déchiffrer et mieux comprendre le fonctionnement de la barrière reproductrice entre *O. sativa* et *O. glaberrima* exercée par le locus *S₁*. Les résultats obtenus nous ont permis de détecter la présence d'au moins deux autres locus génétiquement liés à *S₁*, de discerner de possibles interactions épistatiques et des incompatibilités alléliques entre ces locus, de construire un modèle génétique explicatif de la semi stérilité femelle des hybrides, d'explorer la région génomique qui porte ces locus chez des variétés *O. glaberrima* et son ancêtre *O. barthii*, et finalement, d'identifier des gènes candidats pour deux des locus (*S₁* et *S_{1A}*). La somme des connaissances générée par ce travail permet d'établir des stratégies visant à faciliter le transfert de gènes utiles d'*O. glaberrima* vers *O. sativa*, et ainsi de mieux exploiter la diversité génétique de l'espèce africaine de riz cultivé.

Genetic and genomic inferences for the characterization of the major sterility system between the two rice cultivated species, *Oryza sativa* and *O. glaberrima*

Abstract: The two cultivated rice species *O. sativa* and *O. glaberrima* possess many interesting and complementary traits, such as high yield and good grain quality for the former, and several alleles for resistance or tolerance to different biotic and abiotic stresses for the latter. It would be desirable to combine these characteristics into interspecific lines. However, a strong reproductive barrier impairs the genetic exchange between the two species. This barrier is the result of the action of several loci that might interact or act separately to render the F₁ hybrids pollen-sterile and partially female-fertile. Among these, the *S₁* locus has the strongest effect over the fertility of the hybrids.

The main aim of this work is to contribute to a better understanding of the genetic bases of the reproductive barrier that separates *O. sativa* and *O. glaberrima*. For this purpose, we applied different genetic and genomic approaches in order to characterize the *S₁* locus.

We made a significant advance in the understanding of the reproductive barrier between *O. sativa* and *O. glaberrima* controlled by the *S₁* locus. We were able to detect the presence of at least two additional loci, highly linked to *S₁*, to infer the possible epistatic interactions and allelic incompatibilities between them, to develop a genetic model that explains the female semi-sterility of the hybrids, to explore the genomic region that bears these loci, into a collection of *O. glaberrima* and its ancestor *O. barthii*, and finally, to identify possible candidate genes for two of the loci (*S₁* and *S_{1A}*). The knowledge generated by this work will contribute to the establishment of new strategies to introgress more efficiently useful genes from *O. glaberrima* to *O. sativa*, allowing breeders to take advantage of the genetic diversity present in the cultivated African rice species.

Mots Clés:

Oryza sativa, *Oryza glaberrima*, Stérilité hybride, Distorsion de ségrégation, Incompatibilités du type Bateson-Dobzhansky-Muller, Croisements interspécifiques, Spéciation.

Laboratoire Génome et Développement des Plantes (UMR 5096 – CNRS – IRD – UPVD),
Institut de Recherche pour le Développement (IRD)

A ma famille et à l'amour de ma vie.

REMERCIEMENTS

La première personne que je tiens à remercier est Alain Ghesquière, pour m'avoir donné l'opportunité de réaliser le présent travail de thèse dans son laboratoire, pour ses conseils et son implication, je lui suis très reconnaissante.

Je tiens aussi à remercier Mathias Lorieux pour m'avoir toujours appuyée, de m'avoir ouvert les portes de l'IRD, pour tous ses conseils, et spécialement pour le temps qu'il a investi pour la correction du présent manuscrit.

J'aimerais remercier Romain Guyot pour son important investissement dans le projet, et pour ses conseils précieux.

Mes remerciements vont également aux membres du Laboratoire Génome et Développement des Plantes, particulièrement ceux basés à l'IRD Montpellier, pour tous leurs conseils et aussi pour leur disponibilité, les discussions scientifiques, et de la vie.

Je présente aussi toute ma reconnaissance au Docteur Jérôme Salse et au Professeur Pascal Touzet, pour avoir accepté de rapporter ma thèse doctorale et de faire partie de mon jury de thèse, ainsi qu'au professeur Jean-Loup Notteghem d'avoir accepté de présider le jury de cette thèse.

Je tiens à exprimer ma reconnaissance envers le Département de Soutien et Formation, de l'IRD, et au CIAT pour m'avoir attribué la bourse qui m'a permis d'entreprendre le présent travail de thèse ainsi qu'au GCP pour le financement des travaux.

Je remercie aussi mon amie, Julie Orjuela, pour être toujours disponible avec la meilleure volonté d'aider, et pour son soutien tout au long de la thèse.

Mis más sinceros agradecimientos al personal de la Unidad de Biotecnología del CIAT, por su enorme colaboración en el trabajo. A Joe Tohme, por estar siempre pendiente durante el transcurso de mis trabajos en el CIAT.

Igualmente mil gracias a Jaime Lozano, a Mabel Morales, así como a Marcos Brito y a Víctor Lozano, por su implicación directa en la obtención del material vegetal; y a Myriam C. Duque, por sus siempre claras respuestas.

Gracias a Andrés Gutiérrez, y a María Fernanda Álvarez, por su buena voluntad, colaboración y amistad. Igualmente a Luisa, Yamid, Jesús, Alicia, Luz Elena y Cristian por los buenos momentos, que hicieron muy agradable el trabajo en el laboratorio.

A mi familia y en especial a mi madre, por sus enseñanzas que siempre fueron fuente de inspiración divina, por darme siempre ánimo y tener confianza en mí.

Un agradecimiento de lo mas profundo de mi corazón, para el amor de mi vida, porque su apoyo, comprensión, compañía fueron la fuerza que me permitió llegar hasta el final. Pour tout cela et plus encore, je t'aime. Esta tesis es por ustedes.

TABLE DES MATIERES

I	Introduction.....	15
1.1	Positionnement général.....	17
1.2	Le riz et son importance dans le monde.....	19
1.2.1	Les différents types de culture du riz.....	19
1.2.2	Production de riz dans le monde.....	19
1.2.3	Importance nutritionnelle.....	21
1.2.4	Importance culturelle.....	21
1.2.5	Importance économique.....	21
1.3	Les principales contraintes de la production du riz.....	23
1.3.1	Maladies infectieuses.....	25
1.3.1.1	Champignons.....	25
1.3.1.2	Bactéries.....	26
1.3.1.3	Virus.....	26
1.3.1.4	Parasites et compétiteurs.....	27
1.3.1.5	Nématodes.....	27
1.3.1.6	Insectes.....	27
1.3.1.7	Adventices.....	28
1.4	Comment faire face aux défis de la riziculture ?.....	28
1.5	La réponse génétique : l'amélioration variétale.....	29
1.5.1	Hybridations interspécifiques <i>versus</i> les hybridations intraspécifiques.....	29
1.5.1.1	Classification taxonomique du genre <i>Oryza</i>	31
1.5.2	Le point sur les hybridations interspécifiques chez le riz.....	33
1.5.2.1	Populations interspécifiques du complexe <i>O. sativa</i> et gènes d'intérêt de ces espèces.....	35
1.5.2.1.1	<i>O. rufipogon</i>	35
1.5.2.1.2	<i>O. nivara</i>	37
1.5.2.1.3	<i>O. glumaepatula</i>	37
1.5.2.1.4	<i>O. longistaminata</i>	39
1.5.2.1.5	<i>O. meridionalis</i>	39
1.5.2.1.6	<i>O. barthii</i>	40
1.5.2.1.7	<i>O. glaberrima</i>	40
1.5.2.1.8	Projet pour le développement de ressources génétiques et CSSL à partir de quatre espèces de riz du génome AA.....	45
1.5.3	Limitations des approches interspécifiques.....	45
1.6	Les barrières reproductives chez les plantes.....	47
1.6.1	Barrières reproductives avant croisement.....	47
1.6.2	Barrières reproductives après croisement, agissant avant la fécondation – cas des barrières pré-zygotiques.....	49

I.6.3	Barrières reproductives après la fécondation, agissant après la formation d'un zygote – cas des barrières post-zygotiques.....	49
I.6.3.1	Létalité et faiblesse hybride de la F_1	49
I.6.3.2	Stérilité hybride	51
I.6.3.3	<i>Hybrid breakdown</i> ou faiblesse de la F_2	51
I.6.3.4	Modèle Bateson-Dobzhansky-Muller (BDM) pour les barrières post-zygotiques	53
I.6.4	Distorsion de la ségrégation	54
I.6.5	Barrières reproductives décrites chez le riz	55
I.7	Importance de l'étude des barrières d'isolement reproductif pour le développement de stratégies d'amélioration	60
I.7.1	Ponts interspécifiques – iBridges	60
I.8	Présentation du sujet de recherche.....	63
II	Une Carte Génétique Universelle pour le Riz	65
	Introduction	67
	A universal core genetic map for rice.....	69
III	Cartographie du locus S_I et modèle génétique de sélection gamétique.....	89
	Introduction	91
	A genetic model for the female sterility barrier between Asian and African cultivated rice species.....	93
IV	Analyse détaillée de la région du locus S_I.....	129
	Introduction	131
	Patterns of evolution of three hybrid sterility loci between Asian and African cultivated rice species.....	133
V	Exploration de la région du locus S_I chez <i>O. glaberrima</i> et son ancêtre <i>O. barthii</i>..	171
	Introduction	173
V.1	Matériels et méthodes	175
V.1.1	Construction de la carte physique du locus S_I dans l'accèsion <i>O. glaberrima</i> TOG5681	175
V.1.2	Séquençage, annotation et comparaison des séquences partielles de TOG5681 avec CG14	175
V.1.3	Analyse de la structure de la région $S_I A - S_I - S_I B$ chez différentes accessions <i>O. glaberrima</i> et <i>O. barthii</i>	177
V.1.4	Evaluations morphologiques des accessions	181
V.1.5	Evaluations de la fertilité des variétés du projet iBridges en croisement avec <i>O. sativa</i>	181
V.2	Résultats.....	183
V.2.1	Carte physique de la région de S_I chez TOG5681	183

V.2.2	Séquençage préliminaire de TOG5681 et comparaison avec CG14.....	183
V.2.3	Evaluation de la région génomique autour de S_I dans une collection d' <i>O. glaberrima</i> et d' <i>O. barthii</i>	185
V.2.4	Sélection allélique et fertilité en croisement des variétés hybrides naturels.....	191
V.3	Discussion.....	193
V.4	Annexe.....	196
VI	Discussion	197
VI.1	Récapitulation des principaux résultats	199
VI.2	Construction, analyse et critique du modèle génétique proposé	200
VI.3	Vers un modèle général de la barrière reproductive entre <i>O. sativa</i> et <i>O. glaberrima</i>	202
VI.4	Vues sur la spéciation : conclusions sur la divergence des complexes <i>O. sativa</i> - <i>O. rufipogon</i> et <i>O. glaberrima</i> - <i>O. barthii</i>	208
VI.5	Conséquences sur l'étude des variétés hybrides naturels	211
VI.6	Conséquences sur les stratégies d'amélioration.....	212
VI.7	Les gènes candidats	213
VI.7.1	Les gènes hypothétiques	214
VI.7.2	Les gènes homologue aux Nodulines	214
VI.7.3	Le gène régulateur de la biogenèse des ribosomes.....	214
VI.7.4	Le Pack-MULE portant un fragment de AP2	215
VI.7.5	Le gène F-box	216
VI.7.6	Candidats pour $S_I A$ et $S_I B$	219
VI.8	Conclusion générale.....	220
VI.9	Perspectives	221
VI.9.1	Cartographie fine de $S_I A$ et $S_I B$	221
VI.9.2	Identification et clonage du facteur S_I	223
VI.9.3	Cartographie fine et identification des autres locus de stérilité entre <i>O. sativa</i> et <i>O. glaberrima</i>	225
VI.9.4	Etude des hybrides naturels	226
VII	Références Bibliographiques.....	229

LISTE DES FIGURES

Figure I-1. Production mondiale du riz et réserves mondiales.....	22
Figure I-2. Relations phylogénétiques entre les espèces du genre <i>Oryza</i>	32
Figure I-3. Les différentes espèces du complexe <i>O. sativa</i> (génom AA).....	36
Figure I-4. Exemples de populations CSSL issues des croisements entre <i>O. sativa</i> et <i>O. glaberrima</i>	42
Figure I-5. Lignées CSSL développées dans le cadre du projet du GCP pour l'exploration de la diversité naturelle du riz	44
Figure I-6. Modèle d'incompatibilité BDM dans les barrières reproductives post- zygotiques.....	52
Figure II-1. A. Distribution of SSR markers that form the Universal Core Genetic Map, arranged in anchors throughout the rice genome	78
Figure III-1. Genetic map of the <i>O. sativa</i> cv. IR64 x <i>O. glaberrima</i> cv. TOG5681 cross, fTRD genomic sites, and previously described sterility loci	98
Figure III-2. Mapping of the female component of the <i>S₁</i> locus using the fTRD measure ..	100
Figure III-3. Fine mapping of the female component of the <i>S₁</i> locus.....	102
Figure III-4. Graphical comparison of the <i>O. sativa</i> and <i>O. glaberrima</i> sequences in the <i>S₁</i> locus	106
Figure III-5. Predicted model for the allelic gamete elimination involving <i>S₁</i>	108
Supplemental Figure III-1. Comparison of the <i>O. sativa</i> x <i>O. glaberrima</i> genetic map with the <i>O. sativa</i> cv. Nipponbare physical map.....	120
Supplemental Figure III-2. Graphical comparison between <i>O. glaberrima</i> cv. CG14 BAC OG-BBa0049I08 and the orthologous <i>O. sativa</i> cv. Nipponbare genomic segment.....	121
Supplemental Figure III-3. Supplemental data: S1 model.xlsx	126
Figure IV-1. Genetic and physical maps of <i>S₁</i> , <i>S₁A</i> and <i>S₁B</i> loci	136
Figure IV-2. Physical map and annotation of the 813kb region of the <i>S₁</i> , <i>S₁A</i> and <i>S₁B</i> loci in <i>O. glaberrima</i> cv. CG14.....	138
Figure IV-3. Orthologous sequence comparisons between the 813 kb and 847 kb from the <i>O. glaberrima</i> cv. GC14 and <i>O. sativa</i> ssp. <i>japonica</i> cv. Nipponbare <i>S₁</i> regions	142
Figure IV-4. Comparisons of the orthologous sequences around the 66E18.3/66E18.4 CG14 genes	144

Figure IV-5. Comparisons of the orthologous sequences around the 88O22.2/88O22.7 CG14 genes	146
Figure IV-6. Representation of Ka and Ks values between 120 <i>O. sativa</i> and <i>O. glaberrima</i> orthologous genes in the <i>S_I</i> regions.....	148
Supplemental Figure IV-1. Schematic representation of microcollinearity relationships between <i>O. glaberrima</i> cv. GC14 and <i>O. sativa</i> cv. Nipponbare <i>S_I</i> regions	152
Supplemental Figure IV-2. Comparison between structural variations identified between orthologous <i>S_I</i> regions in CG 14 and Nipponbare, and the genetic map of the <i>S_I</i> locus between <i>japonica</i> and <i>O. glaberrima</i>	153
Supplemental Figure IV-3. Comparison between structural variations identified between orthologous <i>S_I</i> regions in CG 14 and Nipponbare, and the genetic map of the <i>S_I</i> locus between <i>indica</i> and <i>O. glaberrima</i>	154
Supplemental Figure IV-4. Frequencies of the Ka/Ks values in the <i>S_I</i> regions.....	155
Supplemental Figure IV-5. Nucleotide sequence alignment of F-box genes in <i>S_I</i> and <i>S_IA</i> loci.....	156
Supplemental Figure IV-6. Protein sequence alignment of F-box genes in <i>S_I</i> and <i>S_IA</i> loci..	158
Supplemental Figure IV-7. Phylogenetic tree of the F-box genes located in the <i>S_IA</i> - <i>S_I</i> loci region.....	159
Figure V-1. Carte physique de la région du locus <i>S_I</i> chez <i>O. glaberrima</i> cv. TOG5681	182
Figure V-2. Comparaison graphique entre les contigs de CG14 et TOG5681	184
Figure V-3. Evaluation de la structure de la région <i>S_IA</i> - <i>S_I</i> - <i>S_IB</i> dans une collection d'accessions <i>O. glaberrima</i>	188
Figure V-4. Structure populationnelle des accessions d' <i>O. glaberrima</i> et <i>O. barthii</i> trouvé par l'évaluation des marqueurs d'éléments transposables.	196
Figure VI-1. Co-localisation des différents locus de barrières reproductives chez le riz.	204
Figure VI-2. Stratégie de cartographie fine des locus <i>S_IA</i> et <i>S_IB</i>	222

LISTE DES TABLEAUX

Tableau I-1. Surfaces de riz cultivées dans le monde entre 2004 et 2006.	18
Tableau I-2. Production, rendement et consommation mondiale de riz en 2007.....	20
Tableau I-3. Principales contraintes de production du riz selon les régions productrices.	24
Tableau I-4. Liste et caractéristiques des espèces du genre <i>Oryza</i>	30
Tableau I-5. Croisements interspécifiques entre <i>O. sativa</i> et les espèces du genre <i>Oryza</i> hors du groupe de génome AA.....	34
Tableau I-6. Pourcentage de fertilité paniculaire des hybrides entre espèces du genre <i>Oryza</i> de génome AA.....	46
Tableau I-7. Gènes connus comme causant une létalité hybride	48
Tableau I-8. Gènes connus comme causants de la stérilité hybride.....	50
Tableau I-9. Gènes impliqués dans les barrières reproductives post-zygotiques décrites chez le riz.	56
Tableau I-10. Croisements interspécifiques du projet iBridges	62
Table II-1. P_3 probability that all N_a anchors are polymorphic, given n markers per anchor and P_1 the observed probability that a marker is polymorphic	74
Table II-2. List of the AA-genome <i>Oryza spp.</i> accessions used for evaluation of SSR polymorphism.....	76
Table II-3. Percentages of successful amplification in the cultivated and wild species, calculated on the 489 SSRs used in this study	82
Table II-4. Comparisons of physical and genetic maps sizes in three BC ₁ F ₁ interspecific populations	82
Table III-1. Frequency of S_1^s alleles in the <i>O. sativa</i> x <i>O. glaberrima</i> BC ₁ F ₁ populations. ..	102
Table III-2. Sequence annotation of the <i>O. glaberrima</i> cv. CG14, BAC OG-BBa0049I08. .	104
Supplemental Table III-1. SSR markers and PCR conditions used for saturation of chromosome 6	122
Supplemental Table III-2. New SSR and InDels markers in the S1 locus.....	123
Supplemental Table III-3 List of identified putative TEs in the <i>O. glaberrima</i> BAC OG- BBA0049I08.	124
Supplemental Table III-4. Sequence comparison between orthologous coding sequences in the <i>O. glaberrima</i> cv. CG14 BAC OG-BBa0049I08 and <i>O. sativa</i> cv. Nipponbare	125
Table IV-1. List of identified gene families in the <i>O. glaberrima</i> S_1 regions	140

Supplemental Table IV-1. New molecular markers located in the S_I region.....	160
Supplemental Table IV-2. List of identified genes in the 813-kb of the <i>O. glaberrima</i> var. CG14 S_I region.....	161
Supplemental Table IV-3. List of the different types of TE found in the <i>O. glaberrima</i> S_I region.....	166
Supplemental Table IV-4. Evaluation of the insertion time of intact LTR retrotransposons in the <i>O. glaberrima</i> and <i>O. sativa</i> S_I region.	166
Supplemental Table IV-5. List of identified pack-MULEs and enclosed genes in the <i>O. glaberrima</i> cv. CG14 S_I region.	167
Supplemental Table IV-6. List and size contribution of TE found differentially inserted between the <i>O. glaberrima</i> and <i>O. sativa</i> S_I regions.	167
Supplemental Table IV-7. Sequence comparison between orthologous coding sequences in the <i>O. glaberrima</i> cv. CG14 and <i>O. sativa</i> cv. Nipponbare S_I regions.....	168
Tableau V-1. Marqueurs moléculaires utilisés pour l'évaluation de la structure de la région des locus $S_I A - S_I - S_I B$	176
Tableau V-2. Accessions évaluées pour la région de $S_I A - S_I - S_I B$	178
Tableau V-3. Pourcentage d'identité des séquences protéiques et nucléotidiques entre les gènes de CG14 et TOG5681.	186
Tableau V-4. Analyse préliminaire des caractères morphologiques des formes intermédiaires entre <i>O. glaberrima</i> et <i>O. sativa</i>	190
Tableau V-5. Données de transmission de l'allèle S_I^s et de fertilité à partir des croisements du projet iBridges.	192

ABBREVIATIONS

BAC :	<i>Bacterial Artificial Chromosome</i>
BDM :	Bateson-Dobzhansky-Muller
CSSL :	<i>Chromosome Segment Substitution Lines</i> , Lignées de substitution de segments chromosomiques
DS :	Distorsion de la Ségrégation
fTRD :	<i>female Transmission Ratio Distortion</i>
IRMI :	International Rice Microsatellite Initiative
ISSR :	<i>Inter Simple Sequence Repeat</i>
LEA :	<i>Leaf Area Index</i>
LNA :	<i>Locked Nucleic Acid</i>
LRR :	<i>Leucine - Rich Repeat</i>
LTR :	<i>Long Terminal Repeat</i>
MAAL :	<i>Monosomic Alien Addition Lines</i>
MITE :	<i>Miniature Inverted Repeat Transposable elements</i>
MTP :	<i>Minimum Tiling Path</i> , Chemin Minimal de recouvrement
mTRD :	<i>male Transmission Ratio Distortion</i>
NERICA :	New RICE pour Africa
NIL :	<i>Near-Isogenic Lines</i>
PCR :	<i>Polymerase Chain Reaction</i>
PIC :	<i>Polymorphism Information Content</i>
QTL :	<i>Quantitative Trait Locus</i>
RFLP :	<i>Restriction Fragment Length Polymorphism</i>
RPS :	<i>Rate of polymorphic SSRs</i>
RTD :	<i>Rice Tungro Disease</i>
SNP :	<i>Single-Nucleotide Polymorphism</i>
SSR :	<i>Simple Sequence Repeats</i>
TE :	<i>Transposable Elements</i>
TRD :	<i>Transmission Ratio Distortion</i>
UCGM :	Universal Core Genetic Map, Carte génétique universelle du riz
WCV :	<i>Wide Compatibility Variety</i>

I

INTRODUCTION

I.1 Positionnement général

Les travaux présentés dans ce document portent sur la génétique des barrières de reproduction chez le riz, et plus précisément sur l'étude de la barrière de stérilité entre les deux espèces de riz cultivées, *Oryza sativa* et *Oryza glaberrima*, sur les conséquences de cette barrière reproductive, sur les stratégies d'amélioration du riz et la divergence évolutive de ces deux espèces.

Ces travaux ont été réalisés en alternance entre l'Unité de Biotechnologie du CIAT, Centro Internacional de Agricultura Tropical, à Cali, Colombie, et le laboratoire de Génétique et Génomique du Riz de l'IRD, Institut de Recherche pour le Développement, Montpellier, France, sous la direction d'Alain Ghesquière (DR-IRD), et la co-direction de Mathias Lorieux (CR-IRD).

Pour la réalisation de ce travail de thèse j'ai été bénéficiaire d'une bourse de thèse, attribuée par le Département de Soutien et Formation (DSF) de l'IRD et co-financée par le CIAT. Ces travaux ont été rendus possibles grâce au soutien du projet iBridges, financé par le Generation Challenge Program.

Tableau I-1. Surfaces de riz cultivées dans le monde entre 2004 et 2006.

Continent	Millions d'hectares cultivés	Pourcentage de surface selon le type de culture			
		Irrigué	Pluvial de basse altitude	Pluvial de montagne	Eau profonde (flottant)
Asie	135,03	58,6	32,1	6,7	2,6
Chine	29,04	93,0	5,0	2,0	0
Inde	43,09	52,6	32,4	12,0	3,0
Indonésie	11,71	60,1	25,3	0	14,6
Bangladesh	10,66	40,0	42,0	7,0	11,0
Autres pays	40,53	46,4	47,7	3,2	2,7
Afrique	7,79	22,9	35,4	32,6	9,1
Afrique Sub-Saharienne	7,17	18,5	35,3	40,5	5,7
Afrique du Nord	0,62	100,0	0	0	0
Amérique	6,67	58,2	4,0	37,7	0
Amérique du Sud	5,12	45,8	5,1	49,2	0
Amérique du Nord	1,28	100,0	0	0	0
Amérique Centrale	0,27	96,7	3,3	0	0
Europe	0,58	100,0	0	0	0
Océanie	0,05	100,0	0	0	0
Total mondial	150,11	56,9	30,9	9,4	2,8

Source : IRRI-World Rice Statistics (<http://beta.irri.org/statistics>)

I.2 Le riz et son importance dans le monde

I.2.1 Les différents types de culture du riz

Le riz se cultive dans différentes conditions de régime hydrique, principalement en écosystèmes aquatiques, mais aussi en écosystèmes plus secs. Le type de culture dépend fortement de la source d'eau dont on dispose. La culture dite inondée ou irriguée utilise l'eau des rivières ou des puits, et se caractérise par une inondation constante qui permet plusieurs semis par an. Pour la culture pluviale à basse altitude, ou culture pluviale favorisée, des barrages sont construits pour capturer l'eau qui sera ensuite utilisée dans les rizières. Les champs peuvent alors être inondés pendant une partie de la durée de la culture, mais sont aussi susceptibles de subir de fortes variations hydriques (inondations et sécheresses). Il existe aussi une culture pluviale de montagne ou pluviale stricte, dans des écosystèmes à terrains inclinés, où l'eau de pluie n'est pas retenue. Le dernier type d'écologie pour la culture de riz est en zones d'eaux profondes d'un mètre ou plus, où les plantes peuvent avoir une croissance pratiquement flottante (Maclean *et al.*, 2002). Entre 2004 et 2006, 150 millions de hectares de riz ont été cultivés dans le monde (voir Tableau I-1). Près de 57% de cette surface est destiné à la culture irriguée, 31% à la culture pluviale de basse altitude, tandis que seulement 9% et 3% sont destinés respectivement à la culture en milieu pluvial de montagne et en milieux d'eaux profondes. Plus des 90% de la surface cultivée en riz se trouvent en Asie.

I.2.2 Production de riz dans le monde

Domestiqué depuis au moins 7 000 ans (Fuller *et al.*, 2009), pour des raisons économiques, nutritionnelles et culturelles le riz peut être considéré comme le produit agricole le plus important au monde.

Selon les derniers données de la FAO sur la production (FaoStat ; <http://faostat.fao.org/>), le riz est le premier produit agricole de consommation humaine, avec 659,5 millions de tonnes en 2007 (voir Tableau I-2 dans la page suivante). Sa production est seulement dépassée par la canne à sucre (1590,7 millions de tonnes) et le maïs (791,7 millions de tonnes) qui sont utilisés principalement dans l'industrie de biocombustibles et dans l'alimentation animale. Le principal pays producteur de riz est la Chine, avec près du 29% de la production mondiale, suivi par l'Inde (22%). Au niveau de continents, l'Asie produit à elle seule, 91% du riz dans le monde, très loin devant l'Amérique (5%) et l'Afrique (3%). Les rendements varient fortement selon les régions de production. Les plus haut rendements sont obtenus en Afrique du Nord, plus précisément en Égypte, avec 9,9 tonnes par hectare semé. Puis viennent l'Australie (8,1 t/ha), la Turquie (8,0 t/ha), les Etats Unis qui sont responsables de toute la

Tableau I-2. Production, rendement et consommation mondiale de riz en 2007.

Continent	Millions de tonnes de riz paddy produites	Rendement (t/ha)	Millions de tonnes consommées
Asie	600,9	4,2	378,1
Chine	187,4	6,3	128,9
Inde	144,6	3,2	90,4
Indonésie	57,2	4,6	36,3
Bangladesh	43,1	3,8	30,7
Autres pays	168,6	3,8	91,8
Amérique	33,6	5,3	22,4
Amérique du Sud	22,2	4,4	14,3
Amérique du Nord	9,0	8,0	4,4
Amérique Centrale	2,4	3,9	3,7
Afrique	21,3	2,5	21,0
Afrique Sub-Saharienne	14,4	1,7	17,7
Afrique du Nord	6,9	9,9	3,3
Europe	3,6	5,7	5,7
Océanie	0,2	6,7	5,3
Total mondial	659,6	4,1	426,4

Source : FaoStat (<http://faostat.fao.org/>), et IRRI-World Rice Statistics (<http://beta.irri.org/statistics>)

production de l'Amérique du Nord (8,0 t/ha), et la Grèce (7,8 t/ha). Un résumé des rendements de la production de riz par région géographique est présenté dans le Tableau I-2.

I.2.3 Importance nutritionnelle

Le riz est sans doute la culture la plus importante pour la nutrition humaine. Il représente environ 20% de la consommation en protéines au niveau mondial, et peut représenter près du 71% dans certains pays d'Asie. En moyenne, le riz équivaut à 30% de la consommation protéique en Asie, 12% en Amérique latine, 8% en Afrique, et moins de 3% dans le reste du monde. Plus de 426 millions de tonnes de riz ont été consommées en 2007 (voir Tableau I-2). Les régions avec le plus fort déficit entre production et consommation sont l'Océanie, qui importe 96% de son riz, l'Europe (37%), l'Amérique Centrale (35%), et l'Afrique subsaharienne (18%), ce qui fait de ces régions des importateurs majeurs.

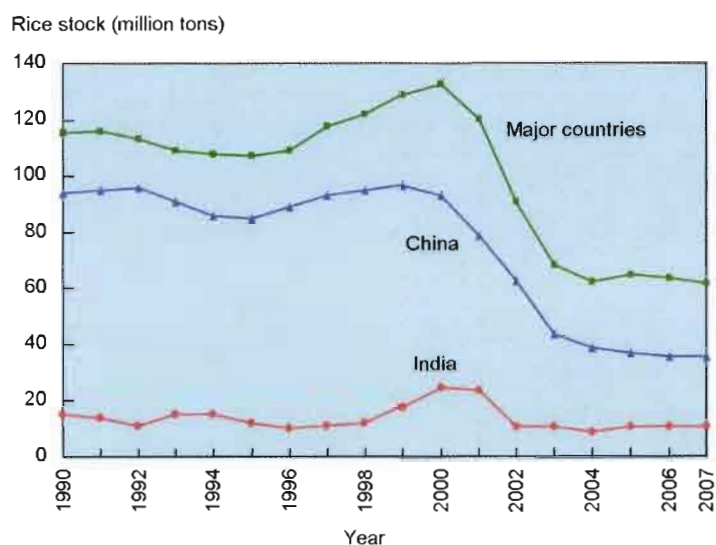
I.2.4 Importance culturelle

Le riz a joué un rôle décisif dans la culture de l'humanité depuis sa domestication il y a environ 10 000 ans. Aucun aliment n'a nourri plus de personnes dans le monde et dans toute l'histoire que le riz. La culture du riz a eu une forte influence sur les habitudes et les croyances religieuses chez tous les peuples qui dépendent de lui. Il a été considéré comme une divinité par de nombreux peuples, et associé au sentiment de bien-être tout au long de l'histoire de l'humanité, au point de faire partie de nombreuses cérémonies qui persistent encore de nos jours (Huggan, 1995). Une des traditions culturelles associées au riz les plus largement disséminées dans le monde est l'acte de lancer du riz aux mariés comme signe de vœux de prospérité et bonheur. Cette association du riz avec la prospérité se voit encore reflétée dans la vie quotidienne. Au Japon, par exemple, les noms « Toyota » et « Honda » sont dérivés de caractères chinois qui signifient respectivement « champ de riz abondant » et « champ de riz principal ».

I.2.5 Importance économique

Malgré la faible proportion de la production du riz destinée à la commercialisation dans le monde (4,5%), le riz est le premier produit agricole en importance économique. Approximativement 30 millions de tonnes de riz furent commercialisées en 2007, équivalents à un marché d'environ 11 milliards de dollars (FaoStat). La Thaïlande est le plus grand exportateur de riz, avec 9,3 millions de tonnes ; loin derrière se trouve le Viet Nam avec près de 4,5 millions de tonnes, suivi par l'Inde, les Etats Unis, le Pakistan et la Chine. L'Asie est le plus gros importateur, avec près de 15 millions de tonnes, suivi par l'Afrique (9,3 millions), et l'Amérique latine (3,5 millions).

A



B

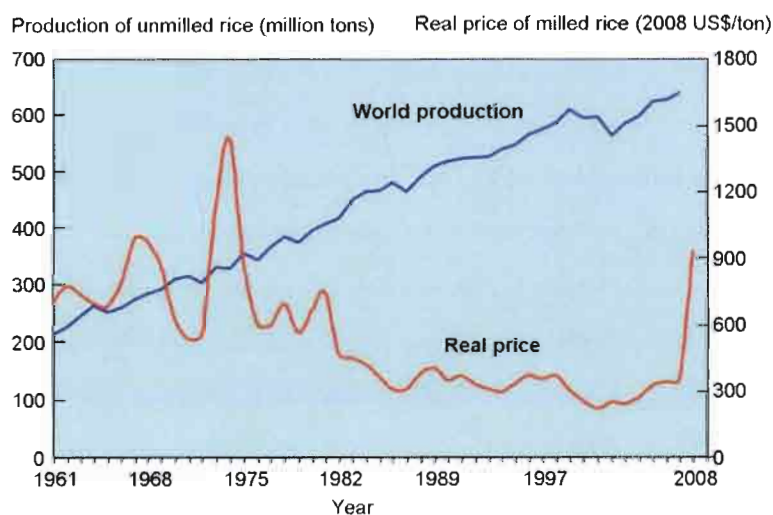


Figure I-1. Production mondiale du riz et réserves mondiales. A. Réserves mondiales du riz entre 1990 et 2007. B. Production mondiale du riz et son prix réel entre 1961 et 2008.
Source : IRRI-World Rice Statistics (<http://beta.irri.org/statistics>)

Mais l'importance économique du riz ne se restreint pas à l'activité commerciale. Tous les processus associés à la production de riz ont besoin d'une importante main d'œuvre. Le travail de centaines de millions de personnes est associé à la riziculture, non seulement en ce qui concerne la production au sens strict, mais aussi pour le traitement post-récolte, comme le battage, l'usinage, le stockage, la distribution et la commercialisation. Des millions de familles en milieu rural tirent une grande partie de leurs revenus de la culture du riz.

Le riz est un produit essentiel pour la sécurité alimentaire dans le monde. Du fait de la faible croissance annuelle de la production de riz (moins de 1% par an), les stocks de riz ont commencé à diminuer de manière significative (voir Figure I-1 A), car la consommation croît à un taux bien plus rapide que la production. Pendant les premiers mois de 2008, le prix du riz a été multiplié par trois (Figure I-1 B), donnant lieu à une forte crise alimentaire. Plusieurs facteurs ont été à l'origine de cette forte augmentation, (1) l'effondrement de la valeur du dollar, (2) les forts prix du pétrole et de ses dérivés, surtout les agrochimiques, (3) la forte restriction des exportations de la part des pays producteurs qui vise à garantir les stocks destinés à la consommation interne, (4) une forte demande des grands importateurs générée par un mouvement de panique dû à la flambée des prix (5) des événements climatiques qui ont touché plusieurs zones productrices de riz (Childs et Kiawu, 2009). Malgré la baisse récente des prix du riz, ils n'ont pas retrouvé les niveaux antérieurs à la crise, indiquant qu'il existe encore un fort déficit dans la production par rapport à la demande mondiale.

I.3 Les principales contraintes de la production du riz

La crise des prix du riz au début 2008 a mis en évidence le fragile équilibre entre la production de riz et sa demande. Des mesures doivent être impérativement prises afin de limiter les risques qu'une situation similaire se présente à nouveau. Il est entre autres nécessaire d'augmenter les rendements d'une manière significative, sans augmenter l'utilisation des ressources, tout en tenant compte des fortes contraintes liées à la riziculture (Zeigler et Barclay, 2008).

Le panorama est vraiment inquiétant, surtout si l'on considère que selon le programme alimentaire mondial des Nations Unies, un sixième de la population mondiale se trouve dans un état de sous-alimentation (FAO, 2009), et que la population mondiale devrait atteindre les 11 milliards vers 2050 si le rythme actuel de croissance est maintenu (<http://www.un.org/esa/population/unpop.htm>). Un tel incrément est le plus grand défi pour la riziculture.

Tableau I-3. Principales contraintes de production du riz selon les régions productrices.

Région	Contrainte
Asie	Facteurs abiotiques Sécheresse Maladies infectieuses Pyriculariose Bactériose vasculaire Maladie de Tungro Fonte des semis Parasites et compétiteurs Foreur jaune de la tige (<i>Scirpophaga incertulas</i> Walker) Galle de la tige (<i>Orseolia oryzae</i> Wood & Mason) Cicadelle marron (<i>Nilaparvata lugens</i> Stal) Mauvaises herbes : ex. <i>Echinochloa crus-galli</i> L. et le riz rouge (Riz sauvage)
Amérique Latine	Facteurs abiotiques Sécheresse Maladies infectieuses Pyriculariose Fonte des semis Virus de la Hoja Blanca Pourriture des gaines Parasites et compétiteurs Delphacide : <i>Tagosodes orizicolus</i> Muir (Sogata) Mauvaises herbes : ex. <i>Echinochloa colona</i> L., <i>E. crus-galli</i> L. et le riz rouge (Riz sauvage)
Afrique	Facteurs abiotiques Sécheresse Hautes concentrations en Fer Acidité des sols Maladies infectieuses Virus de la panachure jaune Pyriculariose Bactériose vasculaire Parasites et compétiteurs Mauvaises herbes : ex. <i>E. crus-galli</i> et le riz rouge Foreur Africain de la tige : ex. <i>Chilo zacconius</i> Bleszynski Galle Africain de la tige (<i>Orseolia oryzivora</i> Harris & Gagne)

D'après Pathak et Khan, 1994; Carbonell *et al.*, 2001; Jones *et al.*, 2001; Maclean *et al.*, 2002; Calvert *et al.*, 2006; Nwanze *et al.*, 2006; Zeigler et Barclay, 2008

Il est donc urgent d'assurer une croissance soutenue de la production mondiale de riz. Celle-ci ne pas être basée sur l'augmentation de la surface cultivée, dont la disponibilité a tendance à diminuer du fait de la croissance démographique qui empiète sur ces surfaces et représente une forte concurrence pour l'accès aux ressources hydriques.

Le changement climatique représente aussi un important défi. La tendance à la hausse de la température moyenne et les changements dans les patrons des précipitations menacent les ressources hydriques nécessaires pour la culture du riz, et augmentent les risques de sécheresse des écosystèmes qui dépendent de l'eau de pluie. En outre, l'élévation du niveau des eaux marines et les diverses catastrophes climatiques peuvent causer une diminution supplémentaire de la surface cultivable, en provoquant une augmentation de la salinité des terres cultivables. Certaines terres où le riz est cultivé montrent aussi des concentrations très élevées en métaux lourds et des pH très acides ; souvent, les paysans n'ont pas d'autre choix que de continuer à cultiver dans ces conditions. Enfin, la concurrence pour la surface avec d'autres cultures, l'élevage des animaux, et l'industrie des biocombustibles menacent la surface disponible pour la riziculture (Nguyen et Ferrero, 2006).

Mais les contraintes liées à la population humaine et aux facteurs abiotiques ne sont pas les seules à exercer une pression sur la riziculture : un ensemble de facteurs biotiques causent de fortes diminutions des rendements, et des mesures doivent être prises pour réduire leur impact. Ces facteurs incluent les maladies infectieuses, les parasites et les compétiteurs du riz. Un résumé des principales contraintes biotiques et abiotiques dans les régions les plus importantes pour la production de riz est présenté dans le Tableau I-3.

I.3.1 Maladies infectieuses

I.3.1.1 Champignons

Les maladies dites Cryptogamiques, ou causées par des champignons, peuvent se manifester sur les feuilles et la tige (Gnanamanickam, 2009). Parmi les principales maladies, on peut énumérer :

- Les maladies des feuilles : Les principales sont la pyriculariose (Rice blast), causée par *Magnaporthe oryzae* B. Couch sp. Nov. (ou son anamorphe *Pyricularia grisea* Sacc), l'helminthosporiose (Brown leaf spot), causée par *Bipolaris oryzae* Breda de Haan (téléomorphe *Cochliobolus miyabeanus* S. Ito & Kurib.), et la « fonte des semis » (Sheath blight) causée par *Rhizoctonia solani* Kühn (téléomorphe *Thanatephorus cucumeris* Frank).
- Les maladies des tiges : plusieurs champignons attaquent les tiges, gaines et racines en produisant une pourriture qui peut a posteriori enrayer le développement complet des

panicules, causant des pertes importantes. La maladie la plus importante de ce type est la pourriture des gaines (Sheath rot) causée par *Sarocladium oryzae* Sawada.

I.3.1.2 Bactéries

Deux maladies bactériennes ont un très fort impact sur l'état physiologique de la plante et sur son rendement :

- La bactériose vasculaire (Bacterial blight) causée par *Xanthomonas oryzae* pv. *oryzae* est une maladie dont leurs effets les plus sévères se manifestent à l'étape de formation des graines. Elle est présente dans toutes les zones de riziculture, et cause des sérieux dégâts dans la production, selon le stade de croissance de la plante. Elle peut être cataloguée comme la maladie la plus grave pour la culture du riz (Gnanamanickam, 2009).
- La bactériose à stries foliaires (Bacterial leaf streak), dont l'agent pathogène est *Xanthomonas oryzae* pv. *oryzicola*. La bactérie attaque les tissus conducteurs de la plante, et se caractérise par des traces longitudinales transparentes sur les feuilles avec des exsudats.

I.3.1.3 Virus

Les maladies virales du riz sont souvent propagées par des insectes vecteurs, qui constituent dans certains cas l'hôte d'une étape du cycle de multiplication du virus. Leur couverture géographique est donc dépendante de la distribution de ces vecteurs (principalement des cicadelles). Les maladies virales plus importants du riz sont :

- Le Tungro (Rice Tungro Disease - RTD) causé par le complexe entre le RTBV (*Rice Tungro Bacilliform Virus*), un *Caulimovirus* (dsDNA), et le RTSV (*Rice Tungro Spherical Virus*), un Waikavirus (ssRNA+). C'est la maladie virale la plus importante en Asie. Les virus du Tungro ont comme vecteur la cicadelle verte (*Nephotettix virescens* Distant). Les symptômes incluent une décoloration des feuilles, un tallage diminué, une floraison tardive et une stérilité paniculaire (Gnanamanickam, 2009).
- La panachure jaune, ou RYMV (*Rice Yellow Mottle Virus*). Le RYMV est un *Sobemovirus* (ssRNA+). Il a une distribution endémique en Afrique, où il cause des pertes importantes dans la production. Ses vecteurs sont, entre autres, les scarabées chrysomélidés mais ce virus peut aussi se transmettre mécaniquement. Après infection, les feuilles acquièrent des taches jaunes puis marrons, associées à une réduction du tallage et de la taille de la plante (Kouassi *et al.*, 2005).
- Le virus de la feuille blanche RHBV (*Rice Hoja Blanca Virus*) est un *Tenuivirus* (ssRNA en ambisens) dont la distribution se limite aux zones tropicales et subtropicales de l'Amérique. Son vecteur est le delphacide *Tagosodes orizicolus* Muir. Les plantes infectées présentent des taches longitudinales chlorotiques qui donnent une apparence de blancheur aux

feuilles. La production de graines est sévèrement affectée dans les variétés sensibles (Garavito, 2003).

I.3.1.4 Parasites et compétiteurs

I.3.1.5 Nématodes

Les nématodes peuvent attaquer les parties aériennes ou les racines des plantes. Ils se nourrissent en suçant la sève, et peuvent dans certains cas coloniser l'intérieur de la plante. Parmi ceux qui attaquent les feuilles, se trouvent le *Ditylenchus angustus* en rizières d'eaux profondes, et le nématode *Aphelenchoides besseyi* Christie dont la distribution est indépendante de l'écologie de la plante. Les nématodes des racines se classent en ectoparasites et endoparasites. Les espèces de nématodes endoparasites dépendent de l'écologie de la culture. En rizières pluviales, il est courant de trouver différentes espèces de *Meloidogyne* (nématodes à galles) ou le nématode à kystes *Heterodera sacchari*, tandis qu'en conditions irriguées on rencontre plutôt l'espèce *Meloidogyne graminicola* Golden & Birchfield (Prot et Rahman, 1994).

I.3.1.6 Insectes

Les insectes peuvent attaquer toutes les parties de la plante, et causer des dégâts très sérieux qui compromettent la capacité productive de la culture attaquée (Pathak et Khan, 1994). Les attaques peuvent survenir durant la phase végétative, reproductive ou de maturation des graines. Des exemples de ces ravageurs sont :

- Ravageurs de racines : les termites, les fourmis et les larves de coléoptères peuvent réduire significativement le rendement en affaiblissant le développement racinaire, surtout en écologie pluvial de montagne.
- Foreurs de la tige (Stem borers) : les larves de certaines lépidoptères perforent et se nourrissent de la tige, empêchant la formation des panicules. Les pertes causées par ce type d'insecte comptent parmi les plus importantes.
- Galles des tiges : elles sont causées par les petites mouches qui transforment les tiges en galles tubulaires en forme d'oignon.
- Cicadellidés (leafhoppers) : ce sont des insectes volants, dont les nymphes et les adultes sucent la sève du xylème et du phloème des parties aériennes des plantes. Ils peuvent aussi transmettre des maladies virales.
- Delphacidés (planthoppers) : comme les Cicadellidés, ces insectes se nourrissent de la sève, mais obtenue des tiges.
- Ravageurs de feuilles : les dégâts causés par ce type d'insecte sont dus à leur impact sur la surface et capacité photosynthétique de la plante. Dans ce groupe, se trouvent les

Thrips, certaines larves et adultes de coléoptères, les larves de certaines petites mouches, plusieurs types de poux et chenilles de lépidoptères.

- Ravageurs des graines : ces insectes sucent le contenu des graines en formation, ce qui en empêche le remplissage et favorise les infections par divers bactéries et champignons.

I.3.1.7 Adventices

Les mauvaises herbes ou adventices sont une des causes biologiques les plus importantes des pertes de rendement. Ces plantes, de par leur croissance rapide, leur capacité à rester en latence dans les champs grâce à la dormance de leurs graines, et à coloniser rapidement les espaces, entrent en concurrence directe avec le riz pour l'accès aux nutriments, à l'eau et à la lumière (Ampong-Nyarko et Datta, 1991). En outre, elles peuvent servir comme réservoir d'insectes, de virus et de nématodes. Les graines de certains riz adventices peuvent être récoltés avec le riz, ce qui réduit notablement par contamination, la qualité finale des graines.

I.4 Comment faire face aux défis de la riziculture ?

Pour que la riziculture puisse faire face aux défis énoncés plus haut, il est essentiel de trouver les moyens de colmater la brèche entre le potentiel de rendement et les rendements réels obtenus.

Une des mesures à prendre est l'adoption systématique de pratiques de gestion intégrée de la culture. Ces pratiques comprennent le développement et l'adéquation de techniques pour la production, pour chacune des étapes de la culture, et de manière spécifique pour chaque écologie. Les recommandations incluent des pratiques de gestion avant semis, pendant la croissance de la plante, et post-récolte, afin de garantir un rendement maximum tout en utilisant un minimum de ressources (Tran et Nguyen, 2006). La gestion intégrée de la culture implique la gestion des maladies et ravageurs, des mauvaises herbes, et des nutriments, ainsi qu'une gestion appropriée des ressources hydriques.

La gestion intégrée des maladies recherche le contrôle des populations de pathogènes par la connaissance de leurs écologies et de leur cycle de vie, avec comme objectif de réduire leurs opportunités d'infection et de colonisation, et de favoriser leurs ennemis naturels. Cette gestion implique aussi une évaluation périodique des niveaux d'infestation pour établir les seuils où un traitement doit être effectué (Razdan et Sabitha, 2009). La gestion des adventices implique quant à elle une éradication manuelle, le contrôle des niveaux d'eau, et, en dernier recours, l'utilisation d'herbicides. La gestion des nutriments se base sur l'application adéquate d'engrais, en quantité et en temps appropriés au regard de l'état physiologique de la culture. Par une bonne gestion intégrée de la culture appropriée, il est possible de réduire les coûts, tout en augmentant la production.

I.5 La réponse génétique : l'amélioration variétale

Malgré l'utilité et le succès de la gestion intégrée de la culture, celle-ci reste limitée, notamment par l'adaptation des variétés à l'écologie de semis. La gestion intégrée est donc complémentée, et dépend de la disponibilité de cultivars adaptés. Ainsi, le développement de variétés tolérantes aux stress abiotiques (en particulier la sécheresse), résistantes aux maladies, compétitives contre les mauvaises herbes, et à haut potentiel de production et rendement, sans négliger la qualité du grain, doit être une priorité pour la riziculture.

Le riz cultivé correspond majoritairement à l'espèce asiatique *Oryza sativa* L., dont la culture fut disséminée par l'homme tout autour du globe. Il existe aussi une autre espèce de riz cultivé, l'espèce africaine *Oryza glaberrima* Steud, dont la culture se limite à certaines zones de ce continent et représente moins de 2% de la surface cultivée dans ces régions. Les variétés de ces deux espèces ont été sélectionnées par l'homme depuis leur domestication pour s'adapter aux conditions écologiques de chaque région de production. Les méthodes de sélection ont évolué, commençant par une sélection et multiplication des plantes présentant les meilleurs rendements, puis utilisant les croisements et la sélection variétale « traditionnelle », enfin exploitant les avancées de la génétique et de la génomique pour développer des programmes de sélection assisté par marqueurs et de transformation génétique.

La sélection variétale a permis de réaliser des sauts quantitatifs conséquents dans la production des céréales. La révolution verte, notamment, a permis de créer des variétés hautement productives, présentant une meilleure réponse aux engrais, un cycle de vie plus court, un tallage plus élevé, une hauteur de plate inférieure, et une meilleure capacité de photosynthèse.

Par ailleurs, les deux sous-espèces d'*O. sativa*, *indica* et *japonica*, présentent une large diversité génétique, et les croisements entre ces deux groupes permettent d'obtenir des hybrides offrant une performance agronomique supérieure. Ce phénomène d'hétérosis ou vigueur hybride est d'un intérêt vital pour l'amélioration variétale.

I.5.1 Hybridations interspécifiques versus les hybridations intraspécifiques

Le goulot d'étranglement causé par la domestication du riz (effet de population fondatrice), et la sélection pour des traits positifs (connue comme balayage sélectif), réduisirent la variabilité génétique dans les régions génomiques où se trouvent ces caractères (Zhu *et al.*, 2007). Il en résulte une diminution du nombre d'allèles disponibles aux gènes d'intérêt, ce qui a pour effet de restreindre le potentiel de l'amélioration variétale basée sur des croisements intraspécifiques (Brar et Khush, 1997).

Tableau I-4. Liste et caractéristiques des espèces du genre *Oryza*

Complexe	Habitats	Cycle de vie	Espèces	Génome	Distribution
<i>O. sativa</i>	Rivières, lacs et sols inondés ou pluviaux	Pérenne et annuel	<i>O. sativa</i> L.	AA	Mondiale
			<i>O. rufipogon</i> Griff	AA	Asie tropicale et subtropicale, Australie tropicale
			<i>O. nivara</i> Sharma et Shastry	AA	Asie tropicale et subtropicale
			<i>O. glaberrima</i>	AA	Afrique de l'est
			<i>O. barthii</i> Chev	AA	Afrique
			<i>O. glumaepatula</i> Steud	AA	Amérique centrale et du sud
			<i>O. longistaminata</i> Chev et Roehr	AA	Afrique
			<i>O. meridionalis</i> Ng	AA	Australie tropical
<i>O. officinalis</i>	Savanes et forêts de plaine, inondées ou pluviales	Pérenne	<i>O. punctata</i> Kotschy ex Steud	BB	Afrique
			<i>O. officinalis</i> Wall ex Watt	CC	Asie tropicale et subtropicale, Australie tropicale
			<i>O. rhizomatis</i> Vaughan	CC	Sri Lanka
			<i>O. australiensis</i> Domin	EE	Australie tropicale
			<i>O. minuta</i> Presl	BBCC	Philippines, Thaïlande, Papouasie-Nouvelle Guinée
			<i>O. alta</i> Swallen	CCDD	Amérique centrale et du sud
			<i>O. grandiglumis</i> (Doll) Prodoehl	CCDD	Amérique centrale et du sud
			<i>O. latifolia</i> Desv	CCDD	Amérique centrale et du sud
			<i>O. eichingeri</i> Peter	CC/ BBCC	Asie du sud et Afrique de l'est
<i>O. ridleyi</i>	Forêts de marécage, en plaines	Pérenne	<i>O. ridleyi</i> Hook	HHJJ	Asie du sud
			<i>O. longiglumis</i> Jasen	HHJJ	Indonésie, Papouasie-Nouvelle Guinée
<i>O. granulata</i>	Forêts d'altitude non inondées	Pérenne	<i>O. meyeriana</i> (Zoll et Moritzi) Baill	GG	Asie du sud et l'est
			<i>O. granulata</i> Nees et Arn ex Watt	GG	Asie du sud
Autres espèces	Sols riches en fer	Pérenne	<i>O. brachyantha</i> Chev et Roehr	FF	Afrique
	Forêts, près des rivières		<i>O. schlechteri</i> Pilger	HHKK	Papouasie-Nouvelle Guinée
	En bord de rivières		<i>O. coarctata</i> (Roxb) Tateoka /	HHKK	Inde
			<i>Porteresia coarctata</i>		

D'après les informations de Vaughan *et al.*, 2008 ; et Ge *et al.*, 2001.

Les espèces sauvages apparentées au riz représentent une source potentielle d'allèles nouveaux pour des caractères d'intérêt. On peut expliquer la sous-utilisation de ces espèces sauvages en amélioration des plantes par le fait qu'elles possèdent plus souvent –du moins en apparence– des allèles « négatifs » que « positifs » pour des caractères comme le rendement. La détection des allèles supérieurs ou « positifs » chez ces espèces n'est souvent possible qu'après croisement et rétrocroisement avec l'espèce cultivée, ce qui permet d'éliminer les allèles « négatifs ». Avant d'aborder la question des hybridations interspécifiques, nous aborderons brièvement la taxonomie et les relations phylogénétiques entre les espèces du genre *Oryza*.

I.5.1.1 Classification taxonomique du genre *Oryza*

Le genre *Oryza* inclut les deux espèces cultivées, *O. sativa* et *O. glaberrima*, et 22 espèces sauvages connues. Ces espèces ont été regroupées en quatre complexes : *O. sativa*, *O. officinalis*, *O. ridleyi*, et *O. granulata*, selon le type de génome qu'elles possèdent (Vaughan *et al.*, 2008). Le genre a une distribution mondiale, en régions tropicales et subtropicales des cinq continents, avec des adaptations à une grande quantité d'habitats et une prédilection pour les environnements aquatiques (Ge *et al.*, 2001). La liste des espèces, leurs génomes, principaux habitats, et distributions géographiques sont résumés dans le Tableau I-4.

Les relations phylogénétiques entre les espèces du genre *Oryza* ont été d'abord établies par les travaux de G. Second, au moyen d'une caractérisation par marqueurs isoenzymatiques (Second, 1986). Ces résultats ont été complétés par des études sur les introns de deux gènes nucléaires et un gène chloroplastique (Ge *et al.*, 1999), les RFLP mitochondriaux (Abe *et al.*, 1999), par l'analyse d'ISSR (Joshi *et al.*, 2000), et l'analyse des insertions des éléments transposables de type MITE (Transposons) (Park *et al.*, 2003). Un exemple d'arbre phylogénétique construit à partir de ces analyses est présenté dans la Figure I-2 (voir page suivante).

En ce qui concerne les deux espèces cultivées et les six espèces sauvages de génome AA, divers auteurs ont établi leurs relation phylogénétiques en utilisant différents outils, comme les marqueurs RAPD nucléaires (Ishii *et al.*, 1996), microsatellites et RFLP (Bautista *et al.*, 2001), SINE (rétrotransposons) (Cheng *et al.*, 2002), SSR (Ren *et al.*, 2003), séquences d'introns (Zhu et Ge, 2005), « transposon display » (Kwon *et al.*, 2006), et les gènes nucléaires, chloroplastiques et mitochondriaux (Duan *et al.*, 2007). En terme général, et malgré quelques différences entre les arbres phylogénétiques obtenus, il est accepté que les premières espèces du groupe qui ont divergé sont *O. meridionalis* et *O. longistaminata*, il y a environ deux millions d'années (Zhu et Ge, 2005). *O. glumaepatula* s'avère proche de

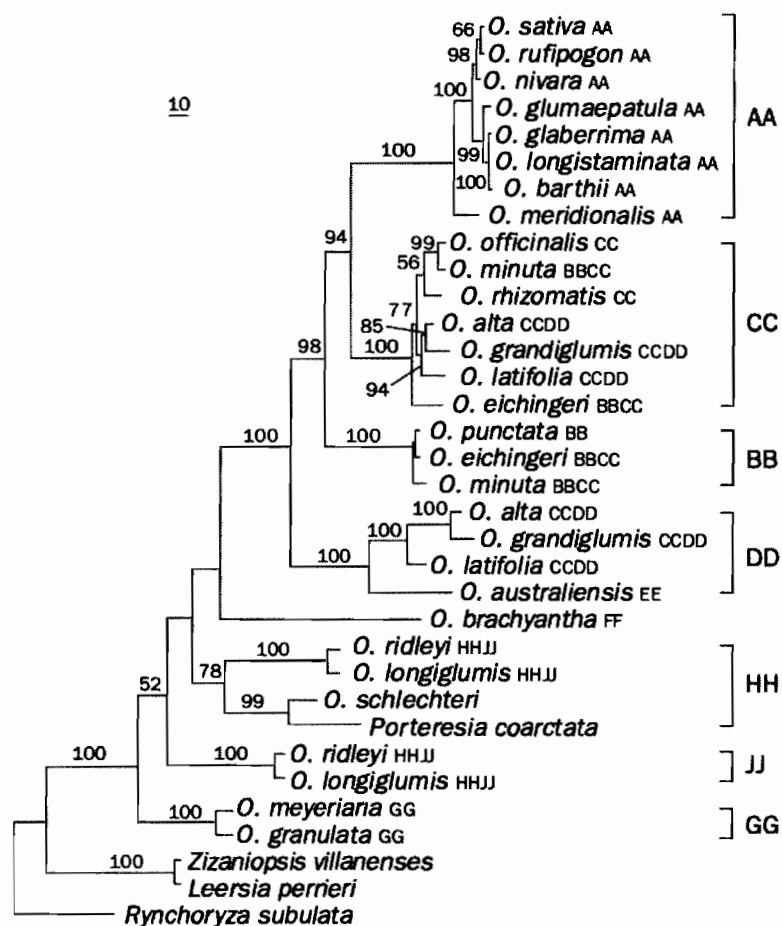


Figure I-2. Relations phylogénétiques entre les espèces du genre *Oryza*. Phylogénie du gène *Adh1*. Le trait horizontal indique un nombre de 10 substitutions. Les valeurs de bootstrap supérieures à 50%. Sont indiquées sur les branches (d'après Ge *et al.*, 1999).

O. barthii et de *O. glaberrima*, tandis que *O. nivara* reste très proche de *O. rufipogon* et *O. sativa*. En outre, la forte relation phylogénétique entre *O. sativa* et *O. rufipogon* d'une part, et *O. glaberrima* et *O. barthii* d'autre part, indique que les deux espèces de riz cultivé ont été domestiquées à partir de ces deux espèces sauvages. La divergence entre *O. sativa* et *O. glaberrima* aurait eu lieu il y a environ 0,67 à 0,7 million d'années (Ma et Bennetzen, 2004; Zhu et Ge, 2005).

La classification et la diversité des espèces cultivées ont également été étudiées en détail. *O. glaberrima* paraît avoir été domestiqué dans le Delta du fleuve Niger à partir de *O. barthii*, il y a environ 2500 à 3000 ans (Portères, 1962; Murray, 2004). Elle est représentée par deux groupes variétaux d'écologies différentes, avec d'une part des variétés de type « pluviales » et d'autre part des accessions adaptées aux zones inondées (Bezançon *et al.*, 1989), avec une diversité limitée (Bezançon *et al.*, 1977; Semon *et al.*, 2005; Barry *et al.*, 2007). *O. sativa* peut être divisée en deux sous-espèces, *indica* et *japonica*. Ces deux sous-espèces semblent avoir été domestiquées à partir de deux populations préalablement différenciées d'*O. rufipogon* (Londo *et al.*, 2006; Xu *et al.*, 2007; Long *et al.*, 2008; Zhao *et al.*, 2009) au Sud de l'Himalaya pour les *indica* et au Nord-Est de la Chine pour les *japonica*. L'époque estimée pour la séparation entre les deux sous-espèces se situe il y a 0,4 à 0,44 million d'années (Ma et Bennetzen, 2004; Zhu et Ge, 2005), longtemps avant leur domestication. Au niveau populationnel, *O. sativa* peut être divisé en cinq groupes variétaux, *indica*, *aus*, *japonica* tempéré, *japonica* tropical, et *aromatiques (basmati)* (Glaszmann, 1987; Garriss *et al.*, 2005).

1.5.2 Le point sur les hybridations interspécifiques chez le riz

Les hybridations entre *O. sativa* et d'autres espèces du genre *Oryza* se réalisent avec une facilité globalement proportionnelle à la distance phylogénétique entre les espèces croisées. Ainsi, des populations issues de croisements entre *O. sativa* et *O. rufipogon* ou *O. nivara* peuvent être obtenues par simples croisements et autofécondations, tandis qu'avec les autres espèces du génome AA il est nécessaire de recourir à des rétrocroisements successifs afin de récupérer une fertilité partielle. Avec les espèces des autres complexes, des méthodologies plus complexes comme le sauvetage d'embryons et la fusion de protoplastes sont nécessaires pour obtenir des hybrides (Brar et Khush, 2002).

L'introduction de gènes d'intérêt d'espèces sauvages vers *O. sativa*, a été réalisée avec succès à plusieurs reprises. Dans le cas des espèces avec des génomes autres que AA, l'introduction de ces caractères a été obtenu par rétrocroisements puis sauvetage d'embryon, technique principalement développée à l'IRRI (International Rice Research Institute). Malgré le faible nombre de recombinaisons entre les chromosomes d'*O. sativa* et ceux des espèces

Tableau I-5. Croisements interspécifiques entre *O. sativa* et les espèces du genre *Oryza* hors du groupe de génome AA

Espèce croisée avec <i>O. sativa</i>	Caractère introgressé vers <i>O. sativa</i>	Référence bibliographique
<i>O. punctata</i> [§]	-	(Yasui et Iwata, 1991)
<i>O. officinalis</i> [§]	-	(Jena et Khush, 1990; Li et al., 2008a)
	Résistance à <i>N. lugens</i>	(Hirabayashi et al., 2003)
	Résistance à <i>Sogatella furcifera</i> Horvath (Cicadelle à dos blanc)	(Brar et Khush, 2006)
	Résistance à <i>X. oryzae</i> pv. <i>oryzae</i>	(Brar et Khush, 2006)
<i>O. rhizomatis</i> [§]	-	(Brar et al., 1991)
<i>O. australiensis</i> [§]	-	(Multani et al., 1994)
	Résistance à <i>N. lugens</i> (gène <i>Bph10(t)</i>)	(Ishii et al., 1994)
	Résistance à <i>X. oryzae</i> pv. <i>oryzae</i>	(Brar et Khush, 2006)
	Floraison précoce	(Brar et Khush, 2002)
<i>O. minuta</i> [§]	-	(Tong et al., 2003)
	Résistance à <i>M. oryzae</i>	(Amante-Bordeos et al., 1992)
	Résistance à <i>X. oryzae</i> pv. <i>oryzae</i>	(Amante-Bordeos et al., 1992)
	Résistance à <i>N. lugens</i>	(Brar et Khush, 2006)
<i>O. alta</i> [§]	-	(Long et al., 1994; Fu et al., 2007)
<i>O. grandiglumis</i>	Augmentation du rendement et résistance à <i>M. oryzae</i>	(Yoon et al., 2006)
<i>O. latifolia</i> [§]	-	(Multani et al., 2003; Yi et al., 2008)
	Résistance à <i>N. lugens</i>	(Haiyuan et al., 2002)
	Résistance à <i>S. furcifera</i>	(Brar et Khush, 2006)
<i>O. eichingeri</i> [§]	-	(Yan et al., 2001; Yan et al., 2002)
<i>O. meyeriana</i> [§]	-	(Xue-Lin et al., 2009)
<i>O. granulata</i> [§]	-	(Brar et al., 1991)
<i>O. brachyantha</i> [†]	-	(Hayashi et al., 1988)
	Résistance à <i>X. oryzae</i> pv. <i>oryzae</i>	(Brar et al., 1996)
<i>O. coarctata</i> [†]	-	(Brar et al., 1997; Jelodar et al., 1999)

§ Croisements faits par MAAL

† Croisements faits par fusion de protoplastes

appartenant aux autres groupes génomiques, l'introduction des caractères a pu être réalisée. La localisation chromosomique des gènes introgressés vers *O. sativa* est obtenue par l'approche de MAAL ou « Monosomic alien addition lines » (Brar et Khush, 1997). Ces lignées possèdent un chromosome additionnel appartenant au génome des espèces sauvages de riz ($2n = 25$), ce qui permet d'identifier le chromosome donneur du caractère. Une liste exhaustive des croisements entre *O. sativa* et des espèces d'*Oryza* de génome non-AA, et des introgressions de gènes d'intérêt vers *O. sativa* est présentée dans le Tableau I-5.

I.5.2.1 Populations interspécifiques du complexe *O. sativa* et gènes d'intérêt de ces espèces

En raison de leur large gamme d'adaptation à différents environnements, les espèces du groupe de génome AA représentent une importante source de gènes pour l'amélioration variétale du riz cultivé. Notamment, ces espèces possèdent plusieurs sources de tolérances aux stress abiotiques, de résistance à différents ravageurs et maladies, et le potentiel d'augmenter le rendement. Comme nous l'avons mentionné plus haut, les hybridations entre les espèces du complexe *O. sativa* s'obtiennent par croisement direct.

Dans ce paragraphe, nous présentons un résumé des gènes d'intérêt connus pour chacune des espèces du génome AA, les croisements réalisés avec *O. sativa*, et les cas d'introgression de ces traits vers *O. sativa*.

I.5.2.1.1 *O. rufipogon*

L'espèce asiatique sauvage *O. rufipogon* se distribue principalement en zones de marécage. Les plantes ont une allure élancée, une floraison tardive, et sont capables de se propager végétativement, ce qui en fait une espèce pérenne (Figure I-3 C) (Cai *et al.*, 2008). Malgré ses performances agronomiques médiocres, *O. rufipogon* est une source reconnue de gènes capables d'augmenter le rendement. Plusieurs chercheurs ont développé des populations hybrides entre les deux espèces afin d'identifier des QTL « Quantitative Trait Loci » liés au rendement, puis les introgresser vers *O. sativa*. Des QTL pour le nombre de grains (Xiao *et al.*, 1998), et des QTL pour le rendement, utiles pour des variétés adaptées à la culture pluviale (Moncada *et al.*, 2001), et applicables chez *indica* (Septiningsih *et al.*, 2003) comme chez *japonica* (Thomson *et al.*, 2003), et pour une large gamme de variétés de riz (Ahn *et al.*, 2003; Bautista *et al.*, 2003; Martinez *et al.*, 2003; Lee *et al.*, 2005; Tian *et al.*, 2006a; Tian *et al.*, 2006b; Tan *et al.*, 2007; Xie *et al.*, 2008) ont été identifiés et dans certains cas introgressés vers *O. sativa*.

O. rufipogon est aussi une source de résistances à divers pathogènes. Des gènes de résistance au Tungro et d'antibiose à l'insecte vecteur correspondant, *N. virescens*, ont été détectés

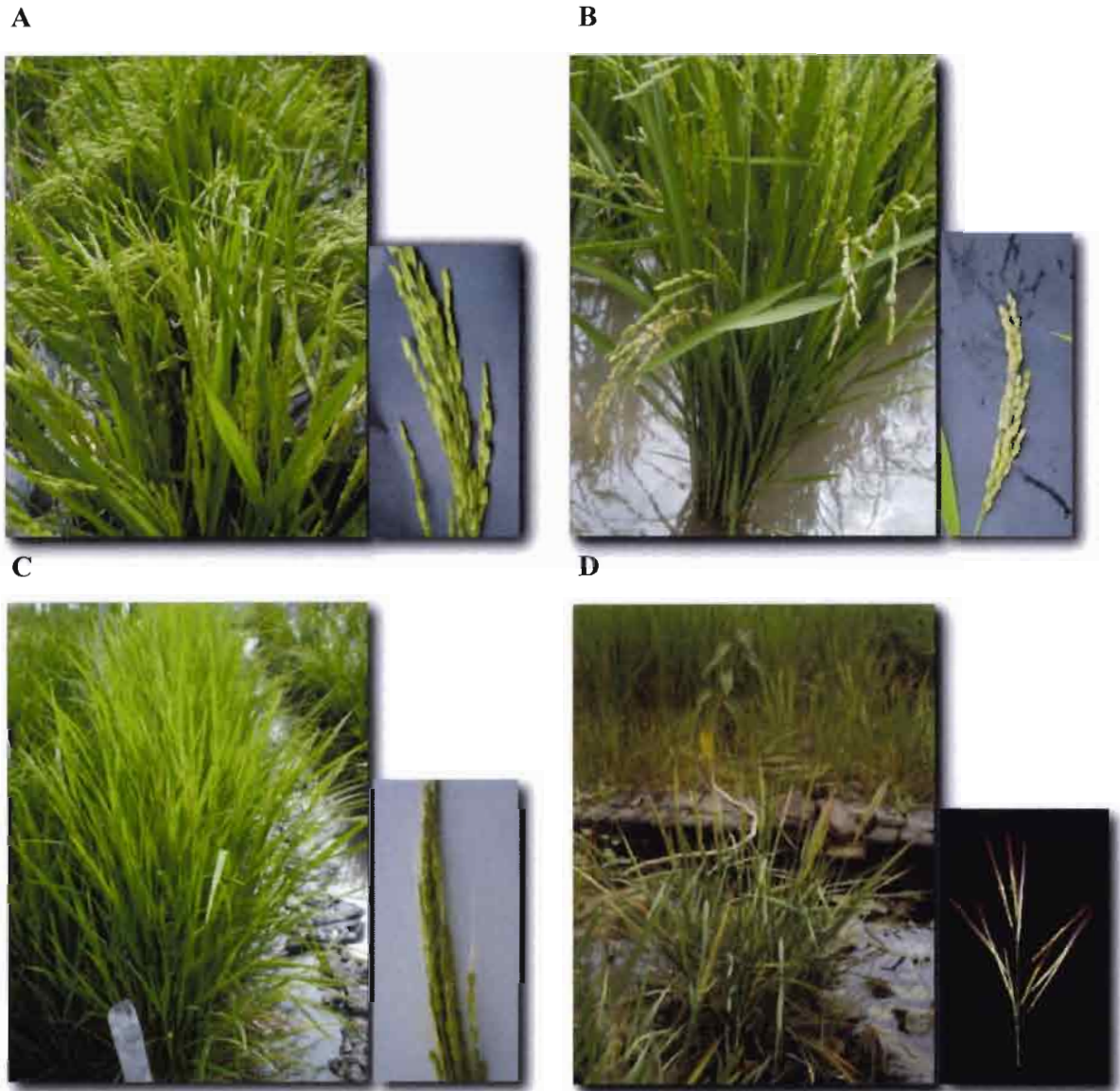


Figure I-3. Les différentes espèces du complexe *O. sativa* (génom AA). A. *O. sativa* ssp. *indica* cv. IR64. B. *O. sativa* ssp. *japonica* cv. Nipponbare. C. *O. rufipogon*. D. *O. nivara*. E. *O. glumaepatula*. F. *O. longistaminata*. G. *O. meridionalis*. H. *O. barthii*. I. *O. glaberrima* cv. TOG5681. J. *O. glaberrima* cv. CG14. Toutes les photographies appartiennent à la collection d'images du groupe de génétique et génomique du riz CIAT, gérée par J. Lozano, sauf pour F. (photo de la plante) et I. prises par nos soins, et D. tirée du site web http://www.gene.affrc.go.jp/databases-plant_images_en.php),

(Kobayashi *et al.*, 1993) et introgressés vers *O. sativa* (Brar et Khush, 2002). D'une façon similaire, la détection du gène de résistance à *X. oryzae* pv. *oryzae*, *Xa23(t)* (Zhang *et al.*, 1998), et son introduction dans le fonds génétique *indica* (Zhou *et al.*, 2009) ont été achevés. Finalement, des locus de tolérance à différents stress abiotiques, comme des hautes concentrations en fer (Mendoza *et al.*, 2003), en aluminium (Nguyen *et al.*, 2003) et de tolérance à la sécheresse (Liu *et al.*, 2004b; Zhang *et al.*, 2006) sont aussi présents dans le génome de différents accessions d'*O. rufipogon*, et représentent une excellente source d'amélioration des variétés d'*O. sativa*.

I.5.2.1.2 *O. nivara*

O. nivara est encore parfois cataloguée dans la littérature comme l'écotype annuel d'*O. rufipogon*. L'espèce est adaptée aux zones de marécage temporaire, qui sèchent selon la saison. Les plantes présentent une faible hauteur, une haute capacité de production et dispersion de graines, ce qui fait de cette espèce une bonne source de gènes pour le rendement (Figure I-3 D) (Cai *et al.*, 2008).

Une population hybride entre IR64 et *O. nivara* a permis de montrer le potentiel de l'espèce sauvage pour l'amélioration du contenu en protéines des graines, qui confère une meilleure qualité à la cuisson (Mahmoud *et al.*, 2007). Il existe plusieurs exemples de gènes de résistance introduits à partir d'*O. nivara* vers *O. sativa*. Le gène dominant *Gs*, qui confère la résistance au RGSV (*Rice Grassy Stunt Virus*) et qui est présent dans plusieurs variétés élités de *O. sativa*, a été le premier gène introduit d'une espèce sauvage vers *O. sativa* (Brar et Khush, 2002). Il provient d'une accession d'*O. nivara*, résistante à ce virus, identifiée et utilisée comme géniteur pour l'introgession de la résistance dans les années 70 (Khush et Ling, 1974). De la même manière, la résistance à *N. lugens* a été introduite dans la variété IR20 (Khush, 1979). Récemment, un gène de résistance à *X. oryzae* pv. *oryzae* fut également répertorié chez *O. nivara* et introduit dans *O. sativa* (Cheema *et al.*, 2008).

I.5.2.1.3 *O. glumaepatula*

O. glumaepatula est l'espèce américaine du complexe *O. sativa* (Figure I-3 E). Elle est annuelle, bien adaptée aux marécages, et plus particulièrement aux zones à inondation rapide de l'Amazonie (Akimoto *et al.*, 1998). Cette adaptation aux variations de niveaux d'eau fait d'*O. glumaepatula* une bonne source de gènes pour la tolérance à l'immersion (Akimoto et Morishima, 2003). Actuellement, au moins deux populations interspécifiques entre *O. sativa* et *O. glumaepatula* ont été développées. La première est dérivée d'un croisement avec la variété de *japonica* Taichung 65 (Sobrizal *et al.*, 1999). Cette population fut utilisée pour cartographier le gène de domestication de l'égrenage (*Sh3-glum*), et pour le port

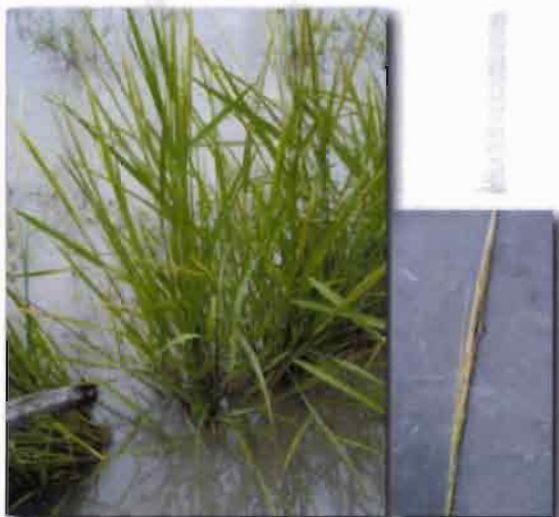
E



F



G



H



I



J

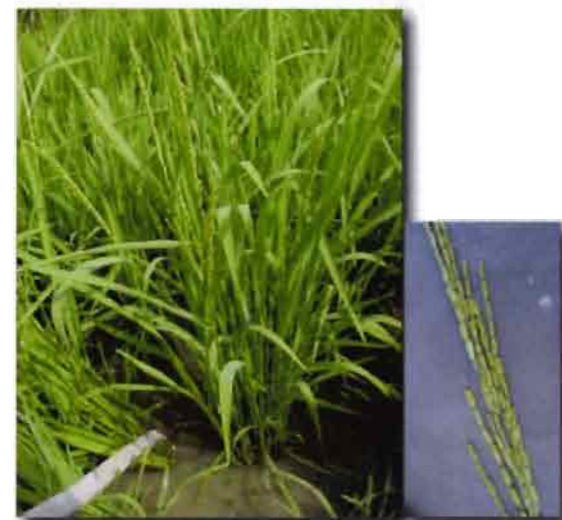


Figure I-3. (Suite)

paniculaire ouvert (*Spr4*) (Doi *et al.*, 2003a). Des gènes de floraison tardive de l'espèce *Lhd1(t)* et *Lhd2(t)* (Sanchez *et al.*, 2003), utiles pour l'adaptation des variétés de riz aux conditions locales de photopériode furent aussi identifiés en utilisant cette population. La deuxième population a été obtenue par rétrocroisements entre *O. glumaepatula* acc. RS-16 et le parental récurrent *O. sativa* acc. BG90-2 (Brondani *et al.*, 2001). L'évaluation de cette population pour différents caractères agronomiques permirent la détection de plusieurs QTL de rendement (Brondani *et al.*, 2002) qui furent ensuite caractérisés et inclus dans le programme d'amélioration variétale de l'Embrapa (Empresa Brasileira de Pesquisa Agropecuária) (Rangel *et al.*, 2008).

I.5.2.1.4 O. longistaminata

O. longistaminata (Figure I-3 F) est une espèce africaine pérenne, allogame, capable de se propager végétativement par rhizomes (Ghesquière, 1986). Elle représente un réservoir de gènes pour le rendement (Brar, 2005), pour la résistance à *M. graminicola* (Soriano *et al.*, 1999), et possède des caractères associés à la tolérance et à la sécheresse (Liu *et al.*, 2004b).

Un rétrocroisement entre une accession d'*O. longistaminata* (acc. WL02) et *O. sativa* (acc. BS125) a permis le développement d'une population interspécifique. Cette population a été étudiée pour différentes caractéristiques agronomiques (Causse et Ghesquiere, 1991). Plus tard, cette population a constitué le matériel biologique pour l'établissement de la première carte génétique hautement saturée du riz à base de marqueurs RFLP (Causse *et al.*, 1994). Une autre population interspécifique a permis de montrer la possibilité d'obtenir des variétés cultivées de riz à rhizomes (Tao *et al.*, 2003). Ce caractère de pérennité pourrait être utile pour stabiliser les sols et éviter l'érosion des couches fertiles. L'introduction dans des lignées élites comme IR64 du gène de résistance à *X. oryzae* pv. *oryzae*, *Xa21* (Khush *et al.*, 1990; Brar et Khush, 2002), et les lignées qui ont hérité d'un gène de résistance au Foreur jaune de la tige (Brar, 2005), sont des exemples d'introgessions réussies à partir de *O. longistaminata*.

I.5.2.1.5 O. meridionalis

O. meridionalis se rencontre dans les marécages temporaires australiens, qui s'assèchent selon la saison. C'est une espèce annuelle, qui produit des panicules compactes avec des graines allongées (Figure I-3 G). Les caractères qu'*O. meridionalis* a acquis au cours de son évolution, sur tout la tolérance à la sécheresse, peuvent représenter un bon réservoir pour l'amélioration de *O. sativa*. Toutefois, l'espèce reste encore peu étudiée, et en conséquence son potentiel pour l'amélioration du riz demeure encore largement inconnu. Une population qui utilise comme parent récurrent la variété de *O. sativa* Taichung 65 a été développée et a

permis la cartographie de deux gènes qui contrôlent l'aristation, *An7* et *An8* (Kurakazu *et al.*, 2001; Doi *et al.*, 2003a).

1.5.2.1.6 *O. barthii*

O. barthii (Figure I-3 H) est une espèce africaine annuelle, adaptée aux zones de marécage pouvant fortement s'assécher pendant l'été. Cette adaptation fait de cette espèce un réservoir de gènes potentiel pour la tolérance à la sécheresse. Des résistances à *P. grisea* (Eizenga *et al.*, 2002) et à *R. solani* (Eizenga *et al.*, 2004) ont été également décrites. Malgré cela, peu de travaux de recherche ont inclut cette espèce comme source de gènes d'intérêt. Le croisement réalisé au CIAT entre *O. sativa* cv. Lemont et *O. barthii* représente à notre connaissance le seul cas de descendance obtenue entre les deux espèces. Ce croisement a permis de mettre en évidence plusieurs QTL pour l'augmentation du rendement (Almeida *et al.*, 2004).

1.5.2.1.7 *O. glaberrima*

O. glaberrima est l'espèce cultivée africaine de riz (Figure I-3 I et J). Malgré sa diversité génétique réduite, *O. glaberrima* s'est bien adaptée aux conditions difficiles de l'agriculture en Afrique. De par ces adaptations environnementales, les accessions d'*O. glaberrima* s'avèrent être des sources de tolérance à une large gamme de stress biotiques et abiotiques. Les variétés d'*O. glaberrima* peuvent être classifiées en pluviales et inondées. Les variétés pluviales sont adaptées à d'éventuelles sécheresses, parfois sévères. Elles possèdent des mécanismes de fermeture rapide des stomates, d'enroulement des feuilles et un système racinaire étendu (Dingkuhn *et al.*, 1999; Asch *et al.*, 2005). Les variétés adaptées à la culture inondée possèdent une bonne tolérance aux inondations prolongées, de par leur capacité d'élongation rapide des entrenœuds et l'augmentation de la capacité photosynthétique des feuilles à la surface de l'eau (Sakagami *et al.*, 2009). *O. glaberrima* est aussi une source de gènes pour la tolérance à de fortes concentrations en fer (Sahrawat et Sika, 2002; Audebert et Fofana, 2009), et pour une meilleure efficacité de l'utilisation de l'azote (Tagawa *et al.*, 2000).

Au niveau des stress biotiques, plusieurs gènes importants ont été répertoriés chez *O. glaberrima*. La résistance de plusieurs accessions de cette espèce au RYMV (Ndjiondjop *et al.*, 1999), est due à une réduction dans le mouvement du virus entre les cellules (Albar *et al.*, 2003). *O. glaberrima* possède aussi des gènes de résistance aux nématodes *Heterodera sacchari* (Reversat et Destombes, 1998) et *M. graminicola* (Soriano *et al.*, 1999). L'espèce dispose par ailleurs d'une grande capacité de compétition contre les adventices, grâce à une croissance végétative rapide, un fort tallage et des feuilles à port « pendant » et présentant un

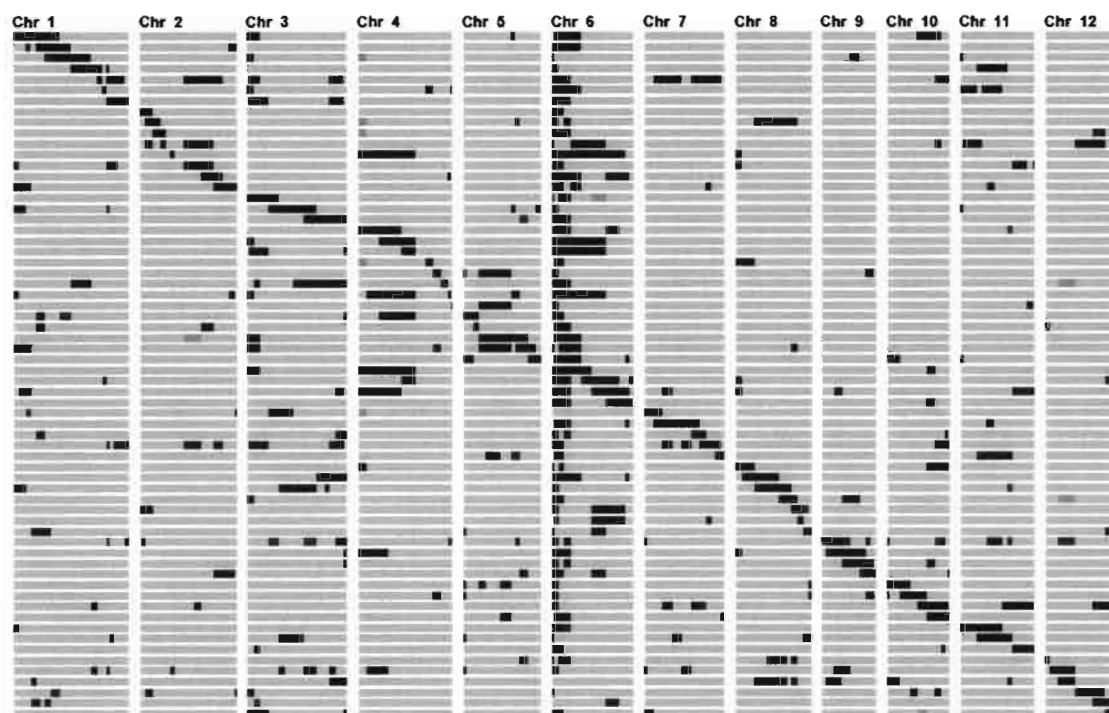
Leaf Area Index (LEA) élevé, et « *droopy leaves* » qui diminuent fortement le rayonnement au sol (Johnson *et al.*, 1998).

Malgré ces caractères avantageux, *O. glaberrima* a été progressivement abandonnée par les agriculteurs depuis l'introduction de *O. sativa*, qui offre des rendements plus élevés. L'égrenage, un petit nombre d'épillets par panicule, un nombre réduit ou nul de ramifications secondaires, une forte tendance à la verse due à la faiblesse de ses tiges en phase reproductive, et la fragilité des graines à l'usinage sont les facteurs d'un rendement faible chez l'espèce africaine (Vietmeyer, 1996). Actuellement, *O. glaberrima* n'est plus cultivée que de manière anecdotique, par exemple dans des parcelles à sols pauvres où *O. sativa* offre des rendements plus faibles, ou à des fins rituelles, ou encore comme nourriture symbolique de pouvoir pour les chefs de tribu (Bezançon, 1994). Cet abandon menace bien évidemment la diversité de l'espèce et les possibilités de l'utiliser comme source de gènes.

Dans le but de caractériser et introgresser divers caractères intéressants d'*O. glaberrima* vers *O. sativa*, différentes populations interspécifiques ont été développées par plusieurs groupes de recherche :

- *O. sativa* ssp. *japonica* acc. Taichung 65 x *O. glaberrima* acc. IRGC104038 : une population BC₁F₁ T65/IRGC104038//T65 a été développée et sa carte génétique a été construite à l'aide de marqueurs RFLP (Doi *et al.*, 1998b). Des QTL pour la stérilité pollinique et le gène *Sh3* d'égrenage et son inhibiteur *iSh3*, ont été répertoriés (Nagai *et al.*, 2002). Egalement, le gène de floraison précoce *Efd1* fut identifié grâce à cette population. Ce gène s'avère utile pour le développement de variétés adaptées aux conditions de jours courts (Doi *et al.*, 2004).
- *O. sativa* cv. V20A x *O. glaberrima* acc. IRGC103544 (MG12) : une population BC₃F₁ a été développée afin de rechercher des QTL associés à la qualité du grain (Li *et al.*, 2004). 12 QTL pour sept caractères de qualité furent détectés.
- *O. sativa* ssp. *japonica* cv. Caiapo x IRGC103544 (MG12) : cette population a été développée au CIAT. Les individus BC₃-HD (haploïdes doublés) de cette population furent évalués pour des composants de la qualité du grain, et des QTL pour 9 de ces caractères ont été répertoriés (Aluko *et al.*, 2004). Des travaux ultérieurs permirent de détecter des QTL additionnels pour le rendement, et un QTL majeur pour la résistance au RSNV (*Rice Stripe Necrosis Virus*) (Gutierrez *et al.*, publication soumise). Par ailleurs, ces lignées ont été utilisées pour définir une librairie de CSSL, ou lignées de substitution de segments chromosomiques, qui portent chacune un fragment chromosomique introgressé et représentent le génome du parent donneur dans le fonds génétique du parent récurrent. Un rétrocroisement additionnel impliquant 59 plantes choisies pour leur bonne représentation du

A



B

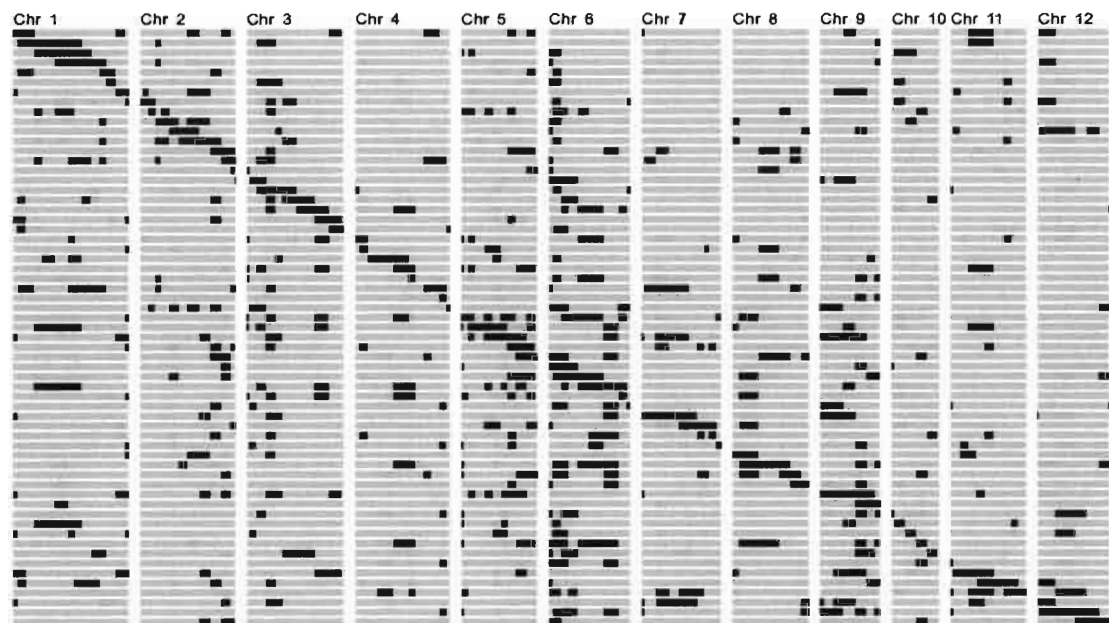


Figure I-4. Exemples de populations CSSL issues des croisements entre *O. sativa* et *O. glaberrima*. Génotypes graphiques des deux populations CSSL en cours de développement au CIAT et à l'IRD. **A.** 64 individus BC₃-HD issus du croisement *O. sativa* ssp. *japonica* cv. Caiapo x *O. glaberrima* cv. MG12. **B.** 61 individus BC₂F₄ et BC₃F₃ issus du croisement *O. sativa* ssp. *indica* cv. IR64 x *O. glaberrima* acc. TOG5681. Le génome du parent récurrent *O. sativa* est représenté en gris clair, et le génome du donneur *O. glaberrima* est représenté en gris foncé. Source : A. Gutierrez, communication personnelle.

génomique d'*O. glaberrima* (voir Figure I-4 A) a été réalisé afin d'en purifier le fonds génétique. L'évaluation génotypique de 4200 descendants BC₄F₂ est en cours.

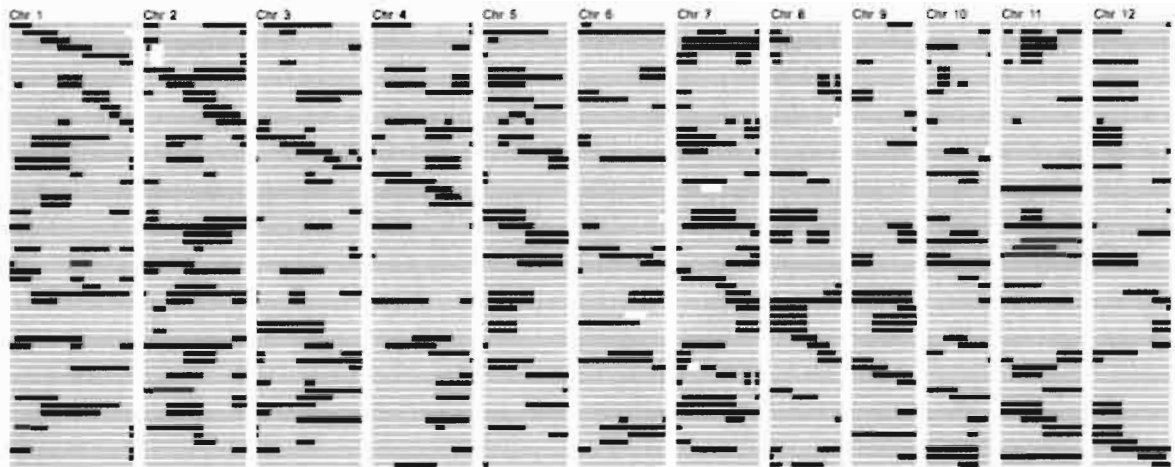
- *O. sativa* ssp. *indica* cv. IR64 x *O. glaberrima* acc. TOG5681 : cette population a été développée par le groupe de génétique et génomique du riz de l'IRD (Ghesquière *et al.*, 1997). A partir des lignées BC₁F₁ de ce croisement, la première carte génétique interspécifique à base de marqueurs SSR impliquant *O. glaberrima* fut développée et des QTL pour la hauteur de plante, la sensibilité à la photopériode et l'aristation furent décrits (Lorieux *et al.*, 2000).

Par la suite, la population a été avancée en génération BC₂F₄ et BC₃F₃, qui constituent des lignées « pré-CSSL ». Le génotypage par marqueurs SSR de la population de 353 lignées a permis d'identifier 61 lignées qui représentent la quasi totalité du génome de TOG5681 (voir Figure I-4 B). Des QTL pour des caractères associés au rendement ont pu être identifiés à partir de l'analyse de la population entière. Par ailleurs, des populations en ségrégation dérivées de ce même croisement (IR64 x TOG5681) ont constitué la base de la cartographie génétique de gènes majeurs pour la résistance au nématode *H. sacchari* (Lorieux *et al.*, 2003), et au virus RYMV (Albar *et al.*, 2003).

- NERICAS : Dans les années 90, le Centre du Riz pour l'Afrique (ADRAO-WARDA) a réalisé des croisements entre huit variétés d'*O. glaberrima* hautement compétitives contre les adventices et tolérantes à la sécheresse, et cinq variétés élites africaines d'*O. sativa*. Parmi les familles obtenues, une fut retenue pour sa fertilité, et avancée en génération afin de fixer les traits d'intérêt ou et dériver des haploïdes doublées (Jones *et al.*, 1997). Les premières variétés NERICA (New Rice for Africa) sont directement issues de ces lignées et représentent donc un ensemble de lignées interspécifiques. Les lignées libérées sont cataloguées comme variétés pluviales ou inondées, selon le parent *O. sativa* utilisé. Les 18 lignées obtenues à partir du croisement WAB56-104, un *O. sativa* ssp. *japonica* de type pluvial et CG14 (*O. glaberrima*), sont adaptées au régime pluvial (Semagn *et al.*, 2006; Semagn *et al.*, 2007), tandis que les 11 provenant du croisement TOG5681 (*O. glaberrima*) x IR64 (*O. sativa* ssp. *indica* de type irrigué) sont adaptées au régime inondées (Ndjondjop *et al.*, 2008).

Ces lignées ont fait l'objet d'une large diffusion en Afrique sub-saharienne. Elles ont été étudiées pour la qualité du grain (Futakuchi *et al.*, 2008; Kishine *et al.*, 2008), le contenu en protéines (Watanabe *et al.*, 2006), la tolérance aux inondations soudaines (Kawano *et al.*, 2008), et le rendement (Sie *et al.*, 2008). Elles représentent un exemple réussi du potentiel des hybrides intraspécifiques dans la production de riz.

A



B

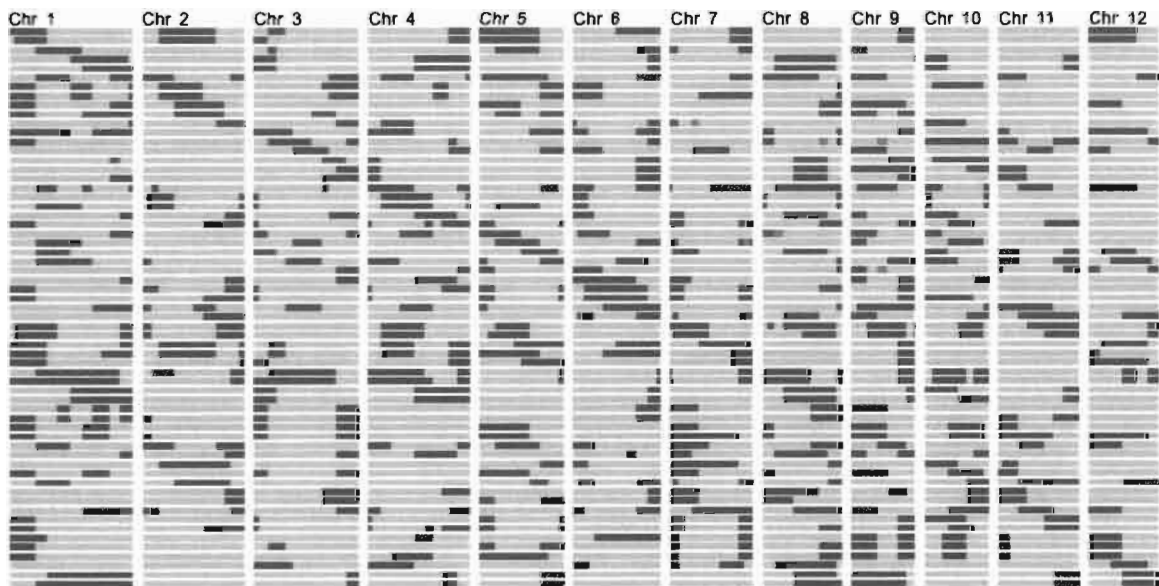


Figure I-5. Lignées CSSL développées dans le cadre du projet du GCP pour l'exploration de la diversité naturelle du riz. Génotypes graphiques des lignées pré-CSSL en la génération BC_2F_1 , pour les croisements Curinga x *O. rufipogon* acc. IRGC105491 **A**, et Curinga x *O. meridionalis* acc. OR44 **B**. En gris clair le génome du parent récurrent Curinga, et en gris foncé le génome du respective donneur. Source : J. D. Arbelaez et L. Moreno, communication personnel.

I.5.2.1.8 Projet pour le développement de ressources génétiques et CSSL à partir de quatre espèces de riz du génome AA

Quatre populations intraspécifiques de type CSSL, en plus de celles obtenues par croisements *O. sativa* x *O. glaberrima*, sont actuellement en cours de développement dans le cadre du projet « Exploring Natural Genetic Variation: Developing Genomic Resources and Introgression Lines for Four AA Genome Rice Relatives ». (http://www.generationcp.org/vw/Download/Competitive_Grant_Proposals/10_TOHME.pdf) visant à l'exploration de la diversité naturelle du riz et financé par le *Generation Challenge Program*. Ce projet est piloté par l'Unité de Biotechnologie du CIAT (Colombie) et par le groupe de Génétique et Génomique du Riz de l'IRD, en partenariat avec l'université Cornell (USA), l'EMBRAPA-CNPAP (Brésil), l'ADRAO-WARDA (Bénin) et Fedearroz (Colombie). L'objectif principal du projet est le développement de ressources biologiques qui permettront l'analyse génétique de caractères provenant de quatre espèces sauvages du groupe *O. sativa*, par la création de lignées CSSL partageant le même fonds génétique (*O. sativa* ssp. *japonica* tropical). Les populations ont été dérivées à partir des croisements suivants :

- *O. sativa* ssp. *japonica* cv. Curinga x *O. rufipogon* acc. IRGC105491 (Fedearroz, CIAT)
- Curinga x *O. meridionalis* acc. OR44 (Fedearroz, CIAT)
- Curinga x *O. glumaepatula* acc. GEN1233 (Embrapa)
- Curinga x *O. barthii* acc. IRGC101937 (CIAT, WARDA)

Le suivi des introgressions a été réalisé à chaque génération à l'aide de marqueurs SSR tirés de la carte génétique universelle du riz (voir Chapitre II). Actuellement, ces populations sont en génération BC₃, et des haploïdes doublées (BC₃HD) dérivés des BC₃F₁ ont été développés pour les deux premiers croisements.

Un exemple de génotypes graphiques pour deux de ces populations est présenté en la Figure I-5.

I.5.3 Limitations des approches interspécifiques

L'importante diversité génétique qui caractérise les espèces apparentées à *O. sativa* a amené les sélectionneurs et généticiens à considérer leur utilisation en l'amélioration variétale comme une approche prometteuse. Cependant, cette approche est fortement limitée par les barrières de reproduction qui séparent les espèces du genre *Oryza*. Les premiers essais d'hybridation interspécifique ont mis en évidence le fait que différents mécanismes empêchaient l'échange de matériel génétique entre les espèces. Des observations cytologiques ont permis de différencier les barrières causées par des défauts d'appariement des chromosomes pendant la méiose (pour les espèces les plus éloignées), des déséquilibres génétiques entre les espèces proches causés par des incompatibilités

Tableau I-6. Pourcentage de fertilité paniculaire des hybrides entre espèces du genre *Oryza* de génome AA.

Parent femelle	Parent male							
	<i>O. sativa</i>	<i>O. rufipogon</i>	<i>O. nivara</i>	<i>O. glaberrima</i>	<i>O. barthii</i>	<i>O. glumaepatula</i>	<i>O. longistaminata</i>	<i>O. meridionalis</i>
<i>O. sativa</i>	-*	26,20	11,90	-*	0,30	1,10	9,60	-
<i>O. rufipogon</i>	57,40	49,60	28,40	5,60	0,70	2,30	6,20	6,60
<i>O. nivara</i>	-	44,80	62,90	0,10	1,30	1,60	0,60	0,90
<i>O. glaberrima</i>	-*	4,70	0,60	-*	82,4	0,80	-	0,10
<i>O. barthii</i>	1,60	2,00	1,10	77,20	79,30	0,40	6,70	0,40
<i>O. glumaepatula</i>	1,40	2,90	5,40	-	0,60	85,60	1,20	0,70
<i>O. longistaminata</i>	-	-	-	-	-	1,40	91,60	-
<i>O. meridionalis</i>	-	8,70	1,00	0,03	0,10	0,40	-	41,50

Pourcentage d'épillets remplis provenant de cinq panicules ensachées. D'après les données publiés par Lu *et al.*, 2003. * Croisement non testé

alléliques (Oka, 1953; Oka, 1957a). Dans les années 70, Chu et Oka réalisèrent des études exhaustives de l'interfécondité (*crossability*) entre les espèces du groupe AA, en croisant les différentes espèces décrites à l'époque (Chu *et al.*, 1969; Chu et Oka, 1970). Plus tard, plusieurs auteurs répétèrent ce type d'analyse pour d'autres espèces du genre comme *O. longistaminata* (Ghesquiere, 1991; Ghesquière et Causse, 1992), *O. glumaepatula* (Naredo *et al.*, 1998), ou *O. meridionalis* (Naredo *et al.*, 1997). Plus récemment, une équipe de chercheurs de l'IRRI a conduit de nouvelles expérimentations, en croisant les espèces actuellement décrites du groupe de génome AA (Lu *et al.*, 2003). Un exemple de mesures d'interfécondité entre les espèces du complexe *O. sativa* est présenté dans le Tableau I-6. Les résultats de ces études mettent en évidence de grandes différences dans la réussite des croisements et de leurs réciproques, ainsi que l'existence de différences dans les mécanismes de barrières d'isolement reproductif entre les espèces du genre *Oryza*. Dans certains cas, les semences hybrides produites montrent un développement réduit de l'embryon ou de l'endosperme ; dans d'autres cas les hybrides se développent normalement, mais leur fertilité est réduite ou nulle.

Afin de mieux comprendre les mécanismes et la génétique des barrières de reproduction entre les espèces apparentées à *O. sativa*, nous présentons dans les paragraphes suivants une vue détaillée des différents types de barrières reproductives et des gènes identifiés qui les contrôlent.

I.6 Les barrières reproductives chez les plantes

Les barrières reproductives peuvent être définies comme les différents mécanismes qui empêchent les flux de gènes d'une population à une autre. Elles jouent un rôle très important dans l'établissement et la conservation des espèces. Elles peuvent agir avant croisement, après croisement mais préalablement à la fécondation, ou postérieurement à la formation des zygotes hybrides. Elles agissent de manière individuelle ou séquentielle, en empêchant l'échange de matériel génétique (Coyne et Orr, 2004).

I.6.1 Barrières reproductives avant croisement

Comme son nom l'indique, ce type de barrières agit préalablement à la rencontre des gamètes mâles et femelles. Elles réunissent une large gamme d'événements écologiques, morphologiques et des caractéristiques inhérentes au système de croisement. Parmi les barrières écologiques les plus importantes, on peut citer les adaptations aux différents habitats ou distributions ecogéographiques (Ramsey *et al.*, 2003), l'attraction de certains types de pollinisateurs (Bradshaw *et al.*, 1995; Hoballah *et al.*, 2007), ou encore les différences de

Tableau I-7. Gènes connus comme causant une létalité hybride

Espèces impliquées	Gène et fonction	Effet sur la viabilité des hybrides	Référence bibliographique
<i>Xiphophorus maculatus</i> et <i>X. helleri</i>	<i>Tu-Sd</i> Récepteur tyrosine-kinase	Induction d'un mélanome malin chez les poissons hybrides causé par l'absence de répression de l'oncogène par le locus <i>R</i>	(Wittbrodt <i>et al.</i> , 1989).
<i>Drosophila simulans</i> et <i>D. melanogaster</i> .	<i>Nup96</i> et <i>Nup160</i> Protéines des pores nucléaires	Les allèles <i>Nup160^{sim}</i> et <i>Nup96^{sim}</i> sont incompatibles avec d'autres composants du complexe des pores nucléaires de <i>D. melanogaster</i> , causant la létalité	(Presgraves <i>et al.</i> , 2003) (Tang et Presgraves, 2009)
<i>D. simulans</i> et <i>D. melanogaster</i>	<i>Lhr</i> et <i>Hmr</i> Protéines de fixation à l'ADN	L'incompatibilité entre <i>Hmr^{mel}</i> et <i>Lhr^{sim}</i> cause la mort des hybrides mâles	(Brideau <i>et al.</i> , 2006; Barbash <i>et al.</i> , 2004)
Diverses accessions d' <i>Arabidopsis thaliana</i>	<i>DM1</i> Domaine NB-LRR, homologue à celui des gènes de résistance	Nécrose des tissus des plantes hybrides, due à une activation du système de défense inadéquate, suite à l'interaction entre <i>DM1</i> et <i>DM2</i>	(Bomblies <i>et al.</i> , 2007; Bomblies et Weigel, 2007)
Diverses accessions d' <i>A. thaliana</i>	<i>HPA1</i> et <i>HPA2</i> paire de gènes paralogues qui ont un rôle dans la synthèse de l'histidine	Du fait d'une évolution divergente, chaque accession a conservé une copie fonctionnelle différente. L'héritage simultané des deux copies non exprimées dans les hybrides cause l'avortement des embryons	(Bikard <i>et al.</i> , 2009)
<i>A. thaliana</i> et <i>A. arenosa</i>	Rétrotransposon ATHILA,	Les effets de dosage et de régulation par silençage de ATHILA et d'autres gènes, qui normalement sont spécifiques de chaque génome parental, se voient sérieusement perturbés dans les embryons hybrides, contribuant à leur avortement.	(Josefsson <i>et al.</i> , 2006)

phénologie de la floraison (Martin *et al.*, 2007). Les changements dans le système de croisement causés par des morphologies incompatibles des fleurs causant l'autogamie (Noland et John, 2007), ou la spécificité des pollinisateurs (Bradshaw *et al.*, 1998), peuvent aussi se cataloguer dans ce type de barrière.

I.6.2 Barrières reproductives après croisement, agissant avant la fécondation – cas des barrières pré-zygotiques

Même lorsque le transfert de pollen a lieu, une autre série d'évènements peut encore éviter la fertilisation, et par conséquent la formation d'un zygote. Chez les plantes, l'échec de la germination du pollen (Hodnett *et al.*, 2005), les déficiences de la croissance du tube pollinique, et de la fertilisation (Escobar-Restrepo *et al.*, 2007) sont autant de barrières pré-zygotiques non compétitives et qui jouent un rôle important. A l'opposé, la préférence pour le pollen provenant de la même espèce, ou provenance conspécifique du pollen (Zhiping *et al.*, 2002; Fishman *et al.*, 2008), constitue un exemple des barrières reproductives compétitives.

I.6.3 Barrières reproductives après la fécondation, agissant après la formation d'un zygote – cas des barrières post-zygotiques

Les barrières post-zygotiques peuvent elles aussi être classées selon leur origine (Coyne et Orr, 2004). Un hybride peut-être inadapté ou non viable du fait de conditions environnementales, s'il présente une faible adaptation à l'habitat normal des individus parentaux (Milne *et al.*, 2003). Dans ce cas, la barrière est dite de type extrinsèque. Au contraire, si la cause de la barrière reproductive est liée à un dysfonctionnement de l'individu, la barrière est classée comme intrinsèque. Parmi les barrières intrinsèques, on distingue la létalité hybride, la stérilité hybride et l'*hybrid breakdown*.

I.6.3.1 Létalité et faiblesse hybride de la F₁

Dans les cas de létalité ou faiblesse hybride, la fécondation est suivie de la formation d'un zygote, mais le développement de l'embryon se voit compromis, et les hybrides présentent une faible vigueur, ou meurent à un stade précoce de leur développement ou avant d'arriver à maturité. La létalité des hybrides chez les plantes peut ainsi se manifester par l'avortement de l'embryon (Phillips *et al.*, 1982), de l'endosperme (Inai *et al.*, 1993) ou des plantules (Mino *et al.*, 2002).

Les bases génétiques de cet effet létal sur les hybrides ont été identifiées pour plusieurs systèmes chez les animaux, mais aussi chez les plantes, recouvrant une large gamme de mécanismes. Malgré les différences entre les mécanismes identifiés, l'interaction entre

Tableau I-8. Gènes connus comme causants de la stérilité hybride

Espèces impliquées	Gène et fonction	Effet sur la fertilité des hybrides	Référence bibliographique
<i>D. mauritiana</i> et <i>D. simulans</i>	<i>Odyseus (OdsH)</i> Facteur de transcription du type Homéobox	Stérilité complète des hybrides mâles	(Ting <i>et al.</i> , 1998; Sun <i>et al.</i> , 2004),
Populations différenciées de <i>Drosophila virilis</i>	<i>Penelope</i> Rétrotransposon	Stérilité mâle due à une dysgenèse hybride. Le transposon induit aussi la mobilisation d'autres familles d'éléments transposables	(Blumenstiel et Hartl, 2005)
Populations différenciées de <i>D. melanogaster</i>	<i>P</i> et <i>I</i> Transposons	Stérilité hybride due à une dysgenèse hybride	(Brennecke <i>et al.</i> , 2008)
<i>D. simulans</i> et <i>D. melanogaster</i>	<i>JYAlpha</i> ATPase nécessaire à la production de gamètes mâles	La copie unique du gène se trouve sur des chromosomes différents selon l'espèce. Les individus qui n'ont hérité d'aucune copie sont stériles	(Masly <i>et al.</i> , 2006)
<i>Drosophila pseudoobscura</i> ssp. <i>pseudoobscura</i> et <i>D. pseudoobscura</i> ssp. <i>bogotana</i>	<i>Overdrive (Ovd)</i> Protéine avec un motif de fixation de l'ADN, MADF	Semi stérilité mâle et distorsion de ségrégation au niveau du chromosome X	(Phadnis et Orr, 2009)
<i>Saccharomyces bayanus</i> et <i>S. cerevisiae</i>	<i>AEP2</i> , facteur translationnel mitochondrial. <i>OLII</i> , Sub-unité de l'ATP synthase	Absence de sporulation dans les hybrides. L'allèle de <i>Sb-AEP2</i> ne peut pas traduire le gène mitochondrial <i>Sc-OLII</i>	(Lee <i>et al.</i> , 2008)
<i>Mus musculus</i> ssp. <i>Musculus</i> et <i>Mus musculus</i> ssp. <i>domesticus</i>	<i>Hst1</i> Histone méthyltransferase	Stérilité mâle. Possible interaction épistatique entre les allèles, qui empêcherait le gène d'accomplir sa fonction d'activation transcriptionnelle de gènes essentiels pour la méiose	(Mihola <i>et al.</i> , 2009)
<i>O. sativa</i> ssp. <i>indica</i> et <i>O. sativa</i> ssp. <i>japonica</i>	<i>S₅</i> Protéase aspartique	Stérilité femelle. L'interaction épistatique des allèles <i>S₅⁻ⁱ</i> et <i>S₅^{-j}</i> cause l'élimination gamétique. L'allèle neutre <i>S₅⁻ⁿ</i> n'interagit pas, du fait d'une localisation erronée de la protéine, évitant ainsi la stérilité.	(Chen <i>et al.</i> , 2008)
<i>O. sativa</i> ssp. <i>indica</i> et <i>O. sativa</i> ssp. <i>japonica</i>	<i>SaM</i> , protéine SUMO E3 ligase. <i>SaF</i> , protéine F-box homologue avec une RALyase	Semi stérilité male. Elimination gamétique due à l'interaction incompatible entre <i>SaF⁺</i> avec <i>SaM</i>	(Long <i>et al.</i> , 2008)

plusieurs gènes, connus ou non, est toujours la cause de la mort des hybrides. Un résumé des gènes de létalité hybride connus à ce jour est présenté dans le Tableau I-7.

A partir de croisements entre différentes espèces du genre *Drosophila*, plusieurs gènes ont été identifiés par une approche de cartographie basée sur les délétions de fragments chromosomiques et leur introduction dans les hybrides. Pour *Arabidopsis*, différentes barrières reproductives ont été également identifiées, certaines agissant entre accessions de la même espèce, et qui peuvent potentiellement conduire à des événements de spéciation. Des mécanismes épigénétiques peuvent aussi être à l'origine d'un isolement reproductif.

La dysgenèse hybride, un exemple de mécanisme épigénétique, consiste en une « *fitness* » réduite des hybrides du fait de la dérégulation de certains éléments transposables. La répression de la mobilité des transposons étant héritée maternellement par l'embryon (Brennecke *et al.*, 2008), le croisement d'individus, utilisés comme femelles et non porteurs d'un certain transposon avec un mâle porteur conduit à la dérégulation du transposon et à son activation dans les hybrides, affectant ainsi leur stabilité génétique.

I.6.3.2 Stérilité hybride

Il y a stérilité hybride lorsque les individus hybrides arrivent à maturité sexuelle, mais sont incapables de produire des gamètes fonctionnels de manière effective. Chez les plantes, ce type d'isolement reproductif a été bien documenté chez la tomate (Moyle et Nakazato, 2008; Moyle et Nakazato, 2009), le riz (voir ci-dessous), et chez *Mimulus* (Sweigart *et al.*, 2006), où il a été notamment démontré que les barrières interspécifiques peuvent varier et présenter du polymorphisme entre populations de la même espèce (Sweigart *et al.*, 2007).

Les réarrangements chromosomiques peuvent aussi causer une forte stérilité hybride, en empêchant le croisement et la recombinaison normale des chromosomes pendant la méiose (Coyne et Orr, 2004). Au niveau génétique, plusieurs gènes et mécanismes impliqués dans la stérilité hybride ont été identifiés, principalement chez la drosophile. En terme général, les gènes de stérilité hybride semblent subir une rapide évolution adaptative par rapport à d'autres types de gènes, à l'exception du gène *Ovd* qui semble avoir une évolution neutre accélérée (Phadnis et Orr, 2009). Un résumé des gènes de stérilité hybride identifiés à ce jour est présenté dans le Tableau I-8.

I.6.3.3 Hybrid breakdown ou faiblesse de la F₂

Ce type de barrière de reproduction se présente lorsque la génération F₂ ou postérieure montre une réduction de viabilité et de fertilité, et un développement perturbé en comparaison avec la F₁. La cause la plus probable de ce type de barrière est attribuée au fait que des allèles récessifs incompatibles ou nocifs deviennent homozygotes par effet de l'autofécondation.

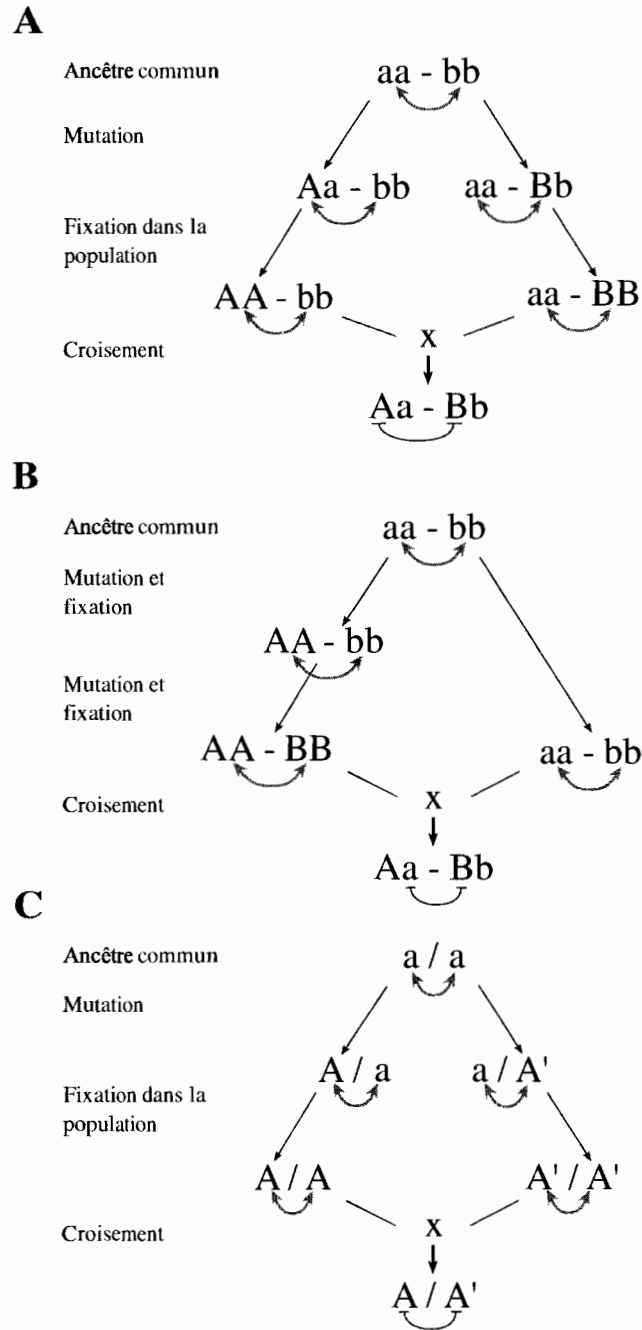


Figure I-6. Modèle d'incompatibilité BDM dans les barrières reproductives post-zygotiques. **A.** A partir d'un ancêtre commun qui possède les allèles a et b pour deux gènes en interaction épistatique, chacun des gènes suit une évolution divergente qui les amène à se fixer dans deux populations sous forme d'allèles A et B incompatibles. La rencontre dans un hybride des allèles incompatibles A et B induit une létalité ou une stérilité. **B.** La divergence allélique a lieu dans une seule des populations, l'autre conserve les allèles ancestraux. Comme dans le modèle (A), l'incompatibilité entre a et B cause stérilité ou létalité dans les hybrides. **C.** Incompatibilité entre des allèles du même locus et leur évolution dans des populations en divergence. Les flèches pointues (vertes) et tronquées (rouges) représentent respectivement les interactions compatibles et incompatibles.

Chez les plantes, et plus particulièrement chez *Mimulus* (Fishman et Willis, 2001; Sweigart *et al.*, 2006) et le riz (voir ci-dessous), l'*hybrid breakdown* est un phénomène bien étudié. Pourtant, seuls les gènes candidats pour un couple de locus ont pu être identifiés. Il s'agit de *hwh1* et *hwh2* qui sont supposés coder respectivement pour une protéine GMC oxydoréductase et pour une protéine transporteur d'hexose (Jiang *et al.*, 2008).

I.6.3.4 Modèle Bateson-Dobzhansky-Muller (BDM) pour les barrières post-zygotiques

Depuis les travaux de Darwin sur la spéciation, les évolutionnistes ont cherché à savoir comment les allèles impliqués dans les barrières reproductives post-zygotiques peuvent s'établir et être conservés par la sélection naturelle. Une réponse à cette question a été donnée de manière indépendante par trois évolutionnistes, William Bateson, Theodosius G. Dobzhansky, et Hermann J. Muller (Bateson, 1909; Dobzhansky, 1936; Muller, 1942). Leur modèle explique que les barrières d'isolement reproductif évoluent graduellement dans le temps, étant le résultat de l'accumulation de changements génétiques, adaptatifs ou neutres, dans la population. Si deux facteurs en interactions épistatique dans un ancêtre commun suivent chacun une divergence génétique au cours du temps, les nouveaux allèles établis dans différentes populations peuvent être incompatibles, et leur combinaison dans les hybrides avoir des effets délétères (voir Figure I-6 A pour plus de détail). L'évolution des deux gènes peut avoir eu lieu dans une seule des deux populations divergentes, l'autre population conservant les allèles ancestraux (Coyne et Orr, 2004; Cattani et Presgraves, 2009) conduisant au même type d'incompatibilité (Figure I-6 B).

Ce concept d'incompatibilités d'allèles en interactions épistatiques est généralement connu sous le terme « modèle d'incompatibilité BMD » (d'après les noms de ses auteurs), et constitue le modèle explicatif le plus accepté de l'origine et du fonctionnement des barrières reproductives post-zygotiques. Le modèle BDM implique des interactions entre deux locus au moins. Cependant, comme certains auteurs le suggèrent (Bomblies et Weigel, 2007), les incompatibilités peuvent résulter de l'évolution d'un seul locus (Figure I-6 C). Il a été montré chez le riz que deux allèles incompatibles d'un seul locus peuvent coexister, en présence d'un allèle neutre compatible avec chacun des deux allèles incompatibles (Chen *et al.*, 2008).

Récemment, il a été montré chez *Arabidopsis* que l'évolution divergente de gènes dupliqués peut mener à des différences d'expression ou du silençage (*silencing*) des gènes paralogues, de manière différente dans chaque population. Une fois le phénomène est établie, une interaction épistatique peut se produire entre les allèles des ces gènes dans les hybrides, générant des barrières reproductives selon le modèle BDM (Bikard *et al.*, 2009). L'accumulation de différentes incompatibilités de type BDM agissant comme barrières

reproductives, peut conduire à l'isolement reproductif total entre deux populations (Coyne et Orr, 2004).

1.6.4 Distorsion de la ségrégation

La transmission des allèles provenant d'un individu hétérozygote (ex. un hybride) peut dévier significativement des proportions mendéliennes attendues. Ce phénomène, appelé Distorsion de la Ségrégation (DS) ou « Transmission Ratio Distortion » (TRD) est caractérisé par une surreprésentation d'un des allèles dans la population.

Chez les plantes, différents mécanismes peuvent aboutir à une déviation des fréquences alléliques finales par rapport aux fréquences attendues (i.e. mendéliennes) dans une population. La sélection conspécifique du pollen au moment de la fécondation (Fishman *et al.*, 2008), la ségrégation non aléatoire des chromosomes pendant la méiose (Meiotic Drive) (Fishman et Willis, 2005), ou des phénomènes postérieurs à la méiose, comme l'avortement sélectif des gamètes (Koide *et al.*, 2008c), ou l'avortement des zygotes (embryon, endosperme ou plantules) (Ichitani *et al.*, 2007), sont les causes principales de la TRD. L'interaction épistatique entre les allèles présents chez un hybride et les possibles incompatibilités existantes (de type BDM) peuvent donc produire un phénomène de TDR.

Au niveau génétique, plusieurs systèmes qui causent une DS ont été décrits, tous pour des modèles animaux. Chez la souris, par exemple, un complexe de gènes localisés sur une région de 20 cM du chromosome 17 ségrégent en bloc du fait de quatre inversions et causent une TRD. Dans le sperme d'un individu hétérozygote pour ce complexe, appelé l'haplotype-t, la mobilité des gamètes portant l'allèle sauvage est réduite à cause de l'action en *trans* des gènes *Tdc* (Trans-acting t-complex distorters) situés dans la région. Cependant, les gamètes qui portent la totalité de l'haplotype-t récupèrent la mobilité grâce à l'action en *cis* du gène *Tcr* (t-complex responder). *Tcr* fut identifié comme un gène résultant de la fusion entre deux kinases, qui s'exprime tardivement au cours de la spermatogenèse (Schimenti, 2000).

Un autre système chez *D. melanogaster* implique un gène codant pour une protéine de transport nucléaire incomplète, qui est responsable du Meiotic Drive et de la distorsion résultante (Merrill *et al.*, 1999). Egalement chez la Drosophile, un gène qui cause la DS des chromosomes sexuels (sex-ratio) et son suppresseur furent aussi caractérisés. Le gène de la surreprésentation du chromosome X dans le sperme (*Dox*) est un ARN non codant, tandis que son suppresseur (*Nmy*) est un gène dérivé du premier après une transposition, et qui cause son silençage par interférence d'ARN (RNAi) (Tao *et al.*, 2007a; Tao *et al.*, 2007b).

Le lien direct entre stérilité hybride et distorsion de la ségrégation fut montré en utilisant des hybrides issus de croisements entre *D. pseudooscura* ssp. *pseudooscura* (USA) et *D. pseudooscura* ssp. *bogotana* (Bog). L'interaction épistatique entre les allèles « Bog » de

plusieurs gènes localisés sur le chromosome X avec les allèles USA dominants d'autres gènes localisés sur le deuxième et le troisième autosome cause la stérilité des hybrides (Orr et Irving, 2005). Un des gènes localisé sur le chromosome X (*Overdrive-Ovd*) fut caractérisé comme la cause du sex-ratio et de la stérilité des mâles hybrides (Phadnis et Orr, 2009).

Chez les plantes, l'association entre la DS et les locus de stérilité pollinique fut démontrée pour la tomate (Moyle et Graham, 2006), indiquant que la stérilité peut être la conséquence directe de l'élimination sélective des gamètes. Les allèles des gènes qui causent ce phénomène sont alors « égoïstes », car ils sont à l'origine de l'élimination des gamètes qui portent l'allèle opposé. La DS peut donc contribuer à la fixation des barrières reproductives dans les populations, en altérant les fréquences alléliques dans les populations (Coyne et Orr, 2004).

I.6.5 Barrières reproductives décrites chez le riz

Chez le riz, plusieurs barrières d'isolement reproductif appartenant aux différentes catégories commentées plus haut ont été décrites. Pour les espèces à génome non AA, le mécanisme d'isolement reproductif le plus documenté est basée sur les différences structurales dans les chromosomes, principalement des inversions et des translocations survenues au cours de l'évolution des espèces. Ces réarrangements empêchent l'appariement chromosomique et la recombinaison normale dans les hybrides, et les rendent stériles. Pour les espèces à génome AA, une large gamme de mécanismes ont été décrits au cours de ces dernières années.

Au niveau pré-zygotique, la faible adhérence du pollen au stigmate, les problèmes de croissance du tube pollinique, et les problèmes de la fertilisation dus à la faible affinité entre les gamètes, peuvent réduire le succès dans la production d'hybrides entre les espèces du groupe d'espèces à génome AA (Liu *et al.*, 2004a). La préférence pour le pollen provenant de la même espèce a été aussi décrite comme source de barrière reproductrice chez le riz, entre *O. sativa* et son ancêtre *O. rufipogon* (Zhiping et al., 2002). Au niveau post-zygotique, différents groupes de recherche ont décrit de nombreux locus de létalité et de stérilité des hybrides, ainsi que d'*hybrid breakdown*. Une liste exhaustive des gènes qui causent ces barrières post-zygotiques chez le riz est présentée dans le Tableau I-9 (voir pages suivantes).

Les barrières de reproductions les plus étudiés chez le riz, concernent d'une part les sous-espèces *indica* et *japonica*, et d'autre part les espèces *O. sativa* et *O. glaberrima*. Elles résultent des actions conjointes de plusieurs locus. Dans le premier cas, la stérilité des hybrides inter-subspécifiques peut être en grande partie expliquée par la présence du locus S_5 , qui cause une stérilité femelle, et l'action additive des autres locus de stérilité mâle présents selon le croisement effectué. Des allèles neutres (S_5^n) chez des variétés dites de compatibilité large (WCV, pour Wide Compatibility Variety) permettent l'obtention d'hybrides plus

Tableau I-9. Gènes impliqués dans les barrières reproductives post-zygotiques décrites chez le riz.

	Locus	Chromosome	Action	Référence bibliographique
Létalité hybride				
<i>O. sativa</i> x <i>O. rufipogon</i>	<i>Cif, Cim</i>	6	AE, AZ	(Matsubara <i>et al.</i> , 2003; Koide <i>et al.</i> , 2008b)
<i>O. glumaepatula</i> x <i>O. sativa</i>	<i>Rhw</i>	8	W	(Sobrizal <i>et al.</i> , 2003)
<i>O. glaberrima</i> x <i>O. barthii</i>	<i>hwb1/W1, - hwb2/W2</i>		W	(Chu et Oka, 1972)
Stérilité hybride				
<i>O. sativa</i> <i>indica</i> x <i>japonica</i>	<i>S₅[§] / S₂₆</i>	6	SE	(Ji <i>et al.</i> , 2005; Qiu <i>et al.</i> , 2005) (Kubo et Yoshimura, 2001)
	<i>S₇</i>	7	SE	(Yanagihara <i>et al.</i> , 1992; Wan et Ikehashi, 1996)
	<i>S₈</i>	6	SE	(Wan et Ikehashi, 1996)
	<i>S₉</i>	4	SE	(Zhu <i>et al.</i> , 2005b; Zhao <i>et al.</i> , 2006)
	<i>S₁₀</i>	6	P, SE	(Sano <i>et al.</i> , 1994; Zhu <i>et al.</i> , 2005a)
	<i>S₁₁</i>	11	P, SE	(Sawamura et Sano, 1996)
	<i>S₁₅</i>	12	SE	(Wan et Ikehashi, 1996; Wan <i>et al.</i> , 1996)
	<i>S₁₆</i>	1	SE	(Wan et Ikehashi, 1996)
	<i>S₁₇</i>	12	SE	(Wan et Ikehashi, 1996; Wan <i>et al.</i> , 1998)
	<i>S₂₄</i>	5	P, SE	(Kubo <i>et al.</i> , 2000; Kubo <i>et al.</i> , 2008)
	<i>S₂₅</i>	12	P	(Kubo <i>et al.</i> , 2001)
	<i>S₂₉[†]</i>	2	E	(Zhu <i>et al.</i> , 2005b)
	<i>S₃₀</i>	7	E	(Zhu <i>et al.</i> , 2005a)
	<i>S₃₁</i>	5*	E	(Zhao <i>et al.</i> , 2007)
	<i>S₃₂</i>	2*	E	(Li <i>et al.</i> , 2007a)
	<i>S₃₃[†]</i>	3*	P	(Jing <i>et al.</i> , 2007)
	<i>S₃₄[†]</i>	11	P	(Jing <i>et al.</i> , 2007)
	<i>S₃₅</i>	1	P	(Kubo <i>et al.</i> , 2008)
	<i>Sa[§]</i>	1	SE	(Long <i>et al.</i> , 2008)
	<i>Sb</i>	5*	P	(Li <i>et al.</i> , 2006)
	<i>Sc</i>	3*	P	(Chen <i>et al.</i> , 2004)
	<i>Sd</i>	1*	P	(Li <i>et al.</i> , 2008c)
	<i>Se</i>	-	-	(Zhang et Lu, 1996)
	<i>Sf</i>	-	-	(Zhang et Lu, 1996)
	<i>f1</i>	1	WCV	(Wang <i>et al.</i> , 1998)
	<i>f3</i>	3	WCV	(Wang <i>et al.</i> , 1998)
	<i>f5/ pf5/spf5</i>	5*	WCV P, SE	(Wang <i>et al.</i> , 2006) (Song <i>et al.</i> , 2005)
	<i>f8/ spf8</i>	8	WCV P, SE	(Wang <i>et al.</i> , 1998) (Song <i>et al.</i> , 2005)
	<i>pss1</i>	8*	P	(Li <i>et al.</i> , 2007b)
	<i>pfl2</i>	12	P	(Song <i>et al.</i> , 2005)
	<i>qSS-2</i>	2	E	(Wang <i>et al.</i> , 2005)

Tableau I-9. (Suite)

	Locus	Chromosome	Action	Référence bibliographique
	<i>qSS-8a</i>	8	E	(Wang <i>et al.</i> , 2005)
	<i>qSS-8b</i>	8	E	(Wang <i>et al.</i> , 2005)
	<i>gal</i>	6	GA	(Iwata <i>et al.</i> , 1964)
	<i>ga2</i>	3	GA	(Nakagahara, 1972)
	<i>ga3</i>	3	GA	(Nakagahara, 1972)
	<i>ga4</i>	6	GA	(Mori <i>et al.</i> , 1973)
	<i>ga5</i>	6	GA	(Mori <i>et al.</i> , 1973)
	<i>ga6</i>	4	GA	(Maekawa, 1982)
	<i>ga7</i>	1	GA	(Maekawa et Kita, 1985)
	<i>ga8</i>	1	GA	(Maekawa et Kita, 1985)
	<i>ga9</i>	1	GA	(Maekawa et Kita, 1985)
	<i>gal0</i>	4	GA	(Kinoshita et Takamure, 1984)
	<i>gal1</i>	7	GA	(Lin et Ikehashi, 1993)
	<i>gal2</i>	4	GA	(Kinoshita, 1993)
	<i>gal3</i>	12	GA	(Rha <i>et al.</i> , 1994)
	<i>gal4</i>	3	GA	(Lu <i>et al.</i> , 2000)
<i>O. sativa</i> x <i>O. glaberrima</i>	<i>S₁</i>	6*	P, SE	(Sano, 1990; Koide <i>et al.</i> , 2008c)
	<i>S₂</i>	-	P, SE	(Sano, 1983; Sano, 1986)
	<i>S₃</i>	11	P, SE	(Sano, 1983; Sano, 1986)
	<i>S₄</i>	6	P, SE	(Sano, 1986)
	<i>S₁₈</i>	10	P	(Doi <i>et al.</i> , 1998a; Doi <i>et al.</i> , 2003b)
	<i>S₁₉ /</i> <i>qSS-3</i>	3	P, SE	(Taguchi <i>et al.</i> , 1999; Doi <i>et al.</i> , 2003b) (Li <i>et al.</i> , 2008b)
	<i>S₂₀ /</i> <i>qSS-7a</i>	7	P	(Doi <i>et al.</i> , 1999; Doi <i>et al.</i> , 2003b) (Li <i>et al.</i> , 2008b)
	<i>S₂₁</i>	7	P	(Doi <i>et al.</i> , 2003b)
	<i>S_{29(t)†} /</i> <i>qSS-2</i>	2	P	(Hu <i>et al.</i> , 2006) (Li <i>et al.</i> , 2008b)
	<i>S_{33(t)†}</i>	1	P, SE	(Ren <i>et al.</i> , 2005)
	<i>S_{34(t)†}</i>	3	P	(Zhang <i>et al.</i> , 2005)
	<i>qSS-1</i>	1	P	(Li <i>et al.</i> , 2008b)
	<i>qSS-7b</i>	7	P	(Li <i>et al.</i> , 2008b)
<i>O. sativa</i> x <i>O. glumaepatula</i>	<i>S₁₂</i>	-	P	(Sano, 1994)
	<i>S₂₂</i>	2	P	(Sobrizal <i>et al.</i> , 2000a)
	<i>S₂₃</i>	7	P	(Sobrizal <i>et al.</i> , 2000b)
	<i>S₂₇</i>	8	P	(Sobrizal <i>et al.</i> , 2001a)
	<i>S₂₈</i>	4	P	(Sobrizal <i>et al.</i> , 2002)
<i>O. sativa</i> x <i>O. rufipogon</i>	<i>S₆</i>	6	P, SE	(Koide <i>et al.</i> , 2008a)
	<i>S_{21ruf}</i>	7	P	(Miyazaki <i>et al.</i> , 2007)
<i>O. sativa</i> x <i>O. longistaminata</i>	<i>S₁₃</i>	1	P	(Taneichi <i>et al.</i> , 2005)

Tableau I-9. (Suite)

	Locus	Chromosome	Action	Référence bibliographique
Hybrid breakdown				
<i>O. sativa</i>	<i>hwa1</i> ,	-		(Oka, 1957b)
<i>indica</i> x <i>japonica</i>	<i>hwa2</i>	-		
	<i>hwc1</i> ,	1*		(Ichitani <i>et al.</i> , 2007)
	<i>hwc2</i>	4*		(Kuboyama <i>et al.</i> , 2009)
	<i>hwd1</i> ,	10		(Fukuoka <i>et al.</i> , 1998)
	<i>hwd2</i>	7		
	<i>hwe1</i> ,	12		(Kubo et Yoshimura, 2002)
	<i>hwe2</i>	2		
	<i>hwg1</i> ,	6		(Fukuoka <i>et al.</i> , 2005)
	<i>hwg2</i>	11		
	<i>hwh1</i> ,	2		(Jiang <i>et al.</i> , 2008)
	<i>hwh2</i> [§] /	11		
	<i>hbd2</i> ,			(Yamamoto <i>et al.</i> , 2007)
	<i>hbd3</i>			
	<i>hbd1</i> [†]	2		(Matsubara <i>et al.</i> , 2007)
	<i>hsa1</i> ,	12		(Kubo et Yoshimura, 2005)
	<i>hsa2</i> ,	8		
	<i>hsa3</i>	9		
	<i>S</i>	12		(Kubo et Yoshimura, 1998)
<i>O. sativa</i> x <i>O. glumaepatula</i>	<i>hwf1</i> (t)	4		(Sobrizal <i>et al.</i> , 2001b)
<i>O. sativa</i> x <i>O. nivara</i>	<i>hbd1</i> (t) [†]	2*		(Miura <i>et al.</i> , 2008)

† Locus détectés dans différents croisements interspécifiques, de localisations différentes mais portant le même nom.

§ Locus pour lesquels le gène correspondant a été identifié.

* Locus cartographiés finement.

AE, Avortement de l'endosperme.

AZ, Avortement du zygote.

W, Faiblesse des hybrides

P, Stérilité pollinique.

GA, Effet décrit comme sélection gamétophytique.

SE, Stérilité du sac embryonnaire.

E, Stérilité des épillets ; l'origine mâle, femelle ou les deux est inconnue. On ne sait pas non plus s'il s'agit plutôt d'une diminution de l'affinité au cours de l'union des gamètes (Song *et al.*, 2005).

WCV « Wide Compatibility Variety », Allèle provenant d'une variété de large compatibilité.

fertiles, et se sont révélés très utiles pour l'amélioration variétale (Liu *et al.*, 1997). Dans le cas des hybridations entre *O. sativa* et *O. glaberrima*, la forte stérilité des hybrides est aussi la résultante de l'action accumulée de plusieurs locus. Parmi ceux-ci, le locus *S_I* a l'effet le plus marqué de par son action sur les gamètes mâles et femelles (Sano, 1990). Aucun allèle neutre ou à meilleure compatibilité n'a été trouvé pour ce locus.

Les barrières reproductives post-zygotiques chez le riz peuvent être accompagnées de DS plus ou moins prononcée. Un criblage pour rechercher des régions du génome présentant des DS réalisé pour 5 croisements inter-subspécifiques et interspécifiques différents (Xu *et al.*, 1997) révéla qu'au moins 17 locus montrent une ségrégation non mendélienne, dont six sont partagées par au moins trois des croisements. Les études menées par Harushima révèlent près d'une trentaine de régions génomiques avec DS pour chacun des trois croisements *indica-japonica* évalués (Harushima *et al.*, 2002). La plus forte distorsion, commune à plusieurs études, est localisée sur le bras court du chromosome 3 (Harushima *et al.*, 1996; Xu *et al.*, 1997; Harushima *et al.*, 2002; Matsushita *et al.*, 2003) où les gènes gamétophytiques *ga2* et *ga3* ont été décrits (Nakagahara, 1972). En ce qui concerne les croisements entre *O. sativa* et *O. glaberrima*, une forte distorsion est systématiquement présente sur le bras court du chromosome 6 (Doi *et al.*, 1998b; Lorieux *et al.*, 2000; Aluko *et al.*, 2004), précisément au locus *S_I* (Sano, 1990).

Comme mentionné préalablement, de tous les locus de stérilité hybride chez le riz, seuls deux d'entre eux, *Sa* et *S₅*, ont été pleinement identifiés (Chen *et al.*, 2008; Long *et al.*, 2008). De même, seul un couple de gènes candidats (*hwh1* and *hwh2*) a été identifié comme étant une cause d'*hybrid breakdown* (Jiang *et al.*, 2008). Ces gènes, avec les trois gènes d'*Arabidopsis* de létalité hybride, constituent les seuls cas de gènes de barrières reproductives pleinement identifiés chez les plantes.

Malgré l'existence de ces barrières reproductives sur le riz, les flux de gènes sont encore existant entre les différentes espèces du genre (Song *et al.*, 2003; Kuroda *et al.*, 2005). En outre, il a été possible de retrouver des individus hybrides naturels qui conservent leur fertilité. Ces individus proviennent des endroits où les différentes espèces co-existent en sympatrie (qui partagent le même habitat). Cela a été le cas, par exemple, pour les individus hybrides dérivés d'*O. longistaminata* (Ghesquiere, 1988), *O. glaberrima* (Semon *et al.*, 2005; Barry *et al.*, 2007), et les populations entre les deux sous-espèces de *O. sativa* (Garris *et al.*, 2005).

I.7 Importance de l'étude des barrières d'isolement reproductif pour le développement de stratégies d'amélioration

Nous avons vu que les espèces apparentées à *O. sativa* constituent un réservoir allélique au potentiel très intéressant pour l'amélioration du riz. Les nombreux gènes déjà introduits dans les variétés élités, témoignent de l'utilité des approches d'amélioration basées sur les croisements interspécifiques. Même si la production de lignées interspécifiques de type NIL (Near-Isogenic Lines) ou CSSL a beaucoup aidé à la identification et à la cartographie des caractères portés par les espèces sauvages, le chemin jusqu'à leur introgression dans *O. sativa* reste difficile, car souvent les barrières reproductives agissent encore dans les populations avancées, dû à leur particularité d'« égoïstes ». Ces barrières exercent une force naturelle qui va en contre des hybridations, en empêchant la introduction facile des caractères vers *O. sativa*.

On comprend donc l'importance de l'étude des barrières reproductives, et en particulier des barrières post-zygotiques entre les espèces du genre *Oryza* : la connaissance de la localisation exacte des facteurs impliqués dans les barrières reproductives, permet de les contourner efficacement en suivant les allèles compatibles à l'aide de marqueurs moléculaires durant le développement des populations. D'autre part, l'identification des facteurs impliqués, de leurs interactions, et de leur mode d'action, doit renseigner sur le fonctionnement moléculaire des barrières de stérilité, sur leur mode d'apparition et d'évolution, et sur leur contribution à la différenciation des espèces.

Développer des stratégies de contournement des barrières de reproduction permettra à la l'amélioration des plantes et à la riziculture en particulier de mieux profiter du potentiel génétique contenu dans ses espèces apparentées.

I.7.1 Ponts interspécifiques – iBridges

Les lignées NERICA sont un bon exemple d'hybridation interspécifique conduisant à la production de variétés cultivées aux caractéristiques adaptées à un environnement spécifique. Toutefois, ces lignées ne représentent qu'une faible proportion du potentiel génétique de l'espèce *O. glaberrima*, étant dérivées à partir de seulement deux accessions de l'espèce africaine. De plus, elles ne représentent qu'une partie du génome d'*O. glaberrima* (8% en moyenne par lignée) (Semagn *et al.*, 2007; Ndjondjop *et al.*, 2008). De ce fait, la quantité de gènes qui peuvent être utilisées dans les programmes d'amélioration reste limitée. Enfin, même si la fertilité des hybrides *O. sativa* x NERICA semble plus élevée que celle des hybrides *O. sativa* x *O. glaberrima* (Ikeda *et al.*, 2009), le problème majeur de l'utilisation de

lignées d'introgression *O. sativa* x *O. glaberrima* réside dans le fait que, généralement, l'allèle *glaberrima* du locus *S₁* reste largement favorisé par rapport à l'allèle *sativa*, quelle que soit la génération de rétrocroisement. Ce déséquilibre allélique entraîne une forte stérilité pollinique des hybrides entre ces lignées d'introgression et les lignées receveuses *O. sativa* et rend le travail de sélection pratiquement impossible.

On aimerait donc pouvoir développer des lignées d'introgression *O. sativa* x *O. glaberrima* qui produiraient des hybrides fertiles avec *O. sativa*, et qui conserveraient une proportion maximale du génome d'*O. glaberrima*. C'est précisément l'objectif du projet iBridges, (*interspecific bridges*) proposé par l'IRD et financé entre autres par le Generation Challenge Program, et qui vise à produire plusieurs populations interspécifiques, partiellement fertiles en croisement avec *O. sativa* et dérivées à partir d'une vingtaine d'accessions donneuses d'*O. glaberrima* représentant la diversité génétique et phénotypique de l'espèce. Le projet est subdivisé en quatre actions: (1) La sélection d'un ensemble de lignées qui représentent bien le potentiel et la diversité d'*O. glaberrima*. (2) La localisation et la caractérisation du locus majeur de stérilité des hybrides entre les deux espèces, le locus *S₁*. (3) La génération d'un ensemble de populations interspécifiques BC₁F₁ entre les lignées d'*O. glaberrima* sélectionnées et trois cultivars d'*O. sativa*. Parmi ces individus, ceux qui ne possèdent pas l'allèle *glaberrima* au niveau du locus *S₁* –et qui sont donc potentiellement compatibles avec *O. sativa*– sont identifiés à l'aide de marqueurs moléculaires liés au locus, puis testés pour leur fertilité et avancés jusqu'à la génération BC₁F₄. (4) Le génotypage à l'échelle du génome entier des lignées BC₁F₄ obtenues par marqueurs SFP (Single Feature Polymorphism).

Le Tableau I-10 (voir dans la page suivante) résume les résultats obtenus en terme d'identification des lignées iBridges candidates. Le présent travail de thèse correspond à l'action 2 du projet iBridges.

Tableau I-10. Croisements interspécifiques du projet iBridges

Accession <i>O. glaberrima</i>	Origine géographique	Caractères d'intérêt/ Ressources disponibles	Accession d' <i>O. sativa</i> pour croisement		
			IR64	Curinga	IR64
CG14	Casamance	Résistance à <i>M. oryzae</i> et nématodes. Librairie de BAC Parent des NERICA	WARDA	CIAT	WARDA
CG17	Casamance	Tolérance à la sécheresse		WARDA	
DC Kono	Sierra Leone	Tolérance à la sécheresse			IER
IG02	Côte d'Ivoire				WARDA
MG12	Mali	Librairie de BAC Parent population CSSL	CIAT	CIAT	CIAT
RAM111	Mali	Parent populations interspécifiques			INERA
RAM24	Mali	Parent populations interspécifiques			
RAM48	Mali	Parent populations interspécifiques			
RAM55	Mali	Parent populations interspécifiques			IER
RAM63	Mali	Parent populations interspécifiques			
RAM90	Mali	Parent populations interspécifiques	Punjab Uty		
RAM95	Mali	Résistance à <i>M. oryzae</i>			
Saliforeh	Sierra Leone	Tolérance à la sécheresse	PhilRice		
TOG5486	Nigéria	Tolérance à la sécheresse	Punjab Uty		
TOG5666	Nigéria	Tolérance à la sécheresse			
TOG5672	Nigeria	Résistance au RYMV et <i>M. oryzae</i>		CIAT	
TOG5674	Nigéria	Résistance au RYMV			
TOG5681	Nigéria	Librairie de BAC, Parent population CSSL et NERICAS	CIAT	CIAT	CIAT
TOG5882	Libéria	Résistance à <i>Orseolia</i> <i>oryzivora</i> Harris et Gagné	Punjab Uty		INERA
TOG5980	Nigéria	Tolérance à la sécheresse			
TOG6208	Zimbabwe	Tolérance à la sécheresse	PhilRice		
TOG6211	Guinée Bissau	Tolérance à la sécheresse			CIAT
TOG6221	Burkina Faso	Tolérance à la sécheresse	PhilRice		
TOG6308	Libéria	Tolérance à la sécheresse		IER	
TOG7106	Nigéria	Résistance à <i>O. oryzivora</i>	Punjab Uty	IER	

Certains des croisements réalisés au CIAT se font en collaboration avec Fedearroz (Colombie). (Punjab Uty) Punjab Agricultural University, Ludhiana, India. (INERA) Institut de l'Environnement et des Recherches Agricoles, Burkina Faso. (IER) Institut d'Economie Rurale, Mali. (PhilRice) The Philippine Rice Research Institute, Philippines. (WARDA) AfricaRice.

I.8 Présentation du sujet de recherche

Le présent travail se situe dans le cadre du projet iBridges décrit au paragraphe précédent, et qui vise à la création de lignées interspécifiques *O. sativa* x *O. glaberrima*, capables de produire des hybrides fertiles avec *O. sativa*. L'objectif général est de contribuer à déterminer les bases génétiques de la barrière reproductive qui sépare les deux espèces de riz cultivé, *O. sativa* et *O. glaberrima*. La connaissance générée devrait permettre d'établir des stratégies visant à faciliter le transfert de gènes d'*O. glaberrima* vers *O. sativa* et ainsi de mieux exploiter la diversité génétique utile de l'espèce africaine de riz. L'objectif principal fixé pour notre travail de thèse est l'identification et la caractérisation du locus qui exerce l'effet le plus sévère sur la fertilité des hybrides, le locus *S₁*.

Dans ce but, différentes approches de génétique et de génomique ont été utilisées tout au long de l'étude.

(1) Tout d'abord, il a été nécessaire de développer un outil fiable pour la construction et la comparaison de cartes génétiques de différents croisements interspécifiques chez le riz. Cet outil, appelé carte génétique universelle du riz, est présenté dans le suivant chapitre de ce document.

(2) Dans le troisième chapitre, nous abordons le sujet principal d'étude, à savoir la barrière de reproduction entre *O. sativa* et *O. glaberrima* gouvernée par le locus *S₁*. Nous présentons la construction d'une carte génétique en utilisant la carte universelle, la cartographie fine et physique du locus *S₁*, la description d'un modèle explicatif d'élimination gamétique et la caractérisation de la séquence de la région chromosomique autour de *S₁*.

(3) Le quatrième chapitre traite de l'analyse approfondie d'une région de 800 kb qui contient *S₁* et les facteurs putatifs du modèle de stérilité hybride, par une approche de génomique comparative.

(4) Le cinquième chapitre présente une étude visant à tester l'universalité du système de stérilité *S₁* par l'analyse d'une collection d'accèsions d'*O. glaberrima*.

(5) La discussion finale synthétise l'ensemble des résultats obtenus et analyse leur contribution à une meilleure connaissance des barrières reproductives du riz et à l'établissement de stratégies d'amélioration des plantes basées sur les croisements interspécifiques. Enfin, nous tenterons de dégager des perspectives d'études et de méthodologies possibles qui permettraient d'aller plus loin dans la compréhension des mécanismes de stérilité chez le riz et en particulier entre les deux espèces de riz cultivé.

II

UNE CARTE GENETIQUE UNIVERSELLE POUR LE RIZ

Introduction

Nous avons vu dans l'introduction générale que de nombreuses populations interspécifiques ont été développées et utilisées pour construire des cartes génétiques par différents groupes de recherche (Voir par exemple Causse *et al.*, 1994; Doi *et al.*, 2003a; Lee *et al.*, 2005). Ces cartes génétiques utilisent des systèmes ou des ensembles de marqueurs différents, ce qui rend leur comparaison parfois complexe, rendant notamment difficile la recherche de coïncidences entre les QTL détectés sur ces différentes cartes. Dans le cadre des activités du projet international « Exploring Natural Genetic Variation: Developing Genomic Resources and Introgression Lines for Four AA Genome Rice Relatives » (http://www.generationcp.org/vw/Download/Competitive_Grant_Proposals/10_TOHME.pdf) dont l'objectif principal était de générer des populations interspécifiques à partir de cinq espèces à génome AA, il était essentiel de se doter d'un outil facilitant la comparaison entre les cartes obtenues pour chacun des croisements. Le concept d'une carte génétique universelle du riz ou « Universal Core Genetic Map » (UCGM) a été développé. L'UCGM se compose d'un ensemble d'*ancres* qui regroupent de 3 à 5 marqueurs SSR étroitement liés entre eux et qui définissent dans la pratique un même locus, du moins à l'échelle de travail d'une carte génétique peu saturée. Ces ancres sont distribuées régulièrement sur le génome. L'avantage de définir et utiliser de telles ancres plutôt que des marqueurs uniques pour définir les locus de la carte réside dans le fait que la probabilité d'observer un polymorphisme entre deux accessions prises au hasard est bien supérieure pour chaque locus. En effet, il suffit qu'un des marqueurs d'une ancre soit polymorphe pour que l'ancre soit définie comme polymorphe. Ma contribution au développement de la carte universelle a consisté à réaliser la sélection et évaluations préliminaires des marqueurs qui la définissent. Par la suite, les données d'évaluation par marqueurs moléculaires pour la construction de la carte génétique du croisement *O. sativa* x *O. glaberrima* IR64/TOG5681//IR64 ont servi à valider l'UCGM, avec d'autres données acquises à partir de rétrocroisements *O. sativa* x *O. rufipogon* et *O. sativa* x *O. meridionalis* avec *O. sativa*.

Nous présentons ici un article qui expose le principe et les données de la carte génétique universelle du riz. Cet article vient de paraître dans la revue *Theoretical and Applied Genetics* (Orjuela *et al.*, 2009)

A universal core genetic map for rice

**Julie Orjuela^{1,**} • Andrea Garavito^{1,2} • Matthieu Bouniol^{1,2} • Juan David Arbelaez¹ •
Laura Moreno¹ • Jennifer Kimball³ • Gregory Wilson³ • Jean-François Rami⁴ •
Joe Tohme¹ • Susan R McCouch³ • Mathias Lorieux^{1,2,**,*}**

¹ Agrobiodiversity and Biotechnology Project, International Center for Tropical Agriculture (CIAT), A.A. 6713, Cali, Colombia.

² Plant Genome and Development Laboratory (UMR 5096 – CNRS – IRD – UPVD), Institut de Recherche pour le Développement (IRD), 911 Av. Agropolis, 34394 Montpellier Cedex 5, France.

³ Department of Plant Breeding and Genetics, Cornell University, 162 Emerson Hall, Ithaca, NY 14853-1901, USA.

⁴ Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), UMR DAP, TA A-96/03, Avenue Agropolis, Montpellier, France.

*Corresponding author. E-mail: m.lorieux@cgiar.org. Tel: +33 4 67 41 61 53. Fax: +33 4 67 41 61 81

**Present address : Plant Genome and Development Laboratory (UMR 5096 – CNRS – IRD – UPVD), Institut de Recherche pour le Développement (IRD), 911 Av. Agropolis, 34394 Montpellier Cedex 5, France.

Abstract To facilitate the creation of easily comparable, low-resolution genetic maps with evenly distributed markers in rice (*Oryza sativa* L.), we conceived of and developed a Universal Core Genetic Map (UCGM). With this aim, we derived a set of 165 *anchors*, representing clusters of three microsatellite or simple sequence repeat (SSR) markers arranged into non-recombining groups. Each anchor consists of at least three, closely linked SSRs, located within a distance below the genetic resolution provided by common, segregating populations (<500 individuals). We chose anchors that were evenly distributed across the rice chromosomes, with spacing between 2 and 3.5 Mbp (except in the telomeric regions, where spacing was 1.5 Mbp). Anchor selection was performed using *in silico* tools and data: the *O. sativa* cv. Nipponbare rice genome sequence, the CHARM tool, information from the Gramene database and the OrygenesDB database. Sixteen AA-genome accessions of the *Oryza* genus were used to evaluate polymorphisms for the selected markers, including accessions from *O. sativa*, *O. glaberrima*, *O. barthii*, *O. rufipogon*, *O. glumaepatula* and *O. meridionalis*. High levels of polymorphism were found for the tested *O. sativa* × *O. glaberrima* or *O. sativa* × wild rice combinations. We developed Paddy Map, a simple database that is helpful in selecting optimal sets of polymorphic SSRs for any cross that involves the previously mentioned species. Validation of the UCGM was done by using it to develop three interspecific genetic maps and by comparing genetic SSR locations with their physical positions on the rice pseudomolecules. In this study, we demonstrate that the UCGM

is a useful tool for the rice genetics and breeding community, especially in strategies based on interspecific hybridisation.

Introduction

Integrating information about gene and quantitative trait locus (QTL) locations on different genetic maps has been a constant issue since molecular markers were discovered. One of the main difficulties lies in the fact that maps of the populations under study do not share the same set of markers. Hence, mathematical projections are the only methods of comparing genetic maps that were constructed using different marker sets. However, this is not an optimal method, considering the variation of recombination frequency and the difference in marker density between populations. In the case of rice, the situation has greatly improved since the genomes of the Nipponbare and 93-11 cultivars have been sequenced (IRGSP 2005; Wu et al. 2002; Yu et al. 2002). It is now possible to accurately compare two genetic maps, looking at the physical positions of the markers in the Nipponbare and 93-11 pseudomolecules. Public databases such as Gramene ([http:// www.gramene.org](http://www.gramene.org)), Oryzabase (<http://www.shigen.nig.ac.jp/rice/oryza-base/top/top.jsp>) or Rice-BRCdb (<http://rice-brcdb.cines.fr>) are tools that make it easier to compile QTL/gene location information on the rice physical map. However, some drawbacks still remain with this approach when developing a new genetic map: (i) finding a set of polymorphic markers that are evenly dispersed on the genome still requires a fair amount of work; (ii) direct comparison between QTL locations on different genetic maps is still cumbersome.

Microsatellites are repeats of short DNA motifs (1-6 bp in length) arranged in tandem, exhibiting high variability in the number of repeats at a given locus. These simple sequence repeats (SSR) are a valuable source of genetic markers (Temnykh et al. 2001). They are codominant and highly polymorphic, making them ideal markers for use in a wide variety of parental combinations. In rice, a high-density SSR map with genome coverage of approximately 2 SSRs per centimorgan cM was reported by the International Rice Microsatellite Initiative (IRMI) (McCouch et al. 2002). This map can be exploited for developing markers that are tightly linked to traits of interest. Subsequently, a total of 18,828 Class 1 (di-, tri- and tetra-nucleotide) SSRs, representing 47 distinctive motif families, were identified and annotated from the complete sequence of the rice genome (IRGSP 2005). More recently, a set of 52,000 SSRs polymorphic between *indica/japonica* were reported (Zhang et al. 2007) (however, the URL provided in the publication does not appear to be functional and we could not obtain information from the authors). More than 80% of the primers designed

for *O. sativa* successfully amplify in other AA-genome species (Lorieux et al. 2000), and find immediate application in the field of diversity and genomic studies.

The genomic information gathered about rice during the past 15 years makes it possible to optimise genetic mapping strategies, which in turn facilitates the search for genes or QTLs that control traits of interest for breeding. A genetic linkage map is a linear representation of the positions of various loci (e.g. markers, genes) along chromosomes. Distances between loci depend on the recombination rates observed in a particular segregating population. Hence, the distance between two specific loci may vary considerably from one population to another. This makes it difficult to compare QTL/gene locations between different genetic linkage maps. Moreover, genetic maps built on sets of “core” markers frequently suffer from poor genome saturation, particularly in chromosomal regions with low recombination frequency. Therefore, it is desirable to identify a set of well-distributed “core” markers for whole genome mapping based on physical rather than on genetic position.

The objective of our study was to develop a UCGM for rice that would be useful in our efforts to simultaneously develop several sets of Chromosome Segment Substitution Lines (CSSL), where each line carried a single or few chromosomal fragments of a donor genotype in a common genetic background (Eshed and Zamir 1995; Jeuken et al. 2008). A complete set of CSSLs represents the whole donor genome divided into small overlapping segments, where interactions between donor alleles are restricted to those falling within substituted regions within a given line. This makes CSSL populations useful materials for identification of QTLs with minor effects (Howell et al. 1996). Several CSSL populations have been developed for rice and found to be widely useful (Doi et al. 1997; Ebitani et al. 2005a; Ebitani et al. 2005b; Ghesquière et al. 1997; Jena et al. 1992; Kubo et al. 2002; Li et al. 2005; Lorieux et al. 2000; Sobrizal et al. 1999; Tian et al. 2006; Wan et al. 2004; Yu et al. 2005; Zhang et al. 2006; Tan et al., 2007).

By developing a UCGM for rice, we aimed to improve the efficiency of our efforts to develop interspecific rice CSSL populations within the framework of a collaborative Generation Challenge Programme project (<http://www.generationcp.org/>). By providing a common set of markers that could be used on multiple populations, we were able to reliably trace the introgression of chromosomal fragments in different populations in different laboratories. In this paper, we explain the concept of Universal Core Genetic Map (UCGM), and demonstrate its value in a distributed research program. This tool represents a set of well dispersed, polymorphic markers that can be used for mapping in almost any interspecific combination of *O. sativa* x wild rice, enhancing the ability to compare genetic maps developed on different populations, and facilitating collaboration among research groups. The concept is valid for

intraspecific crosses as well, though additional markers may be necessary to ensure polymorphism in more closely related materials. We also created a computer program, Paddy Map, which facilitates the selection of standard, well distributed and universal marker sets for any kind of segregating population.

Materials and methods

Concept:

The basic concept of Universal Core Genetic Map is the following: starting from the list of SSRs identified in the rice genomic sequence (IRGSP, 2005), we extracted a set of *anchors*. We defined an anchor as a group of closely linked SSRs located in a specific genomic position. The distance that separates the first and last SSRs within an anchor was chosen to be below the genetic resolution of common segregating populations (<500 individuals), based on the mean Kbp/cM ratio calculated from the genetic distances of a saturated genetic map. This allowed us to consider the anchor as a single, or discrete, locus. Anchors were chosen to be evenly distributed along the genome, with spacing between anchors defined by physical distance (in kb). Taking the high recombination rate in telomeric regions into account, a higher density of anchors was required to avoid large gaps between anchors in the subsequently derived genetic maps. Similarly, the low recombination rate in most centromeric regions required a lower density of anchors in those regions.

The main property of an anchor is to be more polymorphic than a single SSR. Let us rewrite the event “At least one SSR belonging to an anchor is polymorphic” as “An anchor is polymorphic”, since the two events are indeed identical. It is plain to see that the probability of an anchor being polymorphic between two given genotypes is higher than that of a single SSR being polymorphic. The difference between the two probability values is proportional to the number of SSRs per anchor; higher the number of SSRs per anchor, greater the difference. This polymorphism property of the anchors is exploited in generating genetic maps that are no longer based on markers, but on anchors. As we will see in the case of interspecific crosses between AA-genome rice species, anchors made of triplets of SSRs make it possible to generate complete genetic maps that share the same set of polymorphic anchors, with very few exceptions. In this way, the maps can be directly compared, without the need of mathematical projection.

Probabilities of polymorphism:

Let P_1 be the observed probability of an SSR being polymorphic, P_2 the probability of an anchor being polymorphic, and n the number of SSRs per anchor. Then:

$$P_2 = 1 - (1 - P_1)^n$$

According to our observations, P_1 was > 0.8 in every interspecific combination we have tested so far. Taking the value of P_1 at its minimum, 0.8, this gives $P_2 = 0.992$ with $n = 3$. In the case of an *O. sativa* x *O. meridionalis* combination, which we are working on in our laboratory, we got $P_1 = 0.976$ and $P_2 \approx 1.0$.

Let P_3 be the probability of all anchors being polymorphic and Na the total number of anchors along the genome. Then:

$$P_3 = P_2^{Na}$$

In the case of *O. sativa* x *O. meridionalis*, and taking $Na = 165$ and $n = 3$, we get $P_3 = 0.997$ with $n = 3$, which is quite a high value. In Table II-1, we show how P_1 , n and Na modify the P_3 value.

Selection of SSR markers:

We used the following criteria to include an SSR marker in the list of candidates for anchor building (i) it has a unique position in the rice genome, (ii) it belongs to the Class 1 microsatellite family (McCouch et al. 2002), and (iii) it shows a high Polymorphism Information Content (PIC) in *indica-japonica* crosses or in interspecific crosses. In the event of several equivalent candidates for an anchor according to the three above-mentioned criteria, we gave priority to SSRs that were previously used for genotyping in the literature and/or in our laboratory.

The candidate markers were selected based on the list of 18,828 SSRs published by the International Rice Genome Sequencing Project (IRGSP 2005). To verify the position of each SSR on the rice genome sequence, we performed a Blast alignment of each primer pair against the rice pseudomolecules (TIGR, v. 5; <http://rice.plantbiology.msu.edu/>). The Blast searches were carried out using a command-line version of the Primer Blaster tool available on the OrygenesDB database Web page (<http://orygenesdb.cirad.fr>). The search for anchors was performed using both the CHARM database (<http://sliver.plbr.cornell.edu/ssr>) and a Visual Basic script that we designed for this purpose.

Table II-1. P_3 probability that all Na anchors are polymorphic, given n markers per anchor and P_I the observed probability that a marker is polymorphic

P_I	100	150	165	200	250
Na					
$n = 3$					
0.5	0.000002	0.000000	0.000000	0.000000	0.000000
0.705	0.074217	0.020219	0.013688	0.005508	0.001501
0.8	0.447886	0.299744	0.265721	0.200602	0.134251
0.85	0.713145	0.602236	0.572457	0.508575	0.429481
0.9	0.904792	0.860643	0.847824	0.818649	0.778703
0.95	0.987577	0.981424	0.979585	0.975308	0.969231
0.99	0.999900	0.999850	0.999835	0.999800	0.999750
$n = 4$					
0.5	0.001574	0.000062	0.000024	0.000002	0.000000
0.705	0.467565	0.319715	0.285259	0.218617	0.149487
0.8	0.852035	0.786477	0.767811	0.725963	0.670105
0.85	0.950623	0.926856	0.919843	0.903684	0.881091
0.9	0.990049	0.985111	0.983635	0.980198	0.975309
0.95	0.999375	0.999063	0.998969	0.998751	0.998439
0.99	0.999999	0.999999	0.999998	0.999998	0.999998
$n = 5$					
0.5	0.041800	0.008546	0.005308	0.001747	0.000357
0.705	0.799584	0.714984	0.691394	0.639334	0.571689
0.8	0.968502	0.953126	0.948562	0.937995	0.923105
0.85	0.992435	0.988674	0.987548	0.984927	0.981194
0.9	0.999000	0.998501	0.998351	0.998002	0.997503
0.95	0.999969	0.999953	0.999948	0.999938	0.999922
0.99	1.000000	1.000000	1.000000	1.000000	1.000000
$n = 10$					
0.5	0.906917	0.863677	0.851112	0.822499	0.783284
0.705	0.999501	0.999252	0.999177	0.999002	0.998753
0.8	0.999990	0.999985	0.999983	0.999980	0.999974
0.85	0.999999	0.999999	0.999999	0.999999	0.999999
0.9	1.000000	1.000000	1.000000	1.000000	1.000000
0.95	1.000000	1.000000	1.000000	1.000000	1.000000
0.99	1.000000	1.000000	1.000000	1.000000	1.000000

The P_I values were chosen to represent different situations. $P_I = 0.5$: *indica* x *japonica* (general case); $P_I = 0.705$: *indica* x *japonica* (our study); $P_I = 0.8$ to 0.99 : *O. sativa* x wild

Genetic materials:

To evaluate the polymorphism levels of the defined anchors, parental lines from the populations developed within the framework of our research projects were selected. These included three accessions of *O. sativa* ssp. *indica*, four of *O. sativa* ssp. *japonica*, three of *O. glaberrima* and six wild rice biotypes of *O. meridionalis*, *O. rufipogon*, *O. barthii*, and *O. glumaepatula* (Table II-2).

The seeds were provided by the CIAT Rice Product Line research program, Embrapa-CNPAP (Goiania, Brazil) and IRD (Montpellier, France). Twenty seeds of each biotype were sown under greenhouse conditions for subsequent DNA extraction.

Total genomic DNA extraction:

Fresh leaf tissue was collected sixteen days after sowing. Total DNA was extracted on 96-well plates using the cetyltrimethylammonium bromide (CTAB) method (Murray and Thompson 1980) with some modifications. The DNA quality was verified by means of visualization of each sample in a 0.8% agarose gel stained with ethidium bromide, and then quantified using a DyNA Quant 200 fluorometer (Hoefer). Finally, DNA was diluted to 4 ng/μl and stored at -20°C.

PCR optimization:

The DNAs of four control rice accessions were used as a template to choose the best amplification conditions for each primer pair: two *O. sativa* (accessions IR64 and Caíapo) and two *O. glaberrima* (accessions TOG5681 and MG12). These accessions correspond to the progenitors of two CSSL interspecific populations that are being developed in our laboratory. The PCR conditions were optimised in varying the annealing temperature of the primers and the magnesium concentration in the PCR. When no or too faint an amplification was obtained, the PCR was repeated using Nipponbare accession as control to discard the possibility of PCR failing in the control accessions.

Evaluation of SSR primer pairs:

PCR was performed in a final volume of 15 μl containing 20 ng of template DNA, 1 X buffer (Tris-HCL 100 mM, 500 mM of KCl, Triton X-100 1%), 2 ng/μL of MgCl₂, 0.3 μM d-NTP, 2 pmol of each primer, and one unit of Taq polymerase. DNA amplification was performed on an MJ Research PTC-220™ thermocycler, programmed for an initial 3 min at 94°C, followed by 30 cycles of 30 seconds at 94°C, 45 seconds from 50 to 61°C according

Table II-2. List of the AA-genome *Oryza spp.* accessions used for evaluation of SSR polymorphism

Species	Accession
<i>O. sativa ssp. indica</i>	93-11
<i>O. sativa ssp. indica</i>	IR36
<i>O. sativa ssp. indica</i>	IR64
<i>O. sativa ssp. japonica</i>	Nipponbare IRGC12731
<i>O. sativa ssp. tropical japonica</i>	Azucena "C"
<i>O. sativa ssp. tropical japonica</i>	Caíapo
<i>O. sativa ssp. tropical japonica</i>	Curinga
<i>O. glaberrima</i>	TOG5681
<i>O. glaberrima</i>	IRGC103544 (MG12)
<i>O. glaberrima</i>	CG14
<i>O. barthii</i>	IRGC105613 (3232)
<i>O. barthii</i>	IRCG101937 (3239)
<i>O. rufipogon</i>	IRGC105491
<i>O. rufipogon</i>	IRGC105890
<i>O. glumaepatula</i>	GEN1233
<i>O. meridionalis</i>	OR44

to the optimal temperature for each SSR primer, and 1 min at 72°C, with a final extension of 5 min at 72°C. The amplified products were separated electrophoretically using PAGE 4% (Sambrook and Russell 2001) silver stained and photographed. Molecular weights were estimated using a 10-pb DNA ladder (GibcoBRL, Invitrogen). For some primer combinations, the polymorphism was also evaluated using a Li-cor sequencer 4200. For this purpose, 0.5 pmol of M13 IRD700 or IRD800 primers was added by reaction.

Polymorphism Information Content:

We determined the Polymorphism Information Content

(PIC) of 489 SSRs belonging to the anchor list, using the 16 AA-genome rice accessions listed in Table II-2. Allelic information was compiled in a database, Paddy Map. The PIC value was calculated with the help of a small program written in Visual Basic for Applications running in MS Excel 2003.

Testing the UCGM:

The UCGM was used to generate genetic maps of three interspecific BC₁F₁ populations derived from the following crosses: (i) *O. sativa* ssp. *indica* cv. IR64 × *O. glaberrima* acc. TOG5681, (ii) *O. sativa* ssp. *japonica* cv. Curinga × *O. rufipogon* acc. IRGC105491, and (iii) *O. sativa* BRSMG Curinga × *O. meridionalis* acc. OR44 (or W2112). The genetic maps were compared both for marker order and map distances with the expected map based on the genomic sequence. To get equivalent distance units between the physical and genetic maps, we used a conversion rate of 200 Kbp/cM, which is a mean value obtained from the comparison of several rice genetic maps with the chromosome physical sizes.

Results

Finding anchor sets:

A set of 165 well-distributed anchors was identified, each containing at least three, tightly linked SSR markers and representing a total of 517 SSR loci. The average distance covered by each triplet of SSR is 178,501 bp and the mean genomic distance between two anchors was approximately 2.5 megabases (Mb), except in the telomeric regions, where it was approximately 1.5 Mb (Figure II-1). Using this strategy, we were able to cover the whole rice genome with no significant gap between adjacent anchors.

A



Figure II-1. A. Distribution of SSR markers that form the Universal Core Genetic Map, arranged in anchors throughout the rice genome (chromosomes 1 to 6). Graphical map obtained with Mapdisto v 1.7. **B** Distribution of SSR markers that form the Universal Core Genetic Map, arranged in anchors throughout the rice genome (chromosomes 7 to 12)

The precise genomic positions of the 512 SSRs and their 165 anchors, together with the primer sequences, are available as electronic resources at (<http://mapdisto.free.fr/PaddyMap/UCGM/>).

Anchor polymorphism in selected germplasm:

Using 16 diverse accessions (described in Materials and Methods), a total of 3,207 alleles were obtained based on evaluation of 489 of the 512 SSRs markers. The number of alleles per locus varied widely, ranging from 2 (RM6233, RM1063, RM12923, RM15065, RM14327, RM16382, RM20155, RM6652, RM172, RM5999, RM23769, RM26001, RM6327) to 16 (RM8131), with an average of 6.55 alleles/locus. As expected, these values are higher than the values reported in other studies based on isoenzymes (Glaszmann 1987; Second 1982) and restriction fragment length polymorphisms (RFLP) (Zhang et al. 1997). They were similar to the values found in other studies based on SSR markers (Ni et al. 2001, Garriss et al, 2005; Blair et al., 2002). Inclusion of wild species in this study induced the presence of a high number of rare alleles.

Most of the markers produced successful amplification in all tested accessions, with 98.5 % and 98.2 % of positive reactions for *japonica* and *indica* accessions, respectively, and 78.9 % for *O. meridionalis* (accession OR44) and 95.7 % for *O. glaberrima* (accession TOG 5681) (Table II-3). These values are similar to those found by (Song et al. 2003), who evaluated the genetic diversity in *O. rufipogon* populations with 23 SSR markers, and where all of the markers did amplify. However, they are higher than those found in a study of the population structure of *O. glumaepatula* (Brondani et al. 2005).

The observed rate of polymorphic SSRs (RPS) was 85.4 % and 70.4 % for inter- and intraspecific crosses, respectively. The RPS ranged from 79.8 % to 90.5 % for *O. sativa* × wild rice crosses, with a mean of 85.8 %, while RPS was slightly lower in *O. sativa* × *O. glaberrima* crosses, ranging from 83.2 % to 88.8 %. The average RPS between the parents of the six CSSL populations under development was 86.2 %, including two *O. sativa* × *O. glaberrima*, one *O. sativa* × *O. rufipogon*, one *O. sativa* × *O. barthii*, one *O. sativa* × *O. glumaepatula*, and one *O. sativa* × *O. meridionalis* crosses.

The PIC ranged from 0.110 (RM6233, RM15065, RM16382, RM5999) to 0.908 (RM6738, RM8131), with an average of 0.691. This is evidence of high mutation frequencies in the tested accessions, confirming the usefulness of the tested SSRs for diversity analysis and mapping purposes (Olufowote et al. 1997; Garriss et al., 2005). However, the relation between the number of alleles and the PIC is not a direct one, an observation that was already reported in other studies (Huang et al. 2003; Matus and Hayes 2002).

B

Chr 7	Chr 8	Chr 9	Chr 10	Chr 11	Chr 12
RM295 (80) RM6652 (80) RM3394 (80) RM20882 (80.1) RM20891 (80.1) RM20897 (80.1) RM5752 (81) RM5055 (81) RM427 (81) RM20973 (81) RM8909 (82) RM3222 (82) RM3484 (82) RM21238 (83) RM21242 (83) RM6776 (83) RM21323 (84) RM21328 (84) RM3859 (84) RM3456 (85) RM542 (85) RM214 (85) RM5543 (86) RM500 (86) RM5875 (86) RM11 (87) RM3743 (87) RM5420 (87) RM6403 (88) RM455 (88) RM2752 (88) RM234 (89) RM21976 (89) RM18 (89) RM118 (90) RM1364 (90) RM429 (90) RM22105 (90.1) RM22106 (90.1) RM6389 (90.1) RM248 (91) RM420 (91) RM172 (91)	RM408 (92) RM7034 (92) RM337 (92) RM3702 (92.1) RM1789 (92.1) RM3309 (92.1) RM1376 (93) RM5428 (93) RM22418 (93) RM310 (94) RM22529 (94) RM547 (94) RM8243 (95) RM85694 (95) RM22696 (95) RM3481 (95) RM22837 (96) RM6471 (96) RM22839 (96) RM22883 (97) RM5999 (97) RM6990 (97) RM6193 (98) RM27997 (98) RM23001 (98) RM23174 (99) RM23175 (99) RM284 (99) RM3262 (100) RM210 (100) RM23251 (100) RM149 (101) RM3845 (101) RM3452 (101) RM502 (102) RM447 (102) RM23502 (102) RM6948 (102.1) RM7631 (102.1) RM3480 (102.1) RM264 (103) RM477 (103) RM4153 (103)	RM23654 (104) RM23661 (104) RM23662 (104) RM23736 (104.1) RM23742 (104.1) RM23743 (104.1) RM23769 (105) RM5799 (105) RM23779 (105) RM23914 (106) RM23916 (106) RM5526 (106) RM296 (107) RM24085 (107) RM3912 (107) RM409 (108) RM6854 (108) RM24271 (108) RM6839 (108) RM410 (109) RM257 (109) RM6343 (109) RM3249 (110) RM6491 (110) RM160 (110) RM215 (110.1) RM21718 (110.1) RM1026 (111) RM1039 (111) RM205 (111)	RM5095 (112) RM7492 (112) RM6364 (112) RM474 (112.1) RM2381 (112.1) RM24952 (112.1) RM6404 (113) RM6179 (113) RM3882 (113) RM216 (114) RM6646 (114) RM25104 (114) RM25149 (115) RM2504 (115) RM25152 (115) RM239 (116) RM1126 (116) RM311 (116) RM25365 (117) RM25366 (117) RM5689 (117) RM184 (118) RM2432 (118) RM1375 (118) RM25663 (119) RM25664 (119) RM171 (119) RM3510 (119) RM147 (120) RM25788 (120) RM25784 (120) RM3563 (120.1) RM3834 (120.1) RM228 (120.1) RM6673 (121) RM4771 (121) RM590 (121)	RM6327 (122) RM2864 (122) RM7173 (122) RM26001 (122) RM26062 (122.1) RM26063 (122.1) RM2459 (122.1) RM26105 (123) RM337 (123) RM26108 (123) RM26249 (124) RM116 (124) RM6894 (124) RM536 (125) RM26406 (125) RM26513 (126) RM5590 (126) RM7120 (126) RM26643 (127) RM26646 (127) RM26652 (127) RM26643 (127) RM26646 (127) RM26652 (127) RM209 (128) RM26796 (128) RM26797 (128) RM206 (129) RM26998 (129) RM26999 (129) RM27154 (130) RM27162 (130) RM27318 (130.1) RM27326 (130.1) RM3577 (130.1) RM2136 (131) RM7443 (131) RM5926 (131)	RM5568 (132) RM3880 (132) RM3323 (132) RM7315 (132.1) RM5842 (132.1) RM3747 (132.1) RM27542 (133) RM27564 (133) RM217 (133) RM6296 (133) RM27706 (134) RM3146 (134) RM512 (134) RM1036 (135) RM101 (135) RM27877 (135) RM27962 (136) RM27970 (136) RM27973 (136) RM28118 (137) RM465 (137) RM6022 (138) RM5197 (138) RM3448 (138) RM28404 (139) RM463 (139) RM7018 (139) RM28607 (140) RM6265 (140) RM6396 (140) RM1159 (140.1) RM6411 (140.1) RM1300 (140.1) RM6306 (141) RM2313 (141) RM1227 (141)

Figure II 1 Continued

Database:

We developed a simple database, Paddy Map, to provide the scientific community with a user-friendly tool for access to the information described in this paper. Polymorphic SSR markers for any cross between the species analysed in this study can be displayed, as well as their attached anchor and position in the genome. Several other features, such as searching for markers within a specific anchor, or computing various types of statistics, are implemented. Paddy Map is available for download at (<http://mapdisto.free.fr/PaddyMap/>).

Application to the genetic mapping of three interspecific crosses:

To validate the utility of the UCGM, we used it to derive genetic maps of three BC₁F₁ populations representing the first generation development of CSSL libraries. We built these maps to validate the SSR positions, keeping in mind that some of them may be duplicated in certain *Oryza* wild species and could potentially tag unexpected – and unknown – chromosomal fragments. This would result in misleading interpretations of chromosomal position of introgressions and would ultimately lead to confusion when analyzing the phenotypic performance of a number of CSSLs. The three crosses we used were *O. sativa* ssp. *indica* cv. IR64 × *O. glaberrima* acc. TOG5681, *O. sativa* ssp. *tropical japonica* cv. Curinga × *O. rufipogon* acc. IRGC105491, and *O. sativa* BRSMG Curinga × *O. meridionalis* acc. OR44 (or W2112). Backcrosses were made to the *O. sativa* parent in the three populations. Population sizes ranged from 88 to 125 individuals. We used the Paddy Map database to extract one polymorphic SSR per anchor for each cross. The populations were then genotyped for the selected markers, and the genetic maps were built using the MapDisto v. 1.7 software (Lorieux 2007) (<http://mapdisto.free.fr/>). The obtained genetic maps were compared with the expected map using the “Compare Maps” tool implemented in MapDisto. All maps showed a strong colinearity with the expected map, except in the centromeric regions, where a few inversions were observed due to the poor precision of the estimation of the map distances in these low recombining regions. The total map sizes were well correlated with the predicted sizes (Table II-4).

Table II-3. Percentages of successful amplification in the cultivated and wild species, calculated on the 489 SSRs used in this study

Accession groups	Percentage of positive PCR (%)
Cultivated species	
<i>O. sativa</i> ssp. <i>indica</i>	98.296
<i>O. sativa</i> ssp. <i>japonica</i>	98.517
<i>O. glaberrima</i>	95.774
<i>O. sativa</i> + <i>O. glaberrima</i>	97.529
Wild species	
<i>O. barthii</i>	95.706
<i>O. rufipogon</i>	94.683
<i>O. glumaepatula</i>	87.730
<i>O. meridionalis</i>	78.937
All wild	89.264

Table II-4. Comparisons of physical and genetic maps sizes in three BC1F1 interspecific populations

	<i>O. sativa</i> ssp. <i>indica</i> acc. IR64 x <i>O. glaberrima</i> acc. TOG5681	<i>O. sativa</i> ssp. <i>japonica</i> acc. Curinga x <i>O. rufipogon</i> acc. IRGC105491	<i>O. sativa</i> ssp. <i>japonica</i> acc. Curinga x <i>O. meridionalis</i> acc. OR44
Genetic map	1,838	1,884	1,886
Physical map	1,932	1,839	1,843

Genetic distances were computed using Mapdisto v 1.7. Physical distances were converted to centimorgans (cM) using the conversion rate of 200 Kbp/cM

Discussion

On the concept of UCGM:

The strategy used to develop the UCGM and the particular marker sets utilized in this study make it possible for different laboratories to readily generate genetic maps on different populations using a common set of markers. This avoids many of the problems encountered by collaborative projects where different sets of markers are used and comparisons among subsequently generated maps can be difficult. The concept behind the UCGM is that by arranging a series of SSR markers into *anchors* that are represented by triplets of closely linked SSRs, the probability of at least one SSR being polymorphic within a given anchor is much higher than that of a single SSR being polymorphic. By comparing anchor positions instead of marker positions, the UCGM enables a direct comparison of any objects (genes, QTLs, introgressed fragments, etc.) that are positioned on any map generated from the set of anchors.

Another common issue in the construction of genetic maps is the choice of sets of “core” markers that are evenly distributed along a reference genetic map. Due to the uneven distribution of recombination events along the chromosomes, the use of genetic map as the reference results in poor distribution of the markers along the physical genome, leading to a lower number of markers per gene in genomic regions with a low recombination rate. To address this issue, we chose the positions of the UCGM anchors on the basis of the physical map of rice.

Case of interspecific crosses:

The polymorphism observed in 16 accessions representing 6 AA-genome *Oryza* species showed high to very high levels of allelic variation between the cultivated Asian rice, *O. sativa*, and its relatives *O. glaberrima*, *O. meridionalis*, *O. rufipogon*, *O. barthii*, and *O. glumaepatula*. This allowed us to successfully apply the concept of a universal map, which is based on the polymorphism probability of a set of anchors (Table II-1), to the case of three interspecific crosses. The three genetic maps that we generated were validated for both the marker orders/positions and the total map sizes. We therefore continued the development of CSSL libraries with the same anchor set applied to the BC2 and BC3 generations. This way, it will be possible to locate all QTLs or major genes identified using these libraries on the virtually same genetic map.

Case of intrasubspecific crosses:

We see from Table II-1 that, for a value of 0.5 of the observed probability of an SSR being polymorphic, P_1 , which is close to the commonly observed value in the case of intrasubspecific *indica* × *japonica* crosses, it is necessary to increase the number of SSRs per anchor to about $n = 10$, to obtain a fair probability P_3 of all anchors being polymorphic. This objective is feasible, due to the availability of increasing amounts of sequence information on diverse genomes. Moreover, the present study revealed a significant polymorphism level in *indica* × *japonica* crosses (70.4 %). We see from Table II-1 that a new set of anchors made of five SSR loci—an easily reachable goal—should produce a very useable framework, at least for the accessions we tested.

Conclusion

Usefulness of the UCGM:

AA-genome *Oryza* species represent an invaluable genetic reservoir. The development of introgression line libraries is a powerful means of taking advantage of this genetic diversity. The UCGM represents an efficient way to build directly comparable genetic maps and graphical genotyping representations, which will eventually allow us to project all genes and QTLs detected using these populations in the same framework of physically located anchors. We are currently applying this strategy to the development and analysis of six CSSL libraries. The UCGM is also valuable for marker-assisted backcrossing (MAB) in a breeding program. In a separate study, we are using 108 of the 165 anchors to detect and introgress three QTLs for tolerance to the Rice Hoja Blanca Virus (RHBV) and its insect vector, *Tagosodes orizicolus*, in two intrasubspecific populations (*O. sativa* spp. *indica* acc. Fedearroz 50 and acc. Fedearroz 2000 × *O. sativa* spp. *indica* acc. WC366).

Widespread use of the UCGM for genotyping enables us to integrate all of the gene discovery outcomes in a centralized database (Laboratory Information Management System), facilitating the breeders' work in choosing appropriate donors in their selection schemes. This integration will also help us in drastically reducing the time spent in polymorphism screening and it will allow us to make our gene discovery and breeding programs more cost-efficient, as the same set of primers can then be shared by the various rice genetics and breeding projects.

Applicability to other crops:

Now that sequencing a whole genome has become a relatively economical and easy task, developing UCGM for other crops can be envisioned.

Acknowledgements We are grateful to the Generation Challenge Programme for their financial support through the project No G3005.10 “Exploring natural genetic variation: Developing genomic resources and introgression lines for four AA genome rice relatives”.

References

- Blair MW, Hedetale V, McCouch SR (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). *Theor Appl Genet* 105:449-457
- Brondani RPV, Zucchi MI, Brondani C, Rangel PHN, Borba T, Rangel PN, Magalhaes MR, Vencovsky R (2005) Genetic structure of wild rice *Oryza glumaepatula* populations in three Brazilian biomes using microsatellite markers. *Genetica* 125:115-123
- Doi K, Iwata N, Yoshimura A (1997) The construction of chromosome substitution lines of African rice (*Oryza glaberrima* Steud.) in the background of japonica rice (*Oryza sativa* L.). *Rice Genet News* 14:39-41
- Ebitani T., Takeuchi Y., Nonoue Y., Yamamoto T., Takeuchi K., M. Y (2005) Construction and evaluation of chromosome segment substitution lines carrying overlapping chromosome segments of indica rice cultivar “Kasalath” in a genetic background of japonica elite cultivar “Koshihikari”. *Breeding Science* 48:395-399
- Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141:1147-1162
- Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics* 169:1631-1638
- Ghesquière A, Séquier J, Second G, Lorieux M (1997) First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a "contig line" concept. *Euphytica* 96:31-39
- Glaszmann JC (1987) Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics* 74:21-30
- Howell PM, Marshall DF, Lydiate DJ (1996) Towards developing intervarietal substitution lines in brassica napus using marker-assisted selection. *Genome* 39:348-358
- Huang XQ, Wang LX, Xu MX, Roder MS (2003) Microsatellite mapping of the powdery mildew resistance gene Pm5e in common wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 106:858-865
- IRGSP (2005) The map-based sequence of the rice genome. *Nature* 436:793-800
- Jena K, Kochert G, Khush G (1992) RFLP analysis of rice (*Oryza sativa* L.) introgression lines. *Theor Appl Genet* 84:608-616
- Jeuken MJ, Pelgrom K, Stam P, Lindhout P (2008) Efficient QTL detection for nonhost resistance in wild lettuce: backcross inbred lines versus F₂ population. *Theor Appl Genet* 116:845-857

- Kubo T, Aida Y, Nakamura K, Tsunematsu H, Doi K, Yoshimura A (2002) Reciprocal chromosome segment substitution series derived from japonica and indica cross of rice (*Oryza sativa* L.). *Breeding Science* 52:319-325
- Li ZK, Fu BY, Gao YM, Xu JL, Ali J, Lafitte HR, Jiang YZ, Rey JD, Vijayakumar CH, Maghirang R, Zheng TQ, Zhu LH (2005) Genome-wide introgression lines and their use in genetic and molecular dissection of complex phenotypes in rice (*Oryza sativa* L.). *Plant Mol Biol* 59:33-52
- Lorieux M (2007) MapDisto: A free user-friendly program for computing genetic maps. Computer demonstration given at the Plant and Animal Genome XV conference, Jan 13-17 2007, San Diego, CA URL: <http://mapdisto.free.fr/>
- Lorieux M, Ndjonjop M-N, Ghesquière A (2000) A first interspecific *Oryza sativa*×*Oryza glaberrima* microsatellite-based genetic linkage map. *Theoretical and Applied Genetics* 100:593-601
- Matus IA, Hayes PM (2002) Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. *Genome* 45:1095-1106
- McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res* 9:257-279
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321-4325
- Ni J, Colowit PM, Oster JJ, Xu K, Mackill DJ (2001) Molecular markers linked to stem rot resistance in rice. *Theoretical and Applied Genetics* 102:511-516
- Olufowote JO, Xu YB, Chen XL, Park WD, Beachell HM, Dilday RH, Goto M, McCouch SR (1997) Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome* 40:370-378
- Sambrook J, Russell D (2001) *Molecular cloning: a laboratory manual* Third Edition edn. Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, NY 11724.
- Second G (1982) Origin of the genic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. *The Japanese journal of genetics* 57:25-57
- Sobrizal K, Sanchez P, Doi K, Angeles E, Khush GS, Yoshimura A (1999) Development of *Oryza glumaepatula* introgression lines in rice, *O. sativa* L. *Rice Genet News* 16:107-108
- Song ZP, Xu X, Wang B, Chen JK, Lu BR (2003) Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. *Theoretical and Applied Genetics* 107:1492-1499
- Tan L, Liu F, Xue W, Wang G, Ye S, Zhu Z, Fu Y, Wang X and Sun C (2007) Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield-related quantitative trait loci. *J. of Integrative Plant Biology* 49(6):871-884
- Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and Genetic Marker Potential. *Genome Res* 11:1441-1452
- Tian F, Li DJ, Fu Q, Zhu ZF, Fu YC, Wang XK, Sun CQ (2006) Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theor Appl Genet* 112:570-580
- Wan XY, Wan JM, Su CC, Wang CM, Shen WB, Li JM, Wang HL, Jiang L, Liu SJ, Chen LM, Yasui H, Yoshimura A (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. *Theor Appl Genet* 110:71-79
- Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, Mukai Y, Koike K, Yazaki J, Fujii F, Shomura A, Ando T, Kono I, Waki K, Yamamoto K, Yano

- M, Matsumoto T, Sasaki T (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell* 14:525-535
- Yu C, Wan J, Zhai H, Wang C, Jaing L, Xiao Y, Liu Y (2005) Study on heterosis of intersubspecies between indica and japonica rice (*Oryza sativa* L.) using chromosome segment substitution lines. *Chinese Science Bulletin* 50:131-136
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296:79-92
- Zhang Q, Liu KD, Yang GP, Maroof MAS, Xu CG, Zhou ZQ (1997) Molecular marker diversity and hybrid sterility in indica-japonica rice crosses. *Theoretical and Applied Genetics* 95:112-118
- Zhang X, Zhou S, Fu Y, Su Z, Wang X, Sun C (2006) Identification of a drought tolerant introgression line derived from Dongxiang common wild rice (*O. rufipogon* Griff.). *Plant Mol Biol* 62:247-259
- Zhang ZH, Deng YJ, Tan J, Hu SN, Yu J, Xue QZ (2007) A genome-wide microsatellite polymorphism database for the indica and japonica rice. *DNA Research* 14:37-45

III

CARTOGRAPHIE DU LOCUS S_I ET MODELE GENETIQUE DE SELECTION GAMETIQUE

Introduction

La barrière de reproduction qui sépare *O. glaberrima* et *O. sativa* résulte de l'action conjointe de plusieurs gènes, qui a pour effet de rendre les hybrides F_1 totalement mâle stériles et partiellement femelle stériles. Certains de ces gènes ont été cartographiés par différentes équipes de recherche, et montrent une distribution sur tout le génome (voir Tableau I-9). Parmi ces locus, S_I est celui qui a l'effet le plus marqué.

Au début de ces travaux de thèse, la principale source d'information sur le locus S_I était constituée par les publications de Y. Sano (Sano *et al.*, 1979; Sano, 1990) dans lesquelles les différents effets du locus sur la stérilité mâle et femelle de l'hybride *O. sativa* x *O. glaberrima* étaient mis en évidence. Selon Sano, cette stérilité était due à une accumulation d'un locus (S_I) à effet « pollen killer » (purement mâle) avec des modificateurs qui lui sont liés, agissant ensemble comme un « gamete eliminator » (effet sur la production de gamètes mâles et femelles). Sano localisait S_I dans un intervalle assez large sur le bras court du chromosome 6. Des travaux ultérieurs du groupe de génétique et génomique du riz de l'IRD (Lorieux *et al.*, 2000) et de l'ADRAO (Heuer et Miezán, 2003) ont permis de préciser cette cartographie grâce à l'utilisation de marqueurs microsatellites.

Nos travaux sur la cartographie fine du locus S_I ont démarré sur la base de ces résultats. Entre temps, le groupe de Y. Sano a publié de nouveaux résultats (Koide *et al.*, 2008c) montrant que le locus S_I possède au moins deux composants, un premier affectant la fertilité mâle et un autre affectant la fertilité femelle, avec des modificateurs qui augmentent la distorsion de la ségrégation dans les gamètes femelles. Ils cartographient le facteur de stérilité mâle dans un intervalle de 45 kpb sur le génome d'*O. sativa* cv. Nipponbare. Cette nouvelle information nous a amené à nous focaliser sur le facteur de stérilité femelle de S_I , afin d'en préciser sa localisation génomique et de caractériser son mode d'action.

Le manuscrit présenté dans ce chapitre, qui a été soumis à la revue Genetics, se focalise donc sur la cartographie fine du facteur du locus S_I qui affecte la fertilité femelle. Cet article présente tout d'abord la construction d'une carte génétique interspécifique, à l'aide de l'outil UCGM (carte universelle) préalablement développé et décrit dans le chapitre précédent. Cette carte devait nous permettre de rechercher des associations entre la distorsion de la ségrégation de marqueurs moléculaires avec la présence de locus de stérilité entre *O. sativa* et *O. glaberrima*. Nous avons ensuite développé une méthodologie de cartographie de gènes de stérilité, basée sur l'effet de la sélection gamétique sur la ségrégation des marqueurs qui leur sont génétiquement liés. L'application de cette approche sur quatre populations interspécifiques (BC_1F_1) nous a permis de cartographier avec précision le facteur du locus S_I

qui affecte la fertilité femelle. Par ailleurs, la méthodologie développée nous a aussi permis de mettre en évidence deux autres facteurs qui interagissent probablement en épistasie avec S_I dans le contrôle du phénotype de stérilité femelle des hybrides.

Dans le but de trouver des gènes ou facteurs candidats pour le locus S_I , une étude de la séquence génomique a été entreprise. La construction d'une carte physique de la région chez *O. glaberrima*, le séquençage et l'annotation du clone BAC qui contient S_I sont présentés dans la publication. D'autre part, nous proposons un modèle génétique explicatif de la stérilité femelle, construit sur la base de nos données et des informations disponibles dans la bibliographie.

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^{*}, Mathias Lorieux^{*,†}

^{*}Plant Genome and Development Laboratory (UMR 5096 – CNRS – IRD – UPVD), 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, CEA, 91057 Evry, France

[§]Plant Genome and Development Laboratory (UMR 5096 – CNRS – IRD – UPVD), Université de Perpignan, 66860 Perpignan, France.

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

Keywords:

Oryza sativa; *Oryza glaberrima*; Hybrid sterility; Transmission Ratio Distortion; Bateson-Dobzhansky-Muller incompatibilities

Abstract

S_I 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 separately male and female fertility. 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. In order to fine map the factor of the S_I locus affecting female fertility, we developed a new mapping approach based on the evaluation of the degree of female Transmission Ratio Distortion (fTRD) of markers. By the implementation of this methodology in four *O. sativa* x *O. glaberrima* crosses, the female component of the S_I 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 that interact with S_I was also found. In light of the data available, a model where incompatibilities in epistatic interactions between S_I 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_I . According to our model, the recombination ratio and allelic combinations between these factors would determine the final allelic frequencies observed for a given cross.

Introduction

Rice (*Oryza sativa* L.) is the most important crop for human consumption, feeding more than half of the world's population (ZEIGLER and BARCLAY 2008). With the continuous growing food demand and climate changes that menace food production, rice culture is challenged to increase yield in a significant and sustainable manner. As a response to that need, breeding for varieties with high yield, and tolerance to biotic and abiotic stresses, becomes a priority for the national breeding programs and rice research.

A promising way to improve rice consists to identify and introgress interesting traits through interspecific crosses with closely related species. However, despite that the rich genetic diversity found within the *O. sativa* complex (AA), and the high level of heterosis found in the F₁ plants, are both very promising for rice improvement; the frequent hybrid sterility (CHU *et al.* 1969) hinders their common utilization in the different breeding programs. From this genomic group, the African cultivated rice species *O. glaberrima* Steud, represents an interesting source of drought tolerance (SARLA and SWAMY 2005), weed competitiveness (DINGKUHN *et al.* 1998), nematode, and virus resistances (NDJIONDJOP *et al.* 1999; SORIANO *et al.* 1999). Several of these traits have been 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 the action of several loci that might interact or act separately to render the F₁ hybrids pollen sterile and partially female fertile. Among these, the *S₁* 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* x *O. glaberrima* crosses has been possible by performing successive backcrosses followed by selfing (GHESQUIÈRE *et al.* 1997). Nevertheless, this fertility recovery is associated to the presence of the homozygote *O. glaberrima S₁* allele, and is lost again when re-crossing with *O. sativa* (HEUER and MIEZAN 2003). Furthermore, a strong Transmission Ratio Distortion (TRD) of markers linked to *S₁*, in favor of the *O. glaberrima* alleles, settles as a consequence of the systematic elimination of the *O. sativa* alleles from the descendants (ALUKO *et al.* 2004; GHESQUIÈRE *et al.* 1997; LORIEUX *et al.* 2000).

As an alternative to open all the 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* x *O. glaberrima* hybrids that would be fertile when crossed with *O. sativa*. In order to develop these interspecific bridges, genetic factors that affect their fertility must be identified and characterized to better understand the nature of the sterility barrier. Recently, the complex nature of the *S₁* locus was suggested, and the male component

mapped to an interval equivalent to 45 kb on the genome of *O. sativa* cv. Nipponbare (KOIDE *et al.* 2008c). However, the nature of the components affecting female fertility is still unknown.

In this work we describe the fine genetic and physical mapping of the female component of the *S₁* locus, by a new approach based on the evaluation of the degree of TRD caused by the elimination of the *O. sativa* alleles of highly linked markers. This methodology enabled us to identify a 27.8 kb region on the short arm of chromosome 6 as the most probable location of the female component of the *S₁* locus. Additionally, the presence of other factors involved in the sterility was predicted, and their approximated localization deduced. Direct comparison of the *S₁* region in *O. glaberrima* with the *O. sativa* orthologous sequence revealed the presence of several significant differences, and of two possible candidate factors. Based on our data, a model that involves the epistatic interaction of the additional factors with *S₁* was developed to explain the female sterility and the TRD found in the interspecific hybrids.

Materials and Methods

Plant materials:

BC₁F₁ 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). 10-days seedlings were transplanted temporally in the greenhouse and afterwards brought to the field at the International Center for Tropical Agriculture (CIAT, Cali, Colombia) headquarters.

Molecular markers analysis:

In order to derive a new *O. sativa* x *O. glaberrima* genetic map, 140 markers belonging to the version 1 of the Universal Core Genetic Map (UCGM) for rice (ORJUELA *et al.* 2009) (<http://mapdisto.free.fr/PaddyMap/>) were evaluated in the first 125 BC₁F₁ individuals of the

IR64xTOG5681 population. Marker saturation was carried out in the S_I region on chromosome 6 (LORIEUX *et al.* 2000; SANO 1990) (Supplemental Table III-1). The rest of the chromosome was also saturated at a lower density. SSR markers were chosen based on the physical location on the *O. sativa* cv. Nipponbare pseudomolecule (TIGR v. 6 data, (OUYANG *et al.* 2007). Primers were synthesized according to the IRGSP (IRGSP 2005). Further saturation near S_I was done using new additional markers (Supplemental Table III-2) found with the MISA script (THIEL *et al.* 2003), and InDels markers designed by comparison of 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 with an annealing temperature and magnesium concentration ranging from 1.5 to 2.5 mM optimized for each marker (Supplemental Table III-1 and 2), as previously described (ORJUELA *et al.* 2009). Electrophoresis of PCR products was made in 4% agarose and revealed with Ethidium Bromide for polymorphism greater than 12 bp, and in a Li-Cor sequencer (Li-Cor Biosciences) for smaller polymorphism, using a M13 tag (IRD700 and IRD800).

Statistical analysis and mapping approach:

Genetic map construction was made using the MapDisto 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 a segregation χ^2 test. Genotyping error candidates were detected using the tools provided in the Color Genotypes module, using a threshold of 0.005.

In order to find the most probable location of the female component of the S_I locus, the estimation of the female TRD (fTRD) was used as an indicator of the closeness of a given marker from the factor causing the distortion. 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 of the S_I locus in our BC₁F₁ populations depends only from female gametogenesis, the k_f and the k statistic are equivalent, and were thus calculated as the number of heterozygotes over the total number of individuals.

Physical mapping of the *S_I* locus:

For the *O. glaberrima* cv. CG14, several BAC clones of the *S_I* region were chosen based on 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_I* 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 done by the Sanger method. The final BAC sequence was first analyzed using BLAST against different public and local plant nucleotides and proteins databases. Coding regions were *ab initio* predicted using the FGENESH program (SALAMOV and SOLOVYEV 2000) and then confirmed by comparative analysis with annotated genes 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 the comparison with the Nipponbare genome was accomplished using dot-plot alignments of 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 (Ka/Ks) were calculated using the PAML package (YANG 1997), implemented in the PAL2NAL program (SUYAMA *et al.* 2006).

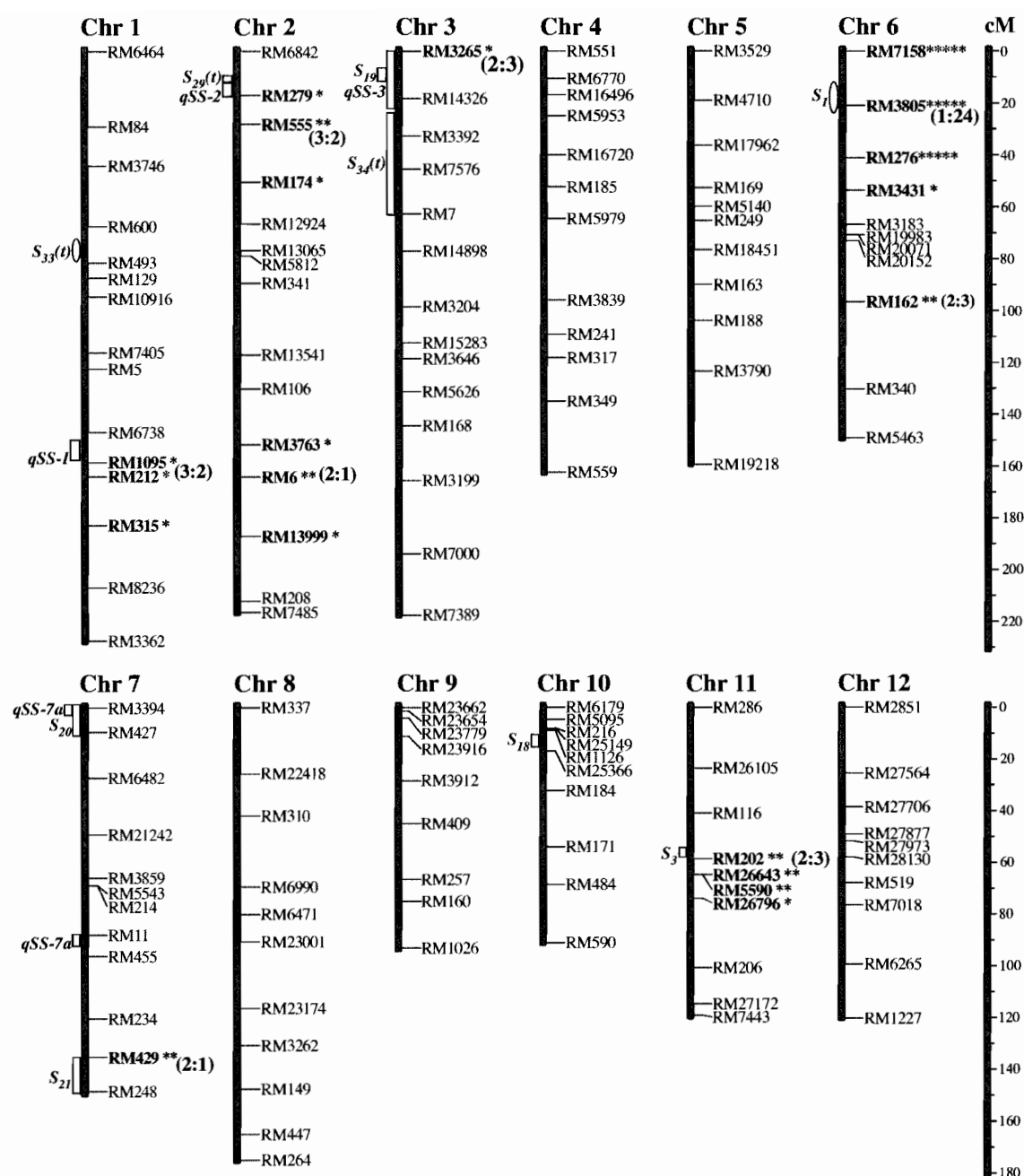


Figure III-1.—Genetic map of the *O. sativa* cv. IR64 x *O. glaberrima* cv. TOG5681 cross, fTRD genomic sites, and previously described sterility loci. Genetic map constructed with 125 individuals of the IR64 x TOG5681 backcross, using 140 anchors from the UCGM. Markers presenting a significant TRD with $P < 0,05$, $P < 0,01$ and $P < 10^{-5}$ are indicated by (*), (**) and (****) respectively 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 pseudomolecule (S_1 (SANO 1990), S_3 (SANO 1983), S_{18} (DOI *et al.* 1998), 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 $qSS7b$ (LI *et al.* 2008a)). Continuous open bars are used to represent loci involved in pollen sterility, while ellipses represent loci affecting male and female fertility.

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 in order to find markers with fTRD and to compare their colocalization with previously described sterility loci between the two species. Using markers from the UCGM a 1879.56 cM map length was obtained (Figure III-1). Possible inversions were detected in centromeric regions (chromosomes 1, 5, 7, 8, 9, 11), and in the short arm of chromosome 10 (Supplemental Figure III-1). Several sites showing a significant deviation from the 1:1 expected segregation ($P < 0.05$) were also found (Figure III-1). The *O. glaberrima* allele was overrepresented on chromosomes 3, 6 and 11, while the *O. sativa* alleles were more common than expected on chromosomes 1, 2 and 7.

Based on TIGR v. 6 information, 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) did colocalize with the different loci described so far (Figure I-1). The strongest fTRD ($P < 10^{-5}$) was found at the S_I 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_I locus by measuring the TRD:

Recently, Koide and colleagues (2008b) proved that the TRD of the Wx allele in pollen produced by heterozygote plants for the S_I region was due to the abortion of gametes carrying the highly linked S_I *O. sativa* allele (denoted in this paper as S_I^s , equivalent to S_I^a (SANO 1990)). They also inferred that the S_I locus has at least two components, one controlling male TRD (mTRD) and the other fTRD. Additionally, the male component of the locus was mapped to a 45 kb region in the Nipponbare pseudomolecule.

In order to map the female component of the S_I locus, we developed a new approach based on the direct measurement of the degree of deviation from the expected Mendelian ratio. As the *O. glaberrima* allele (S_I^g) causes the gametal elimination of S_I^s (and thus of markers linked to it); the degree of TRD of a given marker would be inversely proportional to the genetic distance between the distorter factor and the marker due to meiotic recombination. In consequence, the closest markers to S_I would show the highest TRD. To examine the fTRD level around S_I , a more saturated genetic map was obtained with the 125 IR64xTOG5681 BC₁F₁ individuals. Two succeeding marker saturations in the S_I region

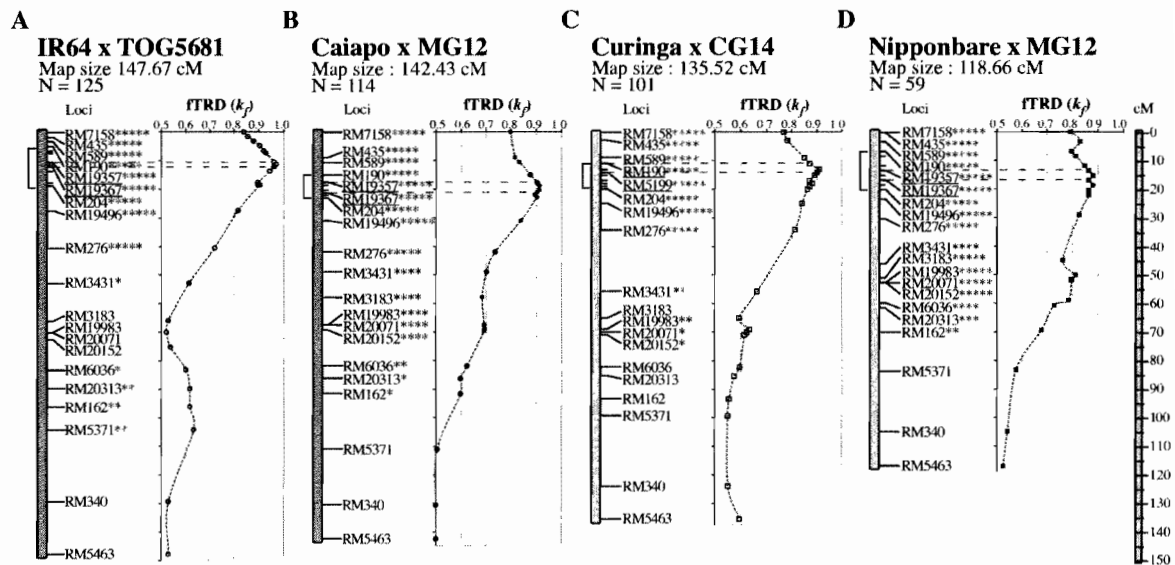


Figure III-2—Mapping of the female component of the S_1 locus using the fTRD measure. Chromosome 6 genetic map for (A) IR64xTOG5681, (B) CAxMG12, (C) CUxCG14 and (D) NIPxMG12 populations. The fTRD value (k) is indicated for each marker. Significant fTRD is shown ($P < 0,05$ (*), $P < 0,01$ (**), $P < 10^{-3}$ (***), $P < 10^{-4}$ (****) and $P < 10^{-5}$ (*****)). Underlined markers limit the fTRD peak for each cross. For each cross, only relevant markers are shown, see **Figure III-3** for the complete list of markers.

(approximately 1 SSR/100 kb and 1 SSR/40 kb) were carried out between the markers showing the highest fTRD in previous works (RM190 and RM204 (LORIEUX *et al.* 2000)). A peak in the fTRD of $k_f = 0.970$ was found, limited by markers RM19357 and RM19367 (Figure 2A). No apparent genomic rearrangements were found in the genetic map compared with the Nipponbare physical map. Based on the fTRD peak data, the location for the female component of S_I was attributed to a 160 kb region in the Nipponbare genome.

Mapping of the female component of the S_I locus in the *O. sativa ssp. japonica* background:

As the *O. sativa* genetic background affects the level of female sterility caused by S_I (KOIDE *et al.* 2008c; SANO 1990), 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* x *O. glaberrima* BC₁F₁ populations were developed, and the chromosome 6 genetic map 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 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_I 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 *japonica* populations, being lower than the one found for the *indica* cross.

Construction of a physical map around the S_I locus in *O. glaberrima*:

In order to compare the region between the two species, a physical map of the S_I locus of *O. glaberrima* was established. Twenty-six BAC clones from the *O. glaberrima* cv. CG14 BAC library were chosen, based on the data available from the OMAP project (KIM *et al.* 2008). PCR amplification with the same markers from the genetic map around S_I 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 confirmation of the overlaps between the clones (Figure III-3 A). Based in the amplification profile, clone OGBBa0049I08 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 X coverage and an error rate below 1 base per 100 kb.

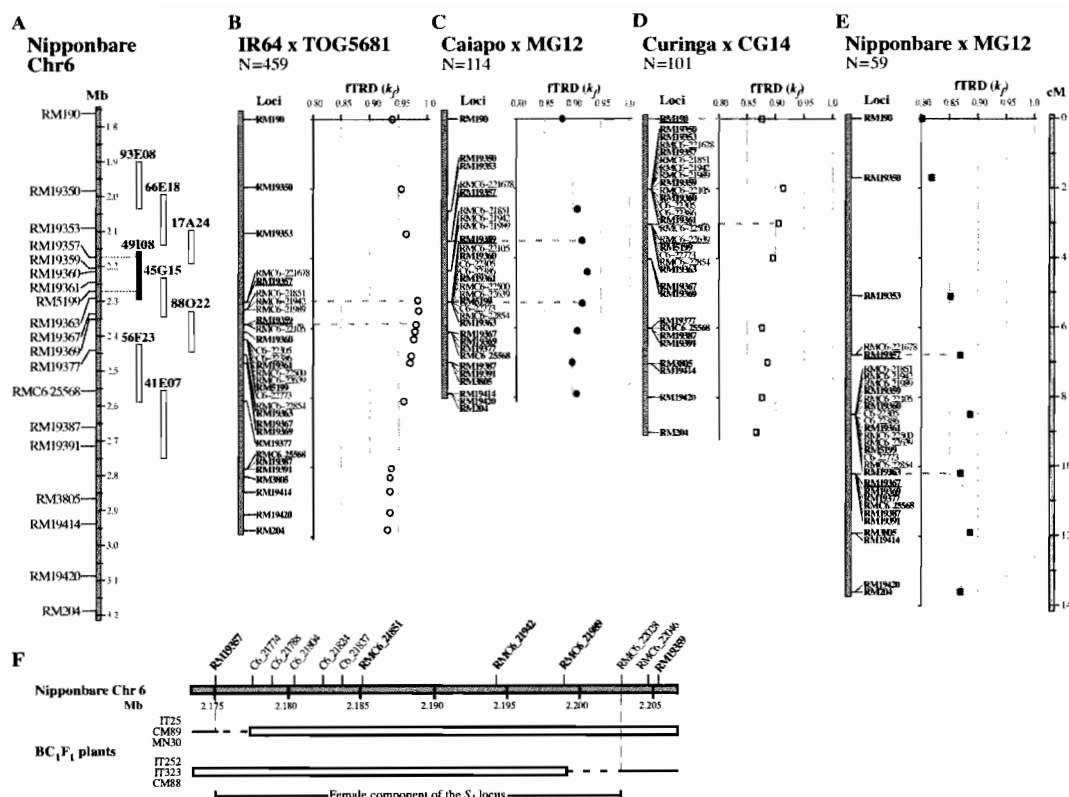


Figure III-3.—Fine mapping of the female component of the S_1 locus. (A) Physical map of the S_1 region in the *O. glaberrima* cv. CG14. Positions of markers around the S_1 locus in the Nipponbare pseudomolecule, and their amplification pattern in the CG14 BAC clones (OG_BB0093E08, OG_BB0066E18, OG_BB0017A24, OG_BB0049I08, OG_BB0045G15, OG_BB0088O22, OG_BB0056F23 and OG_BB0041E07) are shown. Saturation with new markers obtained with the OG_BB0049I08 sequence, in crosses (B) IR64xTOG5681, (C) CAxMG12, (D) CUxCG14 and (E) NIPxMG12. Markers indicated in bold are common with Figure III-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. White bars represent the heterozygote genomic segments, while *O. sativa* homozygote segments are represented by straight lines. Dotted lines symbolize the recombination sites.

Table III-1. Frequency of S_1^s alleles in the *O. sativa* x *O. glaberrima* BC₁F₁ populations.

Populations	Minimal <i>O. sativa</i> inherited region around S_1^s	N	N (S_1^s)	$f(S_1^s)$	$\sigma f(S_1^s)$	$f \pm \sigma f(S_1^s)$	$f \pm 2\sigma f(S_1^s)$
<i>indica</i> x <i>O. glaberrima</i>							
IR64 x TOG5681	RM19350-RM3183	459	9	0.020	0.006	0.013 - 0.026	0.007 - 0.033
<i>japonica</i> x <i>O. glaberrima</i>	RM190-RM3805	274	25	0.091	0.017	0.073 - 0.108	0.056 - 0.126
Caiapo x MG12		114	9	0.079	0.025	0.054 - 0.104	0.028 - 0.129
Curinga x CG14		101	9	0.089	0.028	0.061 - 0.117	0.032 - 0.146
Nipponbare x MG12		59	7	0.119	0.042	0.077 - 0.161	0.034 - 0.203
Additional individuals							
Caiapo x TOG5681		34	1	0.029	0.029	0.000 - 0.058	0.000 - 0.087
Nipponbare x TOG5681		12	1	0.083	0.080	0.004 - 0.163	0.000 - 0.243
<i>O. sativa</i> x <i>O. glaberrima</i>	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 BC₁F₁ plants in *indica* and *japonica* crosses.

Fine mapping of the female component of S_I :

The factor causing fTRD was mapped in the same genomic region for the two *O. sativa* backgrounds. In order to map it more finely, new polymorphic SSR and InDel markers were designed by sequence comparison. Saturation with the new markers was completed in the existing populations and in 334 new BC₁F₁ 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 III-3 B, C). For the CUxCG14 and NIPxMG12 crosses, the values for k_f were respectively 0.911 (RM190-RMC6_22639) and 0.881 (RM19357-RM1967) (Figure III-3 D, E). Consequently, the segment between markers RM19357 and RM19359 was defined as the most probable location for the factor affecting female fertility of the hybrids.

In order to reduce the size of the candidate region, the recombination events were examined around the fTRD peak using new markers. Surprisingly, recombination occurred within the same small segment for the three individuals limiting each side of the interval (Figure III-3 F). This approach allowed us to reduce the S_I female component region to 50.3 kb in *O. glaberrima* and to 27.8 kb in *O. sativa*. This interval 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_I 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.

Allelic state of viable female S_I^s gametes produced by the F₁ hybrids:

The allelic state of the S_I^s viable female gametes produced by the F₁ hybrid was examined at all the chromosome 6 markers. The 36 S_I^s/S_I^s BC₁F₁ plants shared an *O. sativa* chromosomal segment of 7.3 cM ($r = 0.0725$) (Table III-1). The probability of observing no recombination events between the bounds of this segment in 36 lines is $p_0 = (1 - 0.0725)^{36} = 0.066$. It is low enough to raise the hypothesis of the necessary inheritance of other *O. sativa* factors in the vicinity of S_I to ensure the viability of the S_I^s gametes. As S_I is located at 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 the segment (from now on denoted as S_{IA} and S_{IB} , respectively) is suspected.

In the cross with the lowest S_I^s transmission (IR64xTOG5681), the 9 S_I^s/S_I^s plants share a much larger *O. sativa* chromosomal segment (with the exception of a heterozygote < 7 cM

Table III-2. Sequence annotation of the *O. glaberrima* cv. CG14, BAC OG-BBa0049I08.

<i>O. glaberrima</i> gene name	Best BLASTN homology	Best BLASTN EST homology	Best BLASTX homology (swiss prot)	Protein domain	Putative function
OG-BBa0049I08.1 ^p	NM_001063292 <i>O. sativa</i> Os06g0141100 (0.0)	CB618108 <i>O. sativa</i> (0.0)	Q94A82 <i>A. thaliana</i> NADH pyrophosphatase NUDT19 (8e-69)	pfam09296, NUDIX-like; cd03429, NADH pyrophosphatase	Putative NADH pyrophosphatase protein
OG-BBa0049I08.2	CT830197 <i>O. sativa</i> flcDNA (0.0)	CT850081 <i>O. sativa</i> (0.0)	O13801 <i>Schizosaccharomyces pombe</i> RNA-binding protein (2e-9)	pfam00641, Zn-finger	Putative Zinc finger protein
OG-BBa0049I08.3	AK121184 <i>O. sativa</i> flcDNA (2e-131)	EE592322 <i>O. sativa</i> (5e-169)	Q02921 <i>Glycine max</i> Early nodulin 93 (2e-11)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.4	AC136219 <i>O. sativa</i> BAC (0.0)	CT843513 <i>O. sativa</i> (1e-92)	P73714 <i>Synechocystis</i> Zinc metalloprotease (1e-9)	cd06163 zinc metalloproteases	Hypothetical protein
OG-BBa0049I08.45*	AP004329 <i>O. sativa</i> chr. 6 (0.0)	/	/	/	Hypothetical protein
OG-BBa0049I08.5	AK122162 <i>O. sativa</i> flcDNA (2e-148)	EL586675 <i>O. sativa</i> (2e-162)	Q02921 <i>G. max</i> Early nodulin 93 (8e-20)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.6	AK121791 <i>O. sativa</i> flcDNA (1e-157)	CI240527 <i>O. sativa</i> (1e-158)	Q02921 <i>G. max</i> Early nodulin 93 (4e-14)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.7	/	/	/	/	Hypothetical protein
OG-BBa0049I08.75	AK107277 <i>O. sativa</i> flcDNA (1e-115)	/	/	/	Hypothetical protein
OG-BBa0049I08.8	NM_001063299 <i>O. sativa</i> Os06g0142000 (0.0)	CB622551 <i>O. sativa</i> (0.0)	Q9SH88 <i>A. thaliana</i> Ribosome biogenesis regulatory protein (3e-48)	/	Putative protein
OG-BBa0049I08.8.5*	AP002838 <i>O. sativa</i> BAC (2e-156)	CT858525 <i>O. sativa</i> (8e-67)	/	/	Hypothetical protein
OG-BBa0049I08.9*	AP004090 <i>O. sativa</i> BAC (8e-69)	CT844843 <i>O. sativa</i> (3e-68)	/	/	Hypothetical protein
OG-BBa0049I08.10*	AC121363 <i>O. sativa</i> BAC (3e-128)	CB649669 <i>O. sativa</i> (1e-57)	P47927 <i>A. thaliana</i> Floral homeotic protein APETALA 2 (4e-8)	/	Hypothetical protein
OG-BBa0049I08.11	AK073027 <i>O. sativa</i> flcDNA (0.0)	CB658549 <i>O. sativa</i> (0.0)	Q9FJT2 <i>A. thaliana</i> F-box/FBD/LRR-repeat (5e-4)	/	Putative protein
OG-BBa0049I08.12	AK122162 <i>O. sativa</i> flcDNA (2e-175)	CI252819 <i>O. sativa</i> (2e-175)	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.13	CT837557 <i>O. sativa</i> flcDNA (4e-177)	CT856080 <i>O. sativa</i> (4e-177)	Q02921 <i>G. max</i> Early nodulin 93 (1e-14)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.14	CT827995 <i>O. sativa</i> flcDNA (1e-158)	EE592204 <i>O. sativa</i> (5e-176)	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.15	CT827995 <i>O. sativa</i> flcDNA (9e-173)	CI050172 <i>O. sativa</i> (1e-176)	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	pfam03386, Early nodulin 93 ENOD93 protein	Putative ENOD93 protein
OG-BBa0049I08.16	NM_001063304 <i>O. sativa</i> Os06g0142500 (0.0)	CB617832 <i>O. sativa</i> (0.0)	Q9LMN7 <i>A. thaliana</i> Wall-associated receptor kinase 5 (5e-130)	cd00180, Serine/Threonine protein kinases	Putative Serine/Threonine protein kinases
OG-BBa0049I08.17	NM_001063305 <i>O. sativa</i> Os06g0142600 (0.0)	CB635675 <i>O. sativa</i> (0.0)	O82804 <i>A. thaliana</i> Protein Early flowering 3 (1e-55)	/	Putative Early flowering protein
OG-BBa0049I08.18	AP000399 <i>O. sativa</i> BAC (0.0)	CT860562 <i>O. sativa</i> (0.0)	O23081 <i>A. thaliana</i> Cysteine-rich receptor-like protein kinase 41 (2e-57)	cd00180, Serine/Threonine protein kinases	Putative Serine/Threonine protein kinases
OG-BBa0049I08.19	CT834467 <i>O. sativa</i> flcDNA (0.0)	CT848007 <i>O. sativa</i> (0.0)	P00428 <i>Bos taurus</i> Cytochrome c oxidase subunit 5B, mitochondrial (6e-14)	cd00924, Cytochrome c oxidase subunit Vb	Putative cytochrome oxidase
OG-BBa0049I08.20	AK103971 <i>O. sativa</i> flcDNA (0.0)	CA998930 <i>O. sativa</i> (0.0)	Q9SNQ2 <i>O. sativa</i> protein arginine N-methyltransferase (0.0)	cd02440, S-adenosylmethionine-dependent methyltransferases	Putative methyltransferase protein
OG-BBa0049I08.21	AK121920 <i>O. sativa</i> flcDNA (0.0)	EC366141 <i>O. sativa</i> (0.0)	Q38854 <i>A. thaliana</i> 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
OG-BBa0049I08.22	AK071301 <i>O. sativa</i> flcDNA (0.0)	CB672512 <i>O. sativa</i> (0.0)	P22302 <i>Nicotiana glauca</i> Superoxide dismutase chloroplastic (5e-60)	PRK10543, superoxide dismutase; pfam02777, Iron/manganese superoxide dismutases	Putative Superoxide dismutase
OG-BBa0049I08.23	AK061597 <i>O. sativa</i> flcDNA (9e-80)	CI020918 <i>O. sativa</i> (9e-80)	/	/	Putative protein
OG-BBa0049I08.24	AK120946 <i>O. sativa</i> flcDNA (0.0)	CX101001 <i>O. sativa</i> (0.0)	Q9SQI3 <i>Gossypium hirsutum</i> Myristoyl-acyl carrier protein thioesterase, chloroplastic (6e-146)	cd00586, 4-hydroxybenzoyl-CoA thioesterase (4HBT)	Putative Myristoyl-acyl carrier protein thioesterase
OG-BBa0049I08.25	AP003487 <i>O. sativa</i> BAC (0.0) <i>O. sativa</i> BAC (0.0)	CA083999 <i>Saccharum officinarum</i> (1e-99)	O49287 <i>A. thaliana</i> Putative pentatricopeptide repeat-containing protein (2e-93)	pfam01535, PPR repeat	Putative PPR protein
OG-BBa0049I08.26	AP003487 <i>O. sativa</i> BAC (0.0)	CF326731 <i>O. sativa</i> (0.0)	Q80SY5 <i>M. musculus</i> Pre-mRNA-splicing factor 38B (2e-40)	pfam03371, PRP38 family	Putative protein
OG-BBa0049I08.27 ^p	AK067270 <i>O. sativa</i> flcDNA (0.0)	FL869738 <i>Panicum virgatum</i> (0.0)	Q9LW86 <i>A. thaliana</i> Probable sulfate transporter (1e-141)	pfam00916, Sulfate transporter family	Putative sulfate transporter

p: Partial gene, *: pseudogene. The E-value for the BLAST is presented in brackets. Predicted genes in grey are located in the *S_I* locus, between markers RM19357 and RMC6_22028.

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_I . 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 an additional factor, ($S_I C$) located between markers RM204 and RM3183 plays an additional role in the selection of the gametes in this cross. Taken together, these results indicate that two linked factors present in the *O. sativa* genome are necessary for the survival of female gametes carrying S_I^s , and that an additional factor is also required in the IR64xTOG5681 cross.

Sequence annotation of the *O. glaberrima* BAC OG-BBa0049I08:

The 164,664-bp sequence of the *O. glaberrima* BAC OG-BBa0049I08 was carefully annotated and analyzed. From the 28 Non-TE predicted genes, all but one (49I08.7) were validated by sequence alignments with *O. sativa* ESTs and FLcDNA (Table III-2). Three genes (49I08.8.5, 49I08.9, and 49I08.10) were predicted to be non-functional, because the lack of the start exons and frame-shift mutations. 74 TEs derived sequences were identified, accounting for 45 kb of sequence (27,4% of the whole BAC) (Supplemental Table III-3). Interestingly, three transposons enclosed predicted genes 49I08.4, 49I08.9 and 49I08.10 (Supplemental Figure III-2). They were classified as non-autonomous chimeric Mutator-like transposable elements (pack-MULEs) (JIANG *et al.* 2004). Pack-MULEs identified on the BAC showed multiple copy numbers in the Nipponbare genome, and their captured sequences were identified as complete and functional genes (NM_001057082, NM001058016 and NM_001065779 for 49I08.4, 49I08.9 and 49I08.10 respectively) with DNA sequence identity ranging from 75 to 93%.

Direct comparisons between *O. sativa* cv. Nipponbare and *O. glaberrima* cv. CG14 orthologous sequences revealed an overall high-level of conservation. However, the alignment exhibited a series of breaks mainly due to TE insertions in either of the genomes (Supplemental Figure III-2). Gaps were located in the first half of the BAC, while the rest displayed a perfect straight line. The two major breaks were caused by LTR retrotransposons of the Gypsy-type (positions 84,216-95,545 in BAC OG-BBa0049I08, and 2,215,691-2,228,590 in Nipponbare's chromosome 6).

Comparison of the S_I locus between *O. glaberrima* and *O. sativa*:

In order to study the sequence divergence of the S_I locus, fine comparative genomic approaches were carried out between CG14 and Nipponbare. Nucleotide sequences

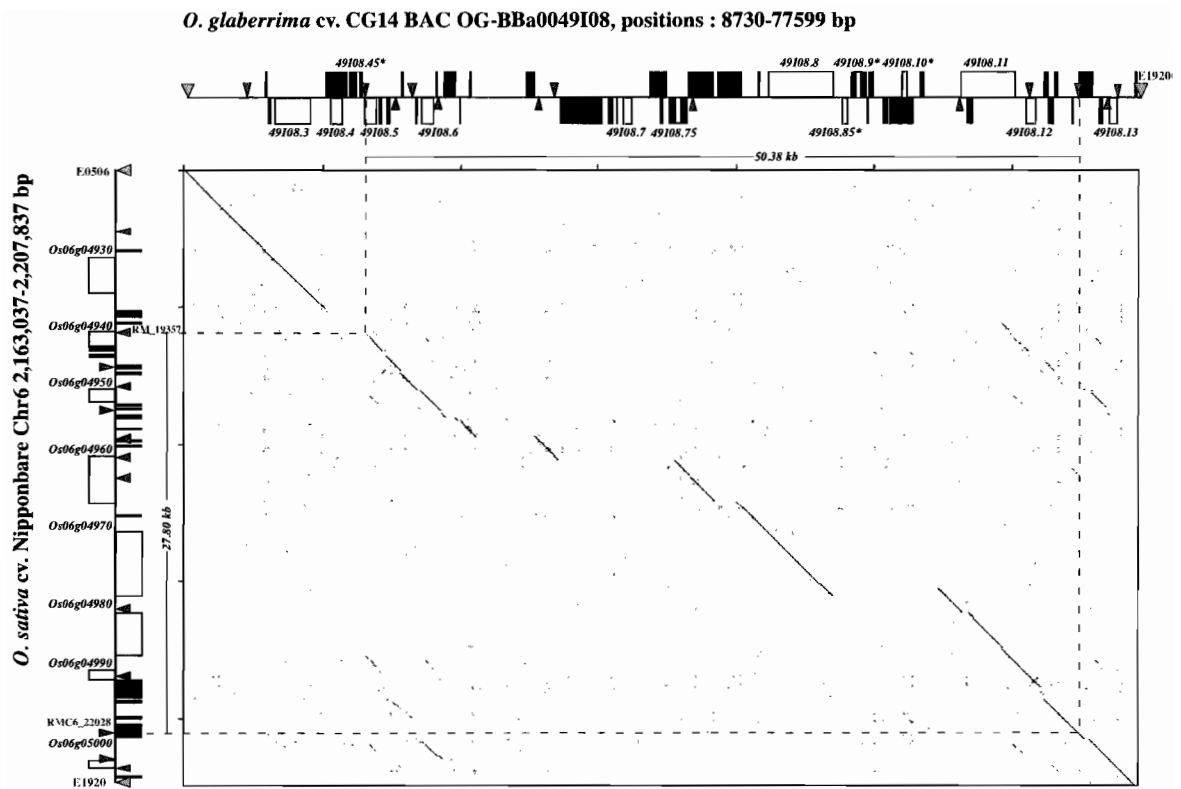


Figure III-4.—Graphical comparison of the *O. sativa* and *O. glaberrima* sequences in the S_1 locus. Pairwise alignments of the respective 44.8 and 68.8 kb sequences of *O. sativa* and *O. glaberrima* in the S_1 locus, limited by markers E0506 and E1920 (KOIDE *et al.* 2008c). Predicted genes (empty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

surrounding the male component of the locus mapped by Koide (2008b) (between markers E506 and E1920) were extracted from both genomes and analyzed (Figure III-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 III-4). This 1.5 X 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 *S₁* locus, six genes and three pseudogenes were predicted (Table III-2). Among the predicted genes, one is a hypothetical protein (49I08.7), three exhibited similarities with “Early nodulins”, one with ribosome biogenesis regulatory proteins (49I08.8), and one with F-box proteins (49I08.11). The three pseudogenes were located between predicted genes 49I08.8 and 49I08.11, and two of them (49I08.9, 49I08.10) corresponded to the pack-MULEs carrying fragments of cellular genes.

The gene content in the segment was found variable between the two species. Neither of the pseudogenes nor the hypothetical protein were found at the collinear positions in *O. sativa* (Supplemental Table III-4 and Figure III-4). In contrast, hypothetical gene Os06g04960 from *O. sativa* was not found to be collinear with *O. glaberrima*. Re-annotation 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 gene 49I08.7. 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 non-autonomous transposon. All the orthologous genes within the segment were found highly conserved at the nucleotide and protein levels, with sequence identities ranging from 96 to 100% and from 82.8 to 100%, respectively.

The Ka/Ks rates were calculated for the 19 orthologous predicted genes present in the CG14 BAC and Nipponbare sequences (Supplemental Table III-4). Most of them appeared to be under purifying selection, with the exception of four genes outside the *S₁* interval that are clearly under strong positive selection ($Ka/Ks \gg 1$ in 49I08.1, 49I08.19, 49I08.21 and 49I08.22). Despite an overall purifying evolution applied to coding regions, four genes show high non-synonymous changes. Three of them are located within the *S₁* interval (49I08.6, 49I08.8 and 49I08.11).

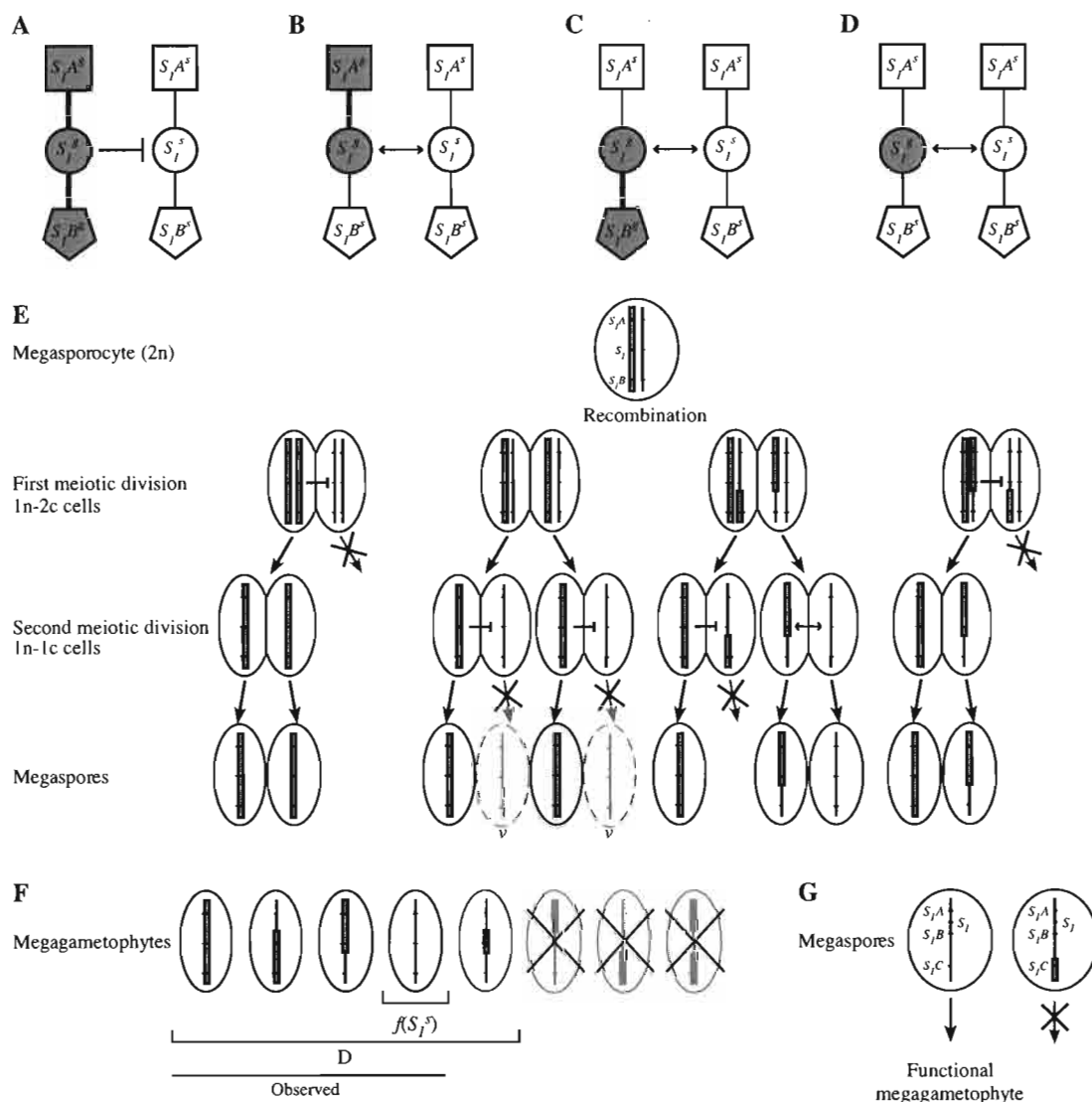


Figure III-5.—Predicted model for the allelic gamete elimination involving S_I . Genetic model based in a BDM incompatibility that explains the allelic elimination of female gametes involving the S_I locus. **(A)** In cells that carry $S_I A^g$, S_I^g and $S_I B^g$ (gray) together with $S_I A^s$, S_I^s and $S_I B^s$ alleles (white), the epistatic interaction between the three *O. glaberrima* factors (thick lines) will cause the abortion (blunt-end arrow) of daughter cells carrying S_I^s . This allelic abortion will thus originate a fTRD. When recombination eliminates $S_I B^g$ **(B)**, $S_I A^g$ **(C)**, or both **(D)**, the epistatic incompatible interaction acting against S_I^s would not be present (open arrow), allowing the development of S_I^s gametes and eliminating the fTRD. **(E)** Model explaining how recombination events between $S_I A$ and $S_I B$, 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 $S_I A^g$, S_I^g and $S_I B^g$, only cells that possess at least one copy of S_I^g will continue their development after each meiotic division. When recombination between $S_I A$ and $S_I B$ occurs, the epistasis against S_I^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_I^s bearing megaspores; however, only the non-recombinant $S_I A^s$ - S_I^s - $S_I B^s$ are able to continue their development. Depending on the alleles confronted in a given cross or the existence of additional factors in the genome, a viability rate of the $S_I A^s$ - S_I^s - $S_I B^s$

Discussion

TRD and its association with interspecific sterility loci:

In plants, post-zygotic barriers such as hybrid sterility are a very common and important phenomena, driving the establishment and conservation of species (WIDMER *et al.* 2009). The strong association between markers showing TRD and pollen sterility QTLs has been already demonstrated in an interspecific cross of Tomato (MOYLE and GRAHAM 2006). This association suggests that hybrid sterility may arise from incompatibilities 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 QTLs causing *O. sativa* x *O. glaberrima* hybrid sterility evidences their colocalization with markers showing fTRD (Figure III-1). Differences in the intensity and the direction of the fTRD were found in comparison with the genetic map obtained by Li (2008) for chromosomes 1, 2, 3, 4, 7 and 12. 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). Co-localization of fTRD sites with pollen sterility loci could imply that these loci have also a role in female gamete development or in post-pollination, as fTRD cannot be caused by loci affecting only male gametogenesis. Occurrence of fTRD in genomic regions where no sterility locus 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 a mechanism other than gamete elimination that bias the segregation. The TRD in the long arm

cells (ν) is applied. As a result of the allelic incompatibilities, only the chalazal-most megaspores that originate from a viable configuration during meiosis would be capable to produce a functional embryony 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 BC₁F₁ plants evaluated. **(G)** Female gamete elimination based on the presence of an additional factor (S_1C), that causes the abortion of the of S_1A^s - S_1^s - S_1B^s - S_1C^g gametes. Gray bars and straight lines represent respectively the *O. glaberrima* and *O. sativa* chromatides.

of chromosome 6, which has also been described for another *O. sativa* x *O. glaberrima* cross (LI *et al.* 2008a), involves different markers in the two *O. sativa* subspecies (Figure III-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, involved in the viability of gametes, zygotes, or endosperm in the *O. sativa* x *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* (KOIDE *et al.* 2008b; MATSUBARA *et al.* 2003), 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 either closely linked or 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. However, only a few factors responsible for inter or intraspecific sterility have been fine mapped (KOIDE *et al.* 2008c; LI *et al.* 2007; LI *et al.* 2006; LI *et al.* 2008b; ZHAO *et al.* 2007), and only two have been recently identified (CHEN *et al.* 2008; LONG *et al.* 2008). Indeed, the mapping strategy that has been used requires the isolation of the genetic region into near isogenic lines and the production and phenotypic analysis of many recombinant plants.

We developed a innovative approach that allows to quickly fine map these genes, based on the effect that they have on their own segregation ratio (PHADNIS and ORR 2009). The approach is similar to the one used to identify induced mutations in genes crucial for male and female gametogenesis in *Arabidopsis* (GRINI *et al.* 1999; HOWDEN *et al.* 1998). It is based on the variations of marker TRD due to the hitchhiking effect with the factor causing the allelic elimination of gametes (CHEVIN and HOSPITAL 2006). These variations are inversely proportional to the genetic distance between markers and the factor, and derive in a maximal marker TRD in the genomic region containing the distorter.

This methodology presents 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 F₂ or BC₁F₁ populations. Second, it is accurate since 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 any TRD locus located in a genomic region with a good recombination ratio, providing that enough markers are available in the

species under study. However, its application limits to cases where highly significant differences in the level of TRD are observed.

The complex nature of the S_I locus:

The sterility barrier due to the S_I 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 the new developed mapping methodology in four BC₁F₁ populations, a region of 27.8 kb in *O. sativa* cv. Nipponbare, and of 50.3 kb in *O. glaberrima* cv. CG14 were identified as the most probable location of the female component of the S_I locus. Surprisingly, this component maps in *indica* and *japonica* 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_I^s gametes in both *O. sativa* backgrounds. In the case of a single factor affecting both gametogenesis, its expression or function may be different, or it may interact with additional gene products in female gametogenesis, in order to cause total elimination of S_I^s from pollen, and partial allelic elimination in megagametophytes.

Based on the allelic structure of the surviving S_I^s female gametes, the presence of two linked elements affecting their viability was detected (S_{IA} and S_{IB}), and evidence for an additional factor (S_{IC}) was also found in one cross. Because S_I alone is capable of inducing mTRD but not fTRD (KOIDE *et al.* 2008c), it is therefore possible that an epistatic interaction between S_I , S_{IA} and S_{IB} is the main cause of fTRD and female sterility in *O. sativa* x *O. glaberrima* hybrids. In this scenario, the additional S_{IC} factor would reinforce the sterility barrier in certain crosses, reducing the viability of recombinant gametes (BARTON and DE CARA 2009). The higher S_I^s elimination found in crosses involving TOG5681 with *indica* and *japonica* accessions (Table II-1), may suggest that the elevated fTRD value does not depend exclusively on the *indica* or *japonica* background, but also on the presence of the S_{IC} incompatibility between certain *O. glaberrima* and *O. sativa* accessions. Nevertheless, this concept needs confirmation due to the low number of plants issued from these additional crosses.

The hypothesis that epistatic interactions between linked factors are responsible for the female sterility and fTRD explains well previous observations, describing how the loss by recombination of the upstream, downstream or both *O. glaberrima* segments around S_I^g , causes the recovery of female fertility and the elimination of fTRD (KOIDE *et al.* 2008c; SANO 1990). Furthermore, it has already been proved that a similar, complex conspecific epistasis causes sterility and TRD in *Drosophila pseudoobscura* inter-subspecific hybrids, where only

males possessing an entire X-chromosome block (and not the ones carrying independently each one of the involved loci) are sterile (ORR and IRVING 2005).

Based on our data, the location for S_1A and S_1B can be predicted between marker 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 inter-subspecific sterility locus S_5 (also called wide compatibility gene) is located within this large region (CHEN *et al.* 2008; QIU *et al.* 2005).

Predicted model for the reproductive sterility barrier between *O. sativa* and *O. glaberrima* mediated by S_1 :

The origin of post-zygotic barriers has been explained by the Bateson-Dobzhansky-Muller (BDM) model as the accumulation of incompatibilities in epistatic interactions between two or more genes (MULLER 1942). As previously mentioned, 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 intervene in the allelic elimination, as the F_1 plants produce neither $S_1A^g-S_1^s$ nor $S_1^s-S_1B^g$ gametes.

In light of the data available, a model where BDM incompatibilities are the cause of the female sterility barrier between *O. sativa* and *O. glaberrima* was composed (Model 1 in Supplemental Figure III-3). The model is based on the necessary inheritance of an *O. sativa* gene block around S_1^s , and on the recombination rates within and around the S_1A-S_1B interval. It predicts the S_1^s and S_1^g frequencies observed in the populations, the observed recombination fractions around S_1 , and the final seed set of the hybrids. The proposed model states that in 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 III-5). When recombination removes S_1A^g or S_1B^g , the epistatic interaction acting against S_1^s would cease, allowing the development of the 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, only 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 to calculate the expected frequency of S_1^s gametes ($f(S_1^s)$) and the total rate of gamete survival (D), based on the probability of observing a recombination within the S_1A-S_1 (r_1) and S_1-S_1B (r_2) segments, the genetic distance between S_1B and the centromere (r_5), and on a viability rate of the $S_1A^s-S_1^s-S_1B^s$ cells (v). This viability rate accounts for the possible differences in the allelic elimination of S_1^s gametes, since their survival may need the interaction with additional factors in the genome. Additionally,

viability could be also affected by intraspecific polymorphism in the epistatic factors (WIDMER *et al.* 2009), which could cause differences in the levels of incompatibility, depending on the alleles confronted in a given cross.

The $f(S_I^s)$, r_1 , r_2 and D values predicted by Model are:

$$f(S_I^s) = (1 - r_3) \left[\frac{r_2}{2} + r_5 \left[v - 2v(r_1 + r_2 - 2r_1r_2) + r_1 - r_2 - 2r_1r_2 \right] \right] / D,$$

$$D = (1 - 2r_3) \left(\frac{r_2 + 1}{2} \right) + 2r_5 \left[\frac{1}{2} + \frac{r_1}{2} - r_1r_2 + v \left(\frac{1}{2} - r_1 - r_2 + 2r_1r_2 \right) \right],$$

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

When applying the model with the recombination rates found in the three *japonica* populations taken together, without the $S_I C$ effect ($r_{1\max} = 0.024$, $r_{2\max} = 0.034$), and with an arbitrary value of v (e.g. 0.125); the calculated $f(S_I^s)$ (9.08%) fits into the $f(S_I^s) \pm \sigma(f(S_I^s))$ range observed in these crosses (Table III-1). The predicted D is 55.0%, which is in accordance with the seed set observed in *japonica* x *O. glaberrima* backcrosses (our unpublished data; SANO 1990).

In the IR64xTOG5681 cross, the $f(S_I^s)$ value also depends on the presence of $S_I C$. When applying the model with the recombination rates of the cross ($r_{1\max} = 0.033$ and $r_{2\max} = 0.049$), and with the additional reduction due to the abortion of the $S_I A^s - S_I^s - S_I B^s - S_I C^s$ gametes (with a $r_3 = 0.349$ that represents the maximal recombination ratio between $S_I B$ and $S_I C$), a $f(S_I^s)$ of 2.23% and a D of 51.1% are predicted, without having to apply a viability factor ($v = 0$). This calculated frequency fits well within the significance interval for the cross (Table III-1). Consequently, with $v = 0$, the formulas for D and $f(S_I^s)$ applicable to the cross convert to:

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

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

and the estimates for r_1 and r_2 to: $\hat{r}'_1 = \frac{r_1 \max}{2D'}$ and $\hat{r}'_2 = \frac{r_2 \max}{2D'}$.

The predictions made by our model support the hypothesis that female sterility in the *O. sativa* x *O. glaberrima* hybrids is caused by the allelic elimination of gametes, due to an incompatible epistatic interaction between S_I and the additional factors. Differences in the observed fTRD between *indica* and *japonica* could be due to polymorphisms within the two subspecies in $S_I A$, S_I , $S_I B$, or $S_I C$, or to the existence of other unlinked epistatic factors. Accordingly, the final allelic frequencies for a given cross would be determined by the alleles confronted in a given cross, the recombination ratio between $S_I A$ and $S_I B$, and by the $S_I C$ -mediated incompatibility.

Genomic variations in the S_I orthologous region, and candidate genes:

Comparative genomic approaches between orthologous regions of closely related species are powerful tools to study genome evolution, to identify novel genes and regulatory elements, and to infer candidate factors for a particular trait from sequence divergences. We used orthologous comparative genomic approaches between *O. glaberrima* and *O. sativa* at the S_I interval, to explore structural changes as putative candidates for the locus. Despite the recent divergence of the Asian and African rice species (approximately 0.6 Mya (AMMIRAJU *et al.* 2008; MA and BENNETZEN 2004)), the S_I interval showed an unusual high number of structural changes, in contrast to the overall BAC sequence. Most of these genomic variations occurred within the *O. glaberrima* intergenic regions, mainly through gene translocations. Insertions of pack-MULEs are the cause of the acquisitions of genes 49I08.9 and 49I08.10, while for genes 49I08.7 and 49I08.8.5 the mechanism of acquisition remains unclear. Pack-MULEs have an important functional role in rice genome evolution, as they capture and relocate gene fragments in other genomic context (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 parental gene expression (HANADA *et al.* 2009). Interestingly, the *O. glaberrima* gene 49I08.10 shows similarities with a putative 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). This pack-MULE could be considered as a putative candidate factor for S_I , as it might affect the expression level of its parental gene in the heterozygotes, in a dose-dependant mechanism similar to hybrid dysgenesis (MICHALAK 2009).

Divergent evolution between closely related species may be at the origin of reproductive isolation. Given that in *Drosophila*, a high level of gene divergence and strong positive selection appeared to be mechanisms that influence the evolution of speciation genes (BARBASH *et al.* 2004; PHADNIS and ORR 2009; WU and TING 2004), we investigated nucleotide divergence of orthologous coding regions between CG14 and Nipponbare, to tentatively deduce candidates for S_I . Recently, 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 insertion 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_I locus in *O. glaberrima* might imply a fast evolution of the segment. Three genes within the S_I interval were affected by high nonsynonymous divergences (Supplemental Table III-4), indicating that significant amino-acid changes occurred between CG14 and Nipponbare. Among these three genes, 49I08.11 belongs to the

super-family of F-box proteins. Members of this family constitute protein complexes involved in the control of a wide range of process, including organogenesis, embryogenesis, and cell cycle progression (XU *et al.* 2009). Incompatibilities between members of such protein complexes can cause sterility, as demonstrated for the complex composed by an F-box and a SUMO E3 Ligase-like protein, that controls the inter-subspecific hybrid sterility mediated by the *Sa* locus in *O. sativa* (LONG *et al.* 2008). A high divergence and similarities with F-box proteins promote the 49108.11 gene as an additional putative candidate factor for *S_I*.

Perspectives:

In this work, we describe and model the complex nature of the allelic female gamete elimination mediated by the *S_I* locus. The knowledge generated by our results will contribute to a better understanding of plant interspecific sterility barriers, and will 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_I*, *S_IA* and *S_IB* 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 fertile progeny when crossed with *O. sativa*. More information is still needed in order to characterize all the factors involved, and the mechanism behind the allelic gamete elimination in hybrids from Asian and African rice species.

Acknowledgements

This work was supported by the Generation Challenge Program. The DSF from IRD, and CIAT provided the PhD scholarship of A. Garavito; 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.

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 Research* **25**: 3389-3402.
- 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.

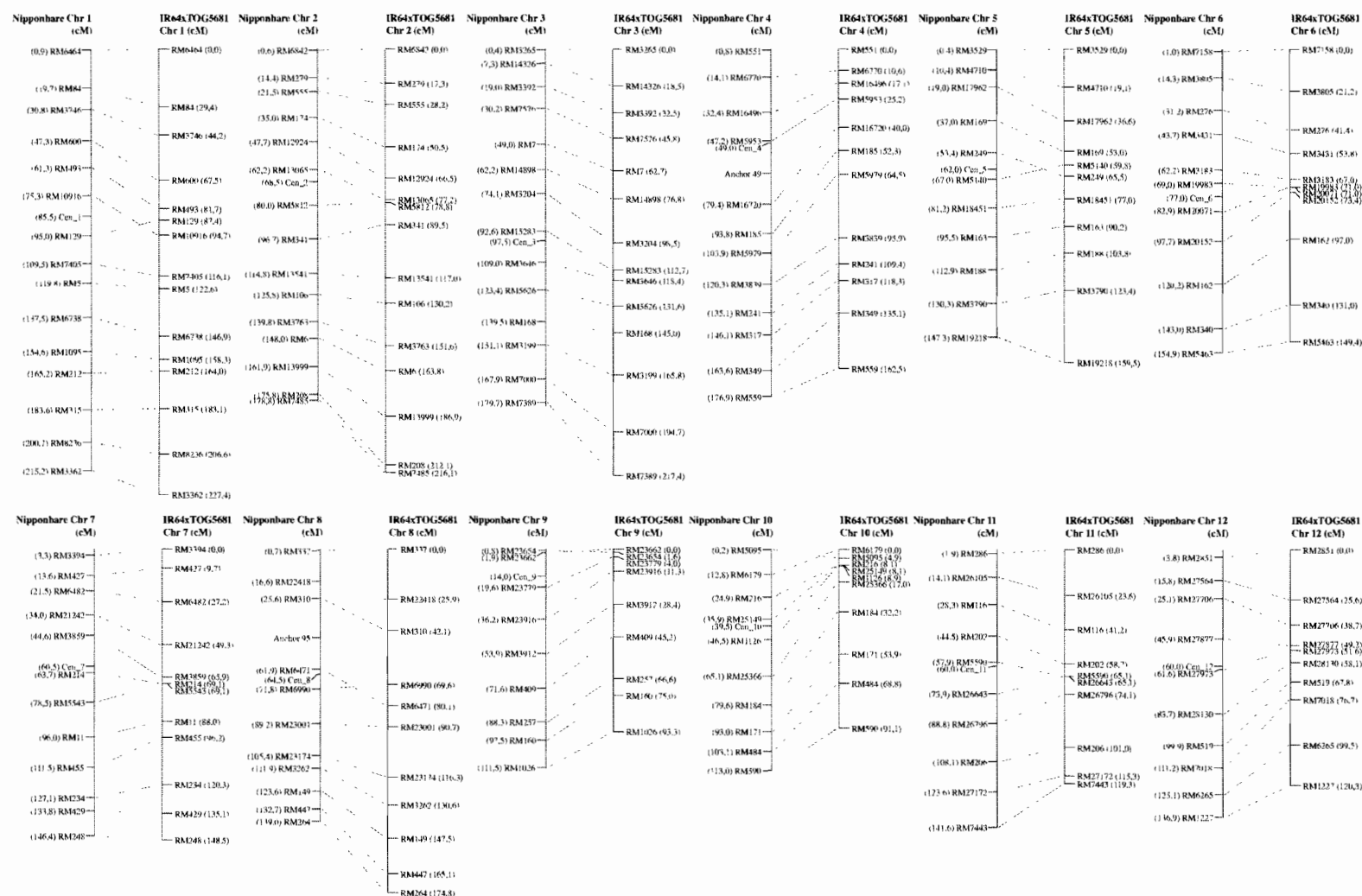
- BARBASH, D. A., P. AWADALLA and A. M. TARONE, 2004 Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biology* **2**: e142.
- BARTON, N. H., and M. A. R. DE CARA, 2009 The evolution of strong reproductive isolation. *Evolution* **63**: 1171-1190.
- 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 & Development* **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. *Proceedings of the National Academy of Sciences* **105**: 11436-11441.
- CHEVIN, L. M., and F. HOSPITAL, 2006 The hitchhiking effect of an autosomal meiotic drive gene. *Genetics* **173**: 1829-1832.
- CHU, Y.-E., H. MORISHIMA and H.-I. OKA, 1969 Reproductive barriers distributed in cultivated rice species and their wild relatives. *Japanese Journal of Genetics* **44**: 207-223.
- 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 Research* **57**: 57-69.
- DOI, K., K. TAGUCHI and A. YOSHIMURA, 1998 A new locus affecting high F₁ pollen sterility found in backcross progenies of Japonica rice and African rice. *Rice Genetics Newsletter* **15**: 146-148.
- DOI, K., K. TAGUCHI and A. YOSHIMURA, 1999 RFLP mapping of *S20* and *S21* for F₁ pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. *Rice Genetics Newsletter* **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 Research* **37**: D992-995.
- GHESEQUIÈRE, 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.
- GRINI, P. E., A. SCHNITTGER, H. SCHWARZ, I. ZIMMERMANN, B. SCHWAB *et al.*, 1999 Isolation of ethyl methanesulfonate-induced gametophytic mutants in *Arabidopsis thaliana* by a segregation distortion assay using the multimarker chromosome 1. *Genetics* **151**: 849-863.
- 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. *Theoretical and Applied Genetics* **107**: 902-909.
- HOWDEN, R., S. K. PARK, J. M. MOORE, J. ORME, U. GROSSNIKLAUS *et al.*, 1998 Selection of T-DNA-tagged male and female gametophytic mutants by segregation distortion in *Arabidopsis*. *Genetics* **149**: 621-631.
- 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 co-linear analysis with *S22* in *O. glumaepatula*. *Euphytica* **151**: 273-278.
- IRGSP, INTERNATIONAL RICE GENOME SEQUENCING P., 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.

- JURKA, J., V. V. KAPITONOV, A. PAVLICEK, P. KLONOWSKI, O. KOHANY *et al.*, 2005 Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**: 462-467.
- 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. *Science* **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 Biology* **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 Phytologist* **179**: 888-900.
- KOSAMBI, D., 1944 The estimation of map distances from recombination values. *Annals of Eugenics* **12**: 172-175.
- 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.). *Theoretical and Applied Genetics* **114**: 515-524.
- LI, J., P. XU, X. DENG, J. ZHOU, F. HU *et al.*, 2008a 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 F₁ pollen sterility in rice (*Oryza sativa* L.). *Chinese Science Bulletin* **51**: 675-680.
- LI, W., R. ZENG, Z. ZHANG, X. DING and G. ZHANG, 2008b Identification and fine mapping of *S-d*, a new locus conferring the partial pollen sterility of intersubspecific F₁ hybrids in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **116**: 915-922.
- 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. *Proceedings of the National Academy of Sciences* **105**: 18871-18876.
- LORIEUX, M., M. N. NDJONDJOP and A. GHESQUIERE, 2000 A first interspecific *Oryza sativa* x *Oryza glaberrima* microsatellite-based genetic linkage map. *Theoretical and Applied Genetics* **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*. *Theoretical and Applied Genetics* **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. *Proceedings of the National Academy of Sciences* **106**: 2071-2076.
- MA, J., and J. L. BENNETZEN, 2004 Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences* **101**: 12404-12410.
- 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.
- MOYLE, L. C., and E. B. GRAHAM, 2006 Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. *Molecular Biology and Evolution* **23**: 973-980.

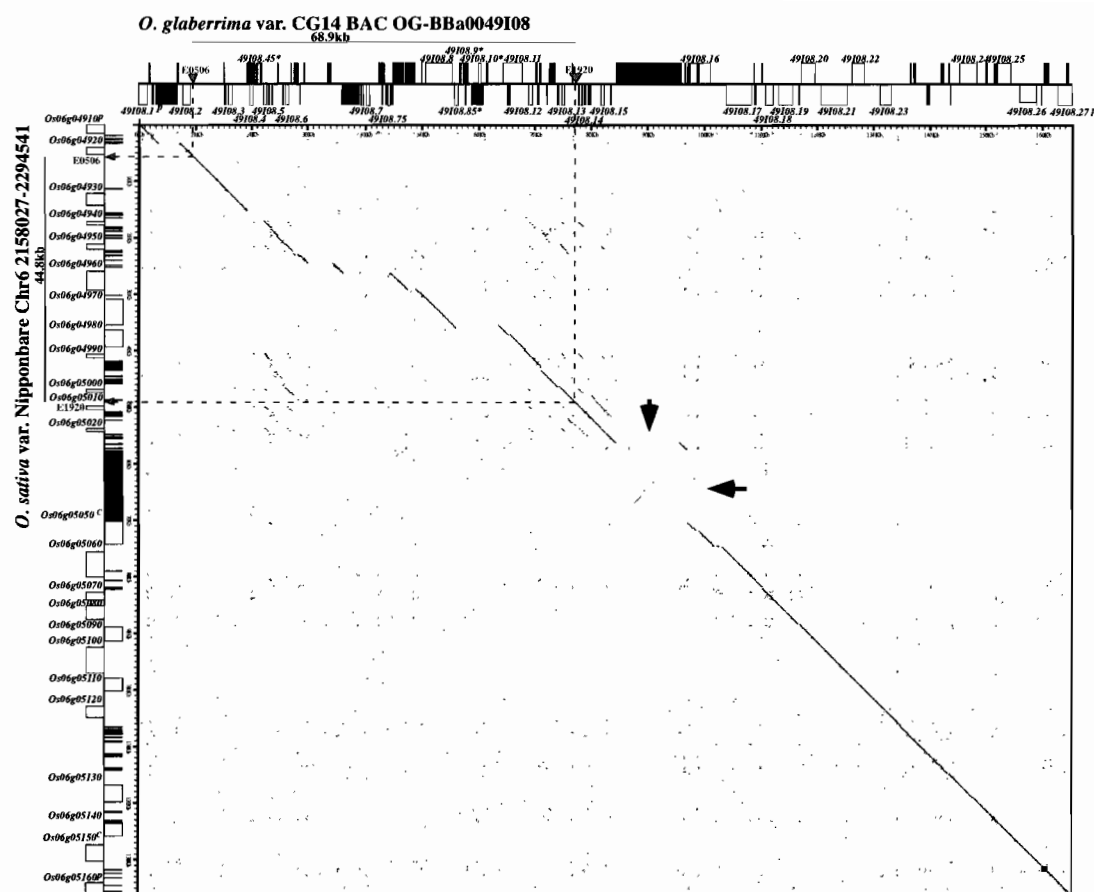
- MULLER, H. J., 1942 Isolating Mechanisms, Evolution and Temperature. Biological Symposium **6**: 71-125.
- NDJIONDJOP, M. N., L. ALBAR, D. FARGETTE, C. BRUGIDOU, M. P. JONES *et al.*, 2003 Mapping a recessive gene conferring resistance to rice yellow mottle virus, pp. 257-259 in *Advances in rice genetics*, edited by G. S. KHUSH, D. S. BRAR and B. HARDY.
- 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.
- ORJUELA, J., A. GARAVITO, M. BOUNIOL, J. ARBELAEZ, L. MORENO *et al.*, 2009 A universal core genetic map for rice. *Theoretical and Applied Genetics* **Doi: 10.1007/s00122-009-1176-1**.
- ORR, H. A., and S. IRVING, 2005 Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. *Genetics* **169**: 671-682.
- 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 Research* **32**: D360-363.
- 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 Research* **35**: D883-887.
- PHADNIS, N., and H. A. ORR, 2009 A Single Gene Causes Both Male Sterility and Segregation Distortion in *Drosophila* Hybrids. *Science* **323**: 376-379.
- QIU, S., K. LIU, J. JIANG, X. SONG, C. XU *et al.*, 2005 Delimitation of the rice wide compatibility gene *S5n* to a 40 kb DNA fragment. *Theoretical and Applied Genetics* **111**: 1080-1086.
- REN, G., P. XU, X. DENG, J. ZHOU, F. HU *et al.*, 2005 A new gamete eliminator from *Oryza glaberrima*. *Rice Genetics Newsletter* **22**: 45-48.
- RICE, P., I. LONGDEN and A. BLEASBY, 2000 EMBOS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **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 Research* **10**: 516-522.
- SANO, Y., 1983 Genetic-studies of speciation in cultivated rice .3. A new gene controlling sterility in F₁ hybrids of 2 cultivated rice species - Its association with photoperiod sensitivity. *Journal of Heredity* **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 F₁ sterility between *O. sativa* L. and *O. glaberrima* STEUD. *Japanese Journal of Genetics* **54**: 121-132.
- SARLA, N., and B. P. M. SWAMY, 2005 *Oryza glaberrima*: A source for the improvement of *Oryza sativa*. *Current Science* **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 Research* **34**: W609-612.
- 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 F₁ pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. Rice Genetics Newsletter **16**: 70-71.
- 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.). Theoretical and Applied Genetics **106**: 411-422.
- TOSHIO, M., and S. FOLKE, 1962 A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiologia 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 Research **35**: W71-74.
- WIDMER, A., C. LEXER and S. COZZOLINO, 2009 Evolution of reproductive isolation in plants. Heredity **102**: 31-38.
- WU, C.-I., and C.-T. TING, 2004 Genes and speciation. Nature reviews; Genetics **5**: 114-122.
- 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. Proceedings of the National Academy of Sciences **106**: 835-840.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences **13**: 555-556.
- ZEIGLER, R., and A. BARCLAY, 2008 The Relevance of Rice. Rice **1**: 3-10.
- ZHANG, Z., P. XU, F. HU, J. ZHOU, J. LI *et al.*, 2005 A new sterile gene from *Oryza glaberrima* on chromosome 3. Rice Genetics Newsletter **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.

SUPPLEMENTAL DATA



Supplemental Figure III-1. —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.



Supplemental Figure III-2.—Graphical comparison between *O. glaberrima* cv. CG14 BAC OG-BBa0049108 and the orthologous *O. sativa* cv. Nipponbare genomic segment. Pairwise alignments of the *O. glaberrima* cv. CG14 BAC OG-BBa0049108 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.

Supplemental Table III-1. SSR markers and PCR conditions used for saturation of chromosome 6

Molecular marker	Position in Nipponbare Chromosome 6 (TIGR V.6)	Annealing Temperature °C	MgCl₂ 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

Supplemental Table III-2. New SSR and InDels markers in the S1 locus

Molecular marker	Position in Nipponbare Chromosome 6 (TIGR V.6)	Forward Primer	Reverse Primer	Annealing Temperature °C	MgCl₂ concentration mM
RMC6_21678	2 167 869	AGGGAGCTCCAGTGTGAGA	TTGTGCCCAAGCTTTAAGGTA	59	1.5
C6_21774	2 177 472	CAGTTTGGGCATTTTCG	ATAGGCGATCCCTAGAGCAA	61	1.5
C6_21788	2 178 838	GCTGCGATGATAAGGCAATA	CCCCTAGCTAGTTGGTGTGA	61	1.5
C6_21804	2 180 401	CCTTACTTCTTTTGAAGTTGG	GCACCACAGTTTCAGGCTCT	61	1.5
C6_21824	2 182 413	CTGCTACCATCCCTCTGCTC	GGAGCAATTGAGAAGCTG	61	1.5
C6_21837	2 183 714	TGGATAGGGCTATAGGAGCA	CTTCCCTTCTCTGGATCGAC	61	1.5
RMC6_21851	2 185 106	AAGGGTCCTTTCCTTAGCC	AGGGCGATTGACGCACAT	59	1.5
RMC6_21942	2 194 257	TCCAATCCCCCTAATTCAG	TTTACCATGGGGAACCTGGA	59	1.5
RMC6_21989	2 198 927	CCTGATGAGCTCTGCTGATG	GTGCGTGCTTTGTGAATCTG	59	1.5
RMC6_22028	2 202 836	CACCATCCTCCTCCATGTTT	CCCTTTCATTTGCCGTCTA	59	1.5
RMC6_22046	2 204 666	AAGAAGGTGAAGACATCAACGA	GCTGCTGAGTTCATGCGTAA	59	1.5
RMC6_22105	2 210 541	GATTTTGCCGTCAGTTTGG	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	CTGCATCCAAAGTTCGAAGG	ATCTGTTCATGGGGGCTGTA	59	1.5
RMC6_22639	2 263 917	TTGCCATTCATGGAATTTAG	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-BBa0049108 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).

Supplemental Table III-3 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

Supplemental Table III-4. Sequence comparison between orthologous coding sequences in the *O. glaberrima* cv. CG14 BAC OG-BBa0049108 and *O. sativa* cv. Nipponbare

<i>O. glaberrima</i> Gene Name	Putative Function	Putative <i>O. sativa</i> (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	Ka (dN)	Ks (dS)	Ka/Ks
OG-BBa0049108.1 ^p	Putative NADH pyrophosphatase protein	LOC_Os06g04910 ^p	/	/	/	/	/	/	/
OG-BBa0049108.2	Putative Zinc finger protein	LOC_Os06g04920	99.3	99.3	100	438	0.0028	0.0266	0.1063
OG-BBa0049108.3	Putative ENOD93 protein	LOC_Os06g04930	61.3	60.9	62.2	/	/	/	/
OG-BBa0049108.4	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.45*	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.5	Putative ENOD93 protein	LOC_Os06g04940	65	54.1	59.5	/	/	/	/
OG-BBa0049108.6	Putative ENOD93 protein	LOC_Os06g04950	89.1	76.8	78.4	376	/	/	/
OG-BBa0049108.7	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.75	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.8	Putative protein	LOC_Os06g04970	98.6	97.7	98.4	2073	0.0112	0.0237	0.4728
OG-BBa0049108.85*	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.9*	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.10*	Hypothetical protein	no orthologous gene	/	/	/	/	/	/	/
OG-BBa0049108.11	Putative F-box protein	LOC_Os06g04980	98.1	96.5	97.7	1542	0.0173	0.0198	0.8715
OG-BBa0049108.12	Putative ENOD93 protein	LOC_Os06g04990	99.7	100	100	348	0.0000	0.0142	0.001
OG-BBa0049108.13	Putative ENOD93 protein	LOC_Os06g05000	100	100	100	348	0.0000	0.0000	99
OG-BBa0049108.14	Putative ENOD93 protein	LOC_Os06g05010	99.4	100	100	351	0.000025	0.0260	0.001
OG-BBa0049108.15	Putative ENOD93 protein	LOC_Os06g05020	99.1	98.3	99.1	348	0.0074	0.0143	0.5175
OG-BBa0049108.16	Putative Serine/Threonine protein kinases	LOC_Os06g05050 ^C	92	87.8	92.1	2291	0.0447	0.0571	0.7828
OG-BBa0049108.17	Putative Early flowering protein	LOC_Os06g05060	99.2	98.8	99.1	2283	0.0057	0.0155	0.3669
OG-BBa0049108.18	Putative Serine/Threonine protein kinases	LOC_Os06g05070	97.6	97.6	98.5	1386	0.0081	0.1700	0.0465
OG-BBa0049108.19	Putative cytochrome oxidase	LOC_Os06g05080	99.8	99.4	99.4	465	0.0032	0.0000	99
OG-BBa0049108.20	Putative methyltransferase protein	LOC_Os06g05090	98.8	98.4	99	1143	0.0071	0.0299	0.2366
OG-BBa0049108.21	Putative 1-deoxy-D-xylulose-5-phosphate synthase	LOC_Os06g05100	99.9	99.6	99.7	2169	0.0018	0.0000	99
OG-BBa0049108.22	Putative Superoxide dismutase	LOC_Os06g05110	99.9	99.6	100	768	0.0016	0.0000	99
OG-BBa0049108.23	Putative protein	LOC_Os06g05120	/	/	/	/	/	/	/
OG-BBa0049108.24	Putative Myristoyl-acyl carrier protein thioesterase	LOC_Os06g05130	97.4	99	99	1284	0.0044	0.0300	0.1457
OG-BBa0049108.25	Putative PPR protein	LOC_Os06g05140	99.8	99.6	99.6	2292	0.0018	0.0037	0.4789
OG-BBa0049108.26	Putative protein	LOC_Os06g05150 ^C	98	97.2	97.5	1296	0.0070	0.0048	1.471
OG-BBa0049108.27 ^p	Putative sulfate transporter	LOC_Os06g05160 ^p	/	/	/	/	/	/	/

p: Partial gene; *****: Pseudogene; **C:** Nipponbare genes manually corrected. Genes in grey are located in the *S₁* locus.

Supplemental Figure III-3. Supplemental data: S1 model.xlsx

Model with S1A and S1B

			v = 0,125												
Recombination event	Class	1n-2c configuration	Configuration frequency without selection	Genotypic frequency without selection	Model 1b										
			$r1max = 0,024$ $r1 = 0,024$ $r4 = 0,1491$ $r1min = 0,0215$	$r2max = 0,0339$ $r2 = 0,0339$ $r5 = 0,3321$ $r2min = 0,0277$	Survival	Frequency									
$r(S1A-S1B) = r3 = 0,0955$ $r(telomere-centromere) = 0,3846$															
No recombination between S1B and centromere AND	1	<table><tr><td>S1A</td><td>S1</td><td>S1B</td></tr><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>s</td><td>s</td></tr></table>	S1A	S1	S1B	s	s	s	s	s	s	0,298	0,075	aborted	0,000
S1A		S1	S1B												
s	s	s													
s	s	s													
No recombination within S1A-S1B	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,075		0,135					
g	g	g													
g	g	g													
Recombination between S1B and centromere AND	2	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	s	g	g	g	g	g	g	0,589	0,147	partially abort	0,033
s		s	s												
g	g	g													
g	g	g													
No recombination within S1A-S1B	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	s	g	g	g	0,147		0,268					
s	s	s													
g	g	g													
Recombination within S1A-S1B (2 chromatides involved) AND	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>s</td></tr><tr><td></td><td>X</td><td></td></tr></table>	s	s	s	g	g	s		X		0,022	0,005	viable	0,010
s		s	s												
g	g	s													
	X														
No recombination between S1B and centromere	<table><tr><td>s</td><td>s</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	g	g	g	g	0,005	aborted	0,010					
s	s	g													
g	g	g													
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr><tr><td></td><td>X</td><td></td></tr></table>	s	s	s	g	s	s		X		0,015	0,004	aborted	0,000
s		s	s												
g	s	s													
	X														
	<table><tr><td>s</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	g	g	g	g	g	0,004		0,007					
s	g	g													
g	g	g													
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>g</td><td>s</td></tr><tr><td></td><td>X</td><td>X</td></tr></table>	s	s	s	s	g	s		X	X	0,001	0,000	viable	0,000
s		s	s												
s	g	s													
	X	X													
	<table><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	s	g	g	g	g	0,000	aborted	0,000					
g	s	g													
g	g	g													
Recombination within S1A-S1B (2 chromatides involved) AND	4	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>s</td><td>g</td></tr><tr><td></td><td>X</td><td></td></tr></table>	s	s	s	s	s	g		X		0,043	0,011	aborted	0,000
s		s	s												
s	s	g													
	X														
Recombination between S1B and centromere	<table><tr><td>g</td><td>g</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	s	g	g	g	0,011	aborted	0,000					
g	g	s													
g	g	g													
	4	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>g</td><td>g</td></tr><tr><td></td><td>X</td><td></td></tr></table>	s	s	s	s	g	g		X		0,030	0,007	viable	0,014
s		s	s												
s	g	g													
	X														
	<table><tr><td>g</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	s	s	g	g	g	0,007		0,014					
g	s	s													
g	g	g													
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>g</td></tr><tr><td></td><td>X</td><td>X</td></tr></table>	s	s	s	g	s	g		X	X	0,002	0,001	aborted	0,000
s		s	s												
g	s	g													
	X	X													
	<table><tr><td>s</td><td>g</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	g	s	g	g	g	0,001	aborted	0,000					
s	g	s													
g	g	g													
Sum			1,000	1,000	Sum	1,000									
					Sum viable	0,5500									
					Sum S1s	0,0908									
					Sum S1g	0,9092									
			$r1 = 0,0240$		$r1 = 0,0218$										
			$r2 = 0,0339$		$r2 = 0,0308$										
					$1 - r1 = 0,9782$										
					$1 - r2 = 0,9692$										
					$D = 0,5500$										
					$S1s = 0,0500$										
					$f(S1s) = 0,0908$										
					$r1 = 0,0218$										
					$r2 = 0,0308$										

Model with S1A, S1B and S1C

telomere - r4 - S1A - r1 - S1 - r2 - S1B - r3 - S1C - centromere
S1B - r5 - centromere

Recombination event	Class	1n-2c configuration	Configuration frequency without selection	Genotypic frequency without selection	Model 1											
					Survival	Frequency	Frequency after gametic selection (S1C)									
		$r1max = 0,033$ $r1 = 0,033$ $r4 = 0,149$ $r1min = 0,0199$		$r2max = 0,0490$ $r2 = 0,0490$ $r5 = 0,3625$ $r2min = 0,0479$ $r3 = 0,3491$												
		$r(S1A-S1B) = r2 = 0,0955$ $r(telomere-centromere) = 0,3846$														
No recombination between S1B and centromere AND	1	<table><tr><td>S1A</td><td>S1</td><td>S1B</td></tr><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>s</td><td>s</td></tr></table>	S1A	S1	S1B	s	s	s	s	s	s	0,232	0,058	aborted	0,000	0,0000
S1A		S1	S1B													
s		s	s													
s		s	s													
No recombination within S1A-S1B	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,058	0,112	0,1133						
g	g	g														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,058	0,112	0,1133						
g	g	g														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,058								
g	g	g														
g	g	g														
Recombination between S1B and centromere AND	2	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	s	g	g	g	g	g	g	0,611	0,153	aborted	0,000	0,0000
s		s	s													
g		g	g													
g		g	g													
No recombination within S1A-S1B	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	s	g	g	g	0,153	0,295	0,2986						
s	s	s														
g	g	g														
	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	s	g	g	g	0,153	0,000	0,0000						
s	s	s														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,153	0,295	0,2986						
g	g	g														
g	g	g														
Recombination within S1A-S1B (2 chromatides involved) AND	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>g</td><td>s</td></tr><tr><td>g</td><td>g</td><td>s</td></tr></table>	s	s	s	g	g	s	g	g	s	0,025	0,006	viable	0,012	0,0080
s		s	s													
g		g	s													
g		g	s													
No recombination between S1B and centromere	<table><tr><td>s</td><td>s</td><td>X</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	X	g	g	g	0,006	0,012	0,0123						
s	s	X														
g	g	g														
	<table><tr><td>s</td><td>s</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	g	g	g	g	0,006	0,000	0,0000						
s	s	g														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,006	0,012	0,0123						
g	g	g														
g	g	g														
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr></table>	s	s	s	g	s	s	g	s	s	0,016	0,004	aborted	0,000	0,0000
s		s	s													
g		s	s													
g		s	s													
	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr></table>	s	s	s	g	s	s	0,004	0,000	0,0000						
s	s	s														
g	s	s														
	<table><tr><td>s</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	g	g	g	g	g	0,004	0,008	0,0080						
s	g	g														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,004	0,008	0,0080						
g	g	g														
g	g	g														
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>g</td><td>s</td></tr><tr><td>s</td><td>g</td><td>s</td></tr></table>	s	s	s	s	g	s	s	g	s	0,002	0,000	viable	0,001	0,0006
s		s	s													
s		g	s													
s		g	s													
	<table><tr><td>s</td><td>g</td><td>s</td></tr><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>s</td><td>g</td></tr></table>	s	g	s	g	s	g	g	s	g	0,000	0,001	0,0009			
s	g	s														
g	s	g														
g	s	g														
	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>s</td><td>g</td></tr></table>	s	s	s	g	s	g	g	s	g	0,000	0,000	0,0000			
s	s	s														
g	s	g														
g	s	g														
	<table><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>s</td><td>g</td></tr></table>	g	s	g	g	s	g	0,000	0,001	0,0009						
g	s	g														
g	s	g														
Recombination within S1A-S1B (2 chromatides involved) AND	4	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>s</td><td>g</td></tr><tr><td>s</td><td>s</td><td>g</td></tr></table>	s	s	s	s	s	g	s	s	g	0,066	0,017	aborted	0,000	0,0000
s		s	s													
s		s	g													
s		s	g													
Recombination between S1B and centromere	<table><tr><td>s</td><td>s</td><td>X</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	X	g	g	g	0,017	0,000	0,0000						
s	s	X														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>s</td></tr><tr><td>g</td><td>g</td><td>s</td></tr></table>	g	g	s	g	g	s	0,017	0,032	0,0324						
g	g	s														
g	g	s														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,017	0,032	0,0324						
g	g	g														
g	g	g														
	4	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>s</td><td>g</td><td>g</td></tr><tr><td>s</td><td>g</td><td>g</td></tr></table>	s	s	s	s	g	g	s	g	g	0,043	0,011	viable	0,021	0,0137
s		s	s													
s		g	g													
s		g	g													
	<table><tr><td>s</td><td>g</td><td>g</td></tr><tr><td>g</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr></table>	s	g	g	g	s	s	g	s	s	0,011	0,021	0,0211			
s	g	g														
g	s	s														
g	s	s														
	<table><tr><td>g</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr></table>	g	s	s	g	s	s	0,011	0,000	0,0000						
g	s	s														
g	s	s														
	<table><tr><td>g</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>s</td></tr></table>	g	s	s	g	s	s	0,011	0,021	0,0211						
g	s	s														
g	s	s														
	3	<table><tr><td>s</td><td>s</td><td>s</td></tr><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>s</td><td>g</td></tr></table>	s	s	s	g	s	g	g	s	g	0,005	0,001	aborted	0,000	0,0000
s		s	s													
g		s	g													
g		s	g													
	<table><tr><td>g</td><td>s</td><td>g</td></tr><tr><td>g</td><td>s</td><td>g</td></tr></table>	g	s	g	g	s	g	0,001	0,000	0,0000						
g	s	g														
g	s	g														
	<table><tr><td>s</td><td>s</td><td>X</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	s	s	X	g	g	g	0,001	0,002	0,0023						
s	s	X														
g	g	g														
	<table><tr><td>g</td><td>g</td><td>g</td></tr><tr><td>g</td><td>g</td><td>g</td></tr></table>	g	g	g	g	g	g	0,001	0,002	0,0023						
g	g	g														
g	g	g														
			Sum 1,000	1,000			1,000									
					Sum viable	0,518	0,988									
					Sum S1s	0,034	0,022									
					Sum S1g	0,966	0,978									

IV

ANALYSE DETAILLEE DE LA REGION DU LOCUS S_l

Introduction

Dans les chapitres précédents (voir section I.6.3), nous avons décrit le petit nombre de gènes identifié et caractérisé à ce jour chez les animaux et les plantes et qui agissent comme barrières reproductives post-zygotiques. Chez la drosophile, des histoires évolutives contrastées sont relatées pour ces gènes. Ainsi, les analyses des gènes de létalité hybride *Nup96* (Presgraves *et al.*, 2003), *Lhr* (Brideau *et al.*, 2006), *Hmr* (Barbash *et al.*, 2004), et du gène de stérilité hybride *OdsH* (Ting *et al.*, 1998) indiquent une évolution adaptative rapide alors qu'une évolution neutre accélérée est suggérée pour le gène de stérilité hybride *Ovd* (Phadnis et Orr, 2009). En dehors de l'évolution des gènes caractérisés, l'évolution des régions génomiques qui contiennent ces gènes reste mal comprise. Seule la région du gène *Lhr* (Brideau *et al.*, 2006), localisée dans une région hétérochromatique qui évolue très rapidement, a permis de suggérer une action cumulée des gènes et de la structure de la chromatine sur la létalité des hybrides. Chez le riz cultivé asiatique, seulement deux locus de stérilité hybride entre les sous-espèces *indica* et *japonica* ont été identifiés pour lesquels les gènes impliqués ont été caractérisés. Cependant, on sait peu de choses sur les forces évolutives qui s'appliquent sur ces gènes et leur région génomique et généralement peu de données sont disponibles sur l'évolution des régions impliquées dans l'isolement post-zygotique chez les plantes. Récemment, une comparaison génomique entre deux variétés de riz cultivé appartenant aux sous-espèces *indica* et *japonica* a montré une mosaïque de blocs génomiques présentant des divergences hétérogènes (Tang *et al.*, 2006). La présence de régions à forte divergence fut associée à des mécanismes d'adaptation écologique, à la présence de gènes de stérilité et à l'effet de la domestication indépendante de ces sous-espèces. Toutefois, à ce jour, l'évolution des régions impliquées dans d'isolement reproductif post-zygotique entre deux espèces différentes de riz reste inconnue et inexplorée.

Dans le chapitre précédent, nous avons montré le caractère complexe de la stérilité femelle causée par S_I : nos résultats suggèrent que l'interaction de deux autres facteurs génétiquement liés à S_I avec celui-ci a un effet délétère sur la plus grande partie des gamètes portant l'allèle *sativa* S_I^s . En outre, nous avons pu expliquer le taux d'élimination gamétique et la distorsion de la ségrégation des marqueurs moléculaires liés à S_I par un modèle génétique basé sur des interactions épistatiques entre les allèles de ces facteurs incompatibles. A partir de nos données génotypiques, nous avons aussi pu déduire la localisation probable des deux locus, S_{IA} et S_{IB} , qui interagissent avec S_I .

La nature de ces facteurs additionnels, dont la caractérisation doit logiquement apporter des informations précieuses sur le mécanisme génétique de la stérilité femelle des hybrides *O. sativa* x *O. glaberrima*, ont fait l'objet d'une investigation particulière. Pour tenter d'identifier ces facteurs, nous avons réalisé le séquençage et l'annotation de ces locus chez *O. glaberrima* cv. CG14, ainsi que la comparaison avec la région orthologue chez *O. sativa* (Nipponbare). De plus, une analyse comparative a été entreprise pour identifier les mécanismes d'évolution et leur possibles implications dans le processus d'isolement reproductif entre deux espèces géographiquement séparées : *O. sativa* et *O. glaberrima*.

Dans le présent chapitre, nous présentons le séquençage, l'annotation, et l'analyse comparative des séquences génomiques des locus orthologues (S_1A , S_1 et S_1B) entre les riz Africains et Asiatiques. Nous avons trouvé que cette région est soumise à plusieurs mécanismes différents d'évolution, tels que des ruptures de colinéarité, des insertions différentielles d'éléments transposables, ainsi qu'une inversion chromosomique interspécifique, impliquant des conséquences génétiques et évolutives dans cette région. Par contre, la comparaison des taux de divergence ne montre pas une évolution homogène des gènes dans cette région, mais plutôt des évolutions isolées de certains gènes. Une analyse de l'évolution des séquences codantes associée à la présence de variations structurales a conduit à l'identification de plusieurs facteurs candidats pour S_1A , S_1 et S_1B .

Les résultats de ces analyses sont présentés sous forme d'article en préparation dont nous reprenons ici les résultats principaux.

Patterns of evolution of three hybrid sterility loci between Asian and African cultivated rice species.

Romain Guyot, Andrea Garavito, Frédérick Gavory, Sylvie Samain, Alain Ghesquière,
Mathias Lorieux

Abstract:

The rice S_I region provides a unique opportunity to study the evolution of reproductive isolation loci between two closely related and geographically isolated species: the Asian rice *O. sativa* and the African rice *O. glaberrima*. The S_I regions are composed of three adjacent loci S_{IA} , S_I and S_{IB} acting as epistatic factors to promote the allelic elimination of female gametes in hybrids. Recently, a putative F-box-LRR gene has been proposed as a candidate factor for the S_I locus while candidates for S_{IA} and S_{IB} loci remain so far undetermined.

The establishment of a genomic sequence of 813 kb covering completely the S_{IA} , S_I loci and partially the S_{IB} locus in *O. glaberrima* cv. CG14, allowed a comparative analysis with *O. sativa* sequences, enabled the study of the pattern of evolution that may potentially result or contribute to reproductive isolation and speciation processes, and assisted the inference of candidate factors for S_{IA} and S_{IB} loci. We studied the gene divergence and evolution in the S_{IA} , S_I and S_{IB} genomic regions (called here S_I regions). Here, the mean level of divergence (measured in Ks) is similar to two other control-loci evaluated. Beside isolated peaks, no cluster of high divergence was observed in S_I and S_{IB} . In contrast, for S_{IA} , high divergent genes were clustered in a block including one putative F-box gene and three highly divergent LRR genes, suggesting here a local accelerated evolution through ecological adaptation of resistance genes. Beside gene evolution, we found that several S_I regions were reshaped through significant structural variations such as a large sequence insertion, a local duplication, a local invasion of transposable elements and a chromosomal inversion of approximately 45 kb. Detailed analyses provide evidences on the mechanisms of chromosomal inversion as well as the genetic consequences of such rearrangement. These multiple rearrangements participate to the structural genomic differentiation of S_I regions and may contribute in addition to the genes that control the reproductive isolation process between Asian and African rice. Altogether, these patterns of evolution reflect the history of the reproductive

isolation process between Asian and African rice and argue in favor of an allopatric speciation model.

Additionally, we found that the putative F-box-LRR gene proposed as a candidate for S_I is duplicated in the $S_I A$ locus both in Asian and African rice. Comparative analysis of F-box-LRR genes located in the S_I and $S_I A$ loci indicates significant non-synonymous changes particularly in the LRR domain, numerous structural rearrangements between *O. sativa* and *O. glaberrima*, and an additional gene duplication in *O. sativa* (Nipponbare), suggesting that this F-box-LRR gene family may represent strong candidate factors for both $S_I A$, and S_I loci.

MATERIALS AND METHODS

Molecular marker analysis

Genetic markers were designed from the comparison of the Nipponbare sequence with its orthologous CG14 sequence (see below). Microsatellites were found with the MISA script (Thiel *et al.*, 2003), and InDels markers designed by direct comparison of the two sequences. 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 finally evaluated in the BC₁F₁ populations developed from our previous work (Garavito *et al.*, 2009). PCR reactions were carried as described by (Orjuela *et al.*, 2009), with an annealing temperature and magnesium concentration optimized for each primer pair (Supplemental Table IV-1). Separation of the PCR products was carried 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 a M13 tail tag (IRD700 and IRD800).

Sequence analysis and gene annotation method

BAC sequencing was done by the Sanger method. Sequence analysis was done as previously described (Garavito *et al.*, 2009). Briefly, coding regions were predicted *ab initio* using the FGENESH program (Salamov et Solovyev, 2000) and then confirmed by comparative analysis with annotated genes 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) and the physical map diagram was drawn using gff2ps software (Abril et Guigo, 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), from the TIGR repeat database (Ouyang et Buell, 2004), and Retronyza (Chaparro *et al.*, 2007), and finally manually corrected. *De novo* prediction of TEs was performed according to the structure of the different classes of TEs. The final annotation of the BAC sequences was performed using the Artemis tool (Rutherford *et al.*, 2000), and the comparison with the Nipponbare genome was accomplished using dot-plot alignments of the Dotter software (Sonnhammer et Durbin, 1995). Estimations of insertion

time of intact LTR retrotransposons were conducted according to (SanMiguel *et al.*, 1998) with the EMBOSS package using the $1.3 \cdot 10^{-8}$ synonymous substitution/site/year rate of substitution established by (Ma et Bennetzen, 2004). Nucleotide and amino-acid alignments were carried out using ClustalX (Thompson *et al.*, 1997).

Orthologous sequences comparison

The orthologous CG14 *S_I* regions were identified by BLASTn against the *O. sativa* ssp. *japonica* cv Nipponbare pseudomolecules downloaded at the MSU web site (ftp://ftp.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/ release 6.1; September 2006), and against the *O. sativa* ssp. *indica* 93-11 downloaded at the Beijing Genomic Institute web site (<http://rice.genomics.org.cn/rice2/link/download.jsp>). Sequence comparisons were carried out using the Dotter program (Sonnhammer et Durbin, 1995), the Artemis Comparison Tool (Carver *et al.*, 2005), and the EMBOSS package. The downloaded *O. sativa* sequences were re-annotated for genes and TE with similar approaches used to annotate the *O. glaberrima* segment. To study microcollinearity between orthologous *O. glaberrima* and *O. sativa* sequences, the nucleotide sequences of non-TE coding genes were extracted for each segment and used as queries for BLAST alignments between each other to generate a distance matrix. Microcollinearity relationships were displayed using GenomePixelizer software (<http://www.atgc.org/GenomePixelizer/>).

Calculation of nonsynonymous and synonymous nucleotide substitution rates

Orthologous *O. sativa* and *O. glaberrima* coding regions were aligned using the Needle tools (Rice *et al.*, 2000) to estimate the degree of gene structure conservations. Calculations for nonsynonymous and synonymous nucleotides substitution rate were done as previously described (Garavito *et al.*, 2009). Orthologous genes with clear distinct gene structure were removed from the analysis. Identical analyses were carried out with two control loci recently sequenced in *O. glaberrima*: ADH (Ammiraju *et al.*, 2008) from chromosome 11 and MOC1 from chromosome 6 (Lu *et al.*, 2009).

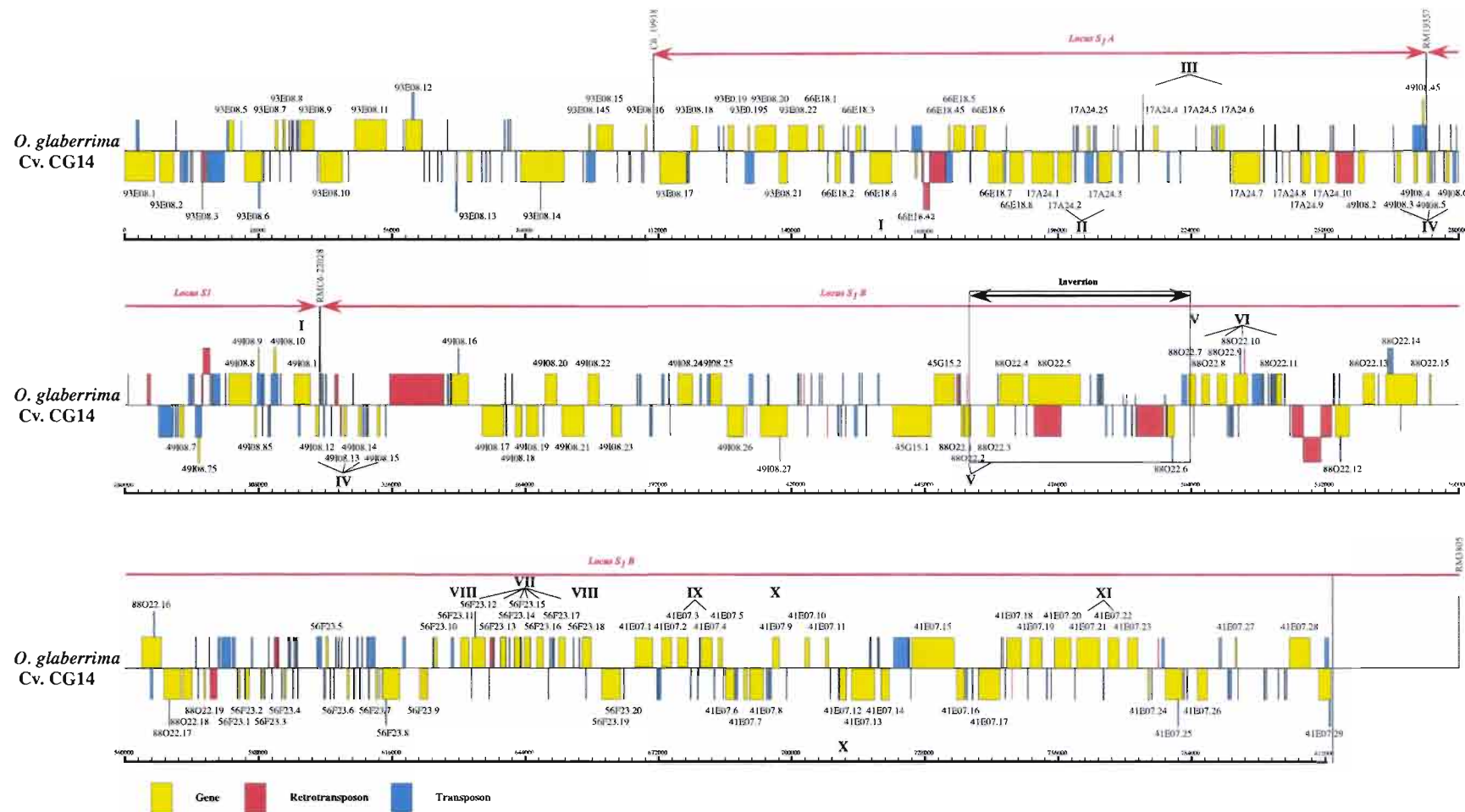


Figure IV-2. Physical map and annotation of the 813kb region of the S_1 , S_1A and S_1B loci in *O. glaberrima* cv. CG14. Yellow, blue and red boxes represent genes, transposons and retrotransposons respectively. TEs nested into others TEs or genes are raised above their insertion sites. Genes are labeled as in the Supplemental Table IV 2. Markers used in the genetic map are indicated in gray. A black arrow indicates a large sequence inversion, relative to *O. sativa* ssp. *japonica* cv. Nipponbare. Regions spanning the S_1 , S_1A and S_1B loci are indicated. Roman numerals indicate duplicated gene families listed in Table IV 1.

RESULTS

Sequencing of the *S_I*, *S_IA* and *S_IB* loci in *O. glaberrima*

In a previous work we described the fine genetic and physical mapping, sequencing, and comparative analysis of the female component of the *S_I* locus (Garavito *et al.*, 2009). Additionally we detected the presence of at least two other loci that interact epistatically with *S_I* to cause the allelic elimination of the gametes produced by the F₁ hybrids. In order to study these two additional loci we sequenced the seven remaining clones from the *O. glaberrima* cv. CG14 BAC library (Kim *et al.*, 2008) that constitute the minimum tiling path (MTP) established around *S_I*. The 8 sequenced clones (including the one from our previous work) account altogether for 1,102 kb of sequences that, once assembled, constitute an 813 kb contig, called here the *S_I* regions. The newly obtained sequences are available with the following EMBL accession numbers: FP340539 (OG-BBa0041E07); FP340540 (OG-BBa0056F23); FP340541 (OG-BBa0088O22); FP340542 (OG-BBa0045G15); FP340544 (OG-BBa0017A24); FP340545 (OG-BBa0066E18); and FP340546 (OG-BBa0093E08).

Limiting the localization of the *S_IA* and *S_IB* loci

In order to find the bounds of the segment that bears the *S_IA* and *S_IB* loci, we designed a set of new polymorphic markers, based on the sequence comparison of the orthologous regions between Nipponbare and CG14. Marker evaluation was carried out in the plants that limit the *S_IA* and *S_IB* in order to find the approximated recombination sites. Using this approach the *S_IA* locus was reduced to 183 kb, while no reduction was obtained for *S_IB* locus, remaining as a 654 kb sequence (Figure IV-1).

Sequence annotation and organization of the 813 kb of the CG14 *S_I* regions

The sequences of the seven CG14 BAC clones were annotated in detail. Gene and transposable element (TE) annotations are indicated in Figure IV-2. In total, 143 non-TE related coding regions were annotated, giving a gene density of about one gene per 6 kb of genomic sequence (Supplemental Table IV-2). Most of predicted genes were confirmed by identification of protein domains, BLASTX homologies in Swiss-Prot database or BLASTN homologies with nucleotide databases. On the 143 predicted genes, 96.3% showed the highest BLASTN homology with *O. sativa* genomic and full-length nucleotide sequences. Seven pseudogenes were identified by the presence of fragmented coding regions lacking start codons, or by the presence of stop codons in the frame of exons. Two of them involved genes

Table IV-1. List of identified gene families in the *O. glaberrima* S_1 regions

Duplicated gene families	Gene Name	Putative <i>O. sativa</i> (Nipponbare) orthologous gene	Putative Function	Positions in S_1 region (bp)	Orientations
I	OG-BBa0066E18.4	LOC_Os06g04690	Putative F-box protein	156499-161044	-
	OG-BBa0049I08.11	LOC_Os06g04980	Putative F-box protein	315435-318859	+
II	OG-BBa0017A24.2	LOC_Os06g04830	Putative LRR protein	195957-198903	-
	OG-BBa0017A24.3	LOC_Os06g04840	Putative LRR protein	204593-207258	-
III	OG-BBa0017A24.4	LOC_Os06g04850	Putative protein	216110-217001	+
	OG-BBa0017A24.6	LOC_Os06g04870	Putative protein	229920-230963	+
IV	OG-BBa0049I08.3	LOC_Os06g04930	Putative ENOD93 protein	267102-267855	-
	OG-BBa0049I08.5	LOC_Os06g04940	Putative ENOD93 protein	273413-274012	-
	OG-BBa0049I08.6	LOC_Os06g04950	Putative ENOD93 protein	277419-277981	-
	OG-BBa0049I08.12	LOC_Os06g04990	Putative ENOD93 protein	320002-320651	-
	OG-BBa0049I08.13	LOC_Os06g05000	Putative ENOD93 protein	325881-326408	-
	OG-BBa0049I08.14	LOC_Os06g05010	Putative ENOD93 protein	328903-329477	-
	OG-BBa0049I08.15	LOC_Os06g05020	Putative ENOD93 protein	332839-333419	-
V	OG-BBa0088O22.2	LOC_Os06g05209	Putative pectate lyase protein	456489-457642	-
	OG-BBa0088O22.3	LOC_Os06g05260	Putative pectate lyase protein	461242-462685	-
	OG-BBa0088O22.7	LOC_Os06g05272	Putative pectate lyase protein	503408-504867	+
VI	OG-BBa0088O22.8	LOC_Os06g05284	Putative Transferase	506199-507791	+
	OG-BBa0088O22.9	LOC_Os06g05300	Putative Transferase	509461-511398	+
	OG-BBa0088O22.10	LOC_Os06g05310	Putative Transferase	512912-515793	+
	OG-BBa0088O22.11	LOC_Os06g05320	Putative Transferase	521812-522540	+
VII	OG-BBa0056F23.12	LOC_Os06g05560	Putative protein	632773-635500	+
	OG-BBa0056F23.13	LOC_Os06g05580	Putative F-box protein	638676-639857	+
	OG-BBa0056F23.14*	LOC_Os06g05590	Putative F-box protein	641606-642834	+
	OG-BBa0056F23.15	LOC_Os06g05600	Putative F-box protein	643777-644970	+
	OG-BBa0056F23.16	LOC_Os06g05610	Putative F-box protein	646377-647636	+
	OG-BBa0056F23.17*	LOC_Os06g05620	Putative F-box protein	650971-652272	+
VIII	OG-BBa0056F23.11	LOC_Os06g05550	GDSL esterase/lipase protein	630402-632063	+
	OG-BBa0056F23.18	LOC_Os06g05630	GDSL esterase/lipase protein	655875-657747	+
IX	OG-BBa0041E07.3	LOC_Os06g05690	putative Cystein synthase protein	676063-678199	+
	OG-BBa0041E07.4	LOC_Os06g05700	putative Cystein synthase protein	681007-683334	+
X	OG-BBa0041E07.9	LOC_Os06g05750	Putative transferase protein	695968-697392	+
	OG-BBa0041E07.12	LOC_Os06g05790	Putative transferase protein	710049-711485	-
XI	OG-BBa0041E07.21	LOC_Os06g05900	Putative Methylase protein	759933-764709	+
	OG-BBa0041E07.22	LOC_Os06g05910	Putative methyltransferases	766462-768723	+

* Pseudogene

showing similarities with putative F-box protein families (56F23.14 and 56F23.17; Supplemental Table IV-2). Of the 143 predicted genes, 35 (25%) belong to eleven distinct duplicated gene families present along the 813 kb that contain *S_I*, coding for F-box (gene families I and VII), LRR proteins (II), putative homeobox (III), Early nodulin-like (IV), Pectate lyase (V), Transferase (VI and X), Esterase/Lipase (VIII), Cystein synthase (IX) and Methylase proteins (XI) (Table IV-1 and Figure IV-2). Copy numbers of duplicated genes vary from two to seven genes (Early nodulin-like family). Most of the duplicated gene families were organized in clusters of relatively adjacent duplicated genes, with the exception of the two genes of the family I (F-box) separated by 154 kb of genomic sequence. The majority of the duplicated gene families were found to be arrayed in tandem while only three families showed duplicated genes in opposite orientations (I, V and X, Table IV-1).

In total 380 known TE were identified representing 18.3% of the genomic sequence. Of the 380 elements, only 37 were classified as class I retroelements (Supplemental Table IV-3). Among all the LTR retrotransposons identified (19 elements), four were found complete (with two intact LTRs) and could be used to estimate their insertion time. Intact LTR retrotransposons showed an insertion time ranged from 0 to 0.89 ($\pm$ 0.269 MYA; Supplemental Table IV-4). Interestingly, ten annotated transposons overlap with predicted coding regions. These transposons were classified as pack-MULE since they showed similarities with Mutator-like elements and contained captured coding sequences (Jiang *et al.*, 2004). Among enclosed coding regions, six were classified as pseudogene due to the presence of frame-shift mutations or based on the absence of start codons. Only three pack-MULE displayed similarities to proteins with known functions (Supplemental Table IV-5).

Sequence comparisons of orthologous *S_I* regions between *O. glaberrima* and *O. sativa*

The orthologous of the CG14 *S_I* regions were identified in the *O. sativa* ssp. *japonica* cv. Nipponbare public sequence as a stretch of 847 kb between coordinates 1,900,806 and 2,748,457 on chromosome 6. The orthologous regions in *O. sativa* ssp. *indica* cv 93-11 sequence was also used for comparative analysis and consist of a stretch of 1077 kb broken by 152 ambiguous segments (with 'N'), that sometimes did not allow comparative analysis. Pairwise comparisons between CG14 and Nipponbare *S_I* regions revealed the presence of stretches of highly conserved segments interrupted by different zones with significant alterations. Most of the sequence variations involved mechanisms of sequence insertions, deletions, duplications, and a large sequence inversion (Figure IV-3). Sequence variations were not limited to intergenic regions since they overlap different segments that include

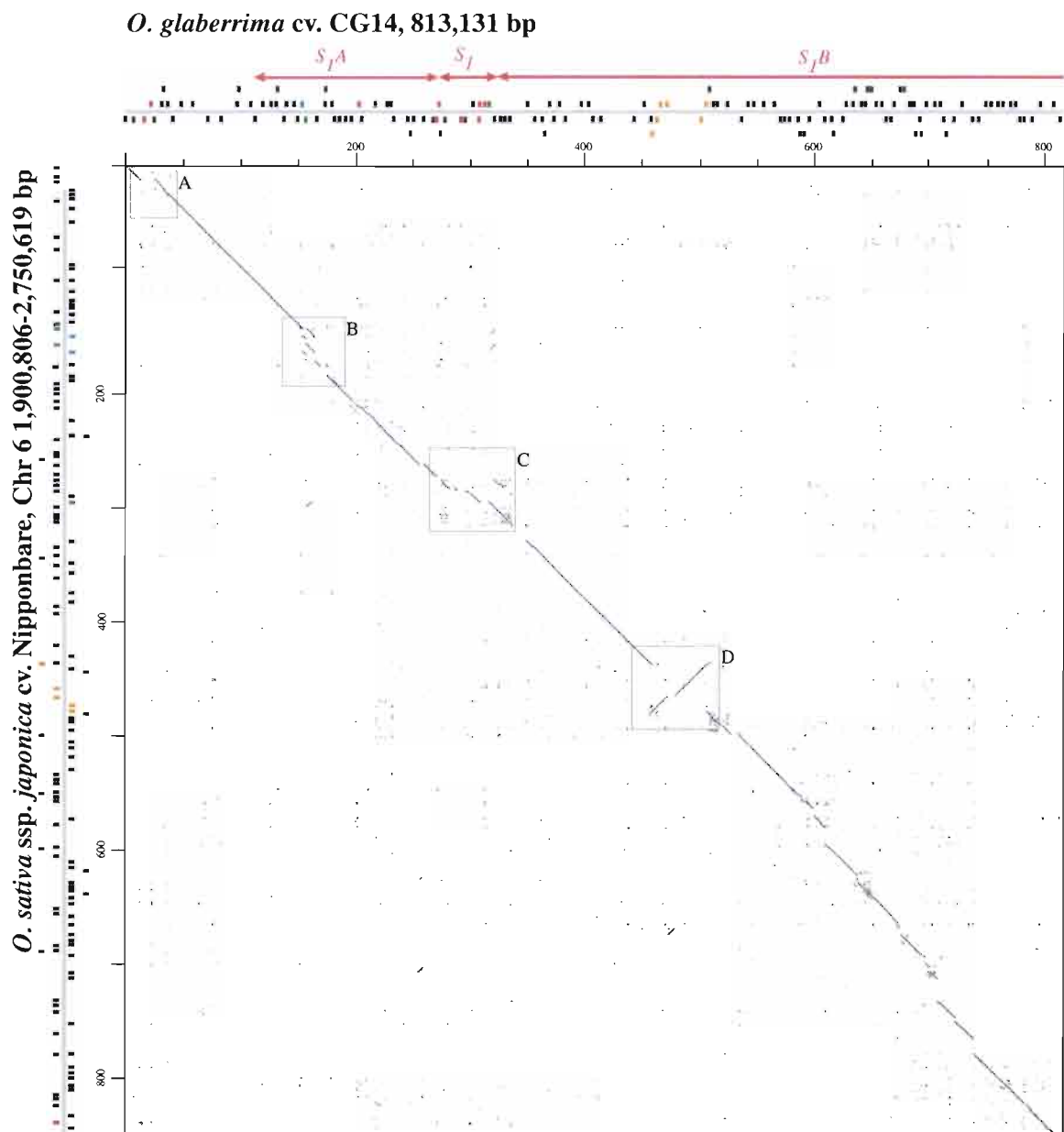


Figure IV-3. Orthologous sequence comparisons between the 813 kb and 847 kb from the *O. glaberrima* cv. GC14 and *O. sativa* ssp. *japonica* cv. Nipponbare S_1 regions. Comparisons were performed using dot plot alignment of the CG14 sequence (horizontal axis) against the Nipponbare sequence (vertical axis; coordinates 1,900,806-2,750,619-bp on chromosome 6). The rough positions of S_1A , S_1 and S_1B loci are indicated along the horizontal axis. Positions and orientation of genes are symbolized by black and colored boxes along X and Y-axes. Colored boxes represent genes that disrupted microcollinearity. Clear boxes in the dot blot underline four large regions showing a strong disruption in the microcollinearity.

coding genes (Figure IV-3). Detailed analysis of the collinearity was then performed between the non-TE genes located within the CG14 and Nipponbare *S_I* regions. All CG14 predicted genes were used as queries to BLAST them against Nipponbare genes within the *S_I* region to generate a distance matrix of genes. Relationships between orthologous and paralogous genes were then drawn using the computed distances to study microcollinearity (Supplemental Figure IV-1). Most of the genes (>90%) were found to be conserved in the same order and orientation between the two orthologous sequence at the *S_I* regions. Respectively 17 and 11 predicted genes were found involved in mechanisms that disrupt the microcollinearity between CG14 and Nipponbare. First, nine genes were present in CG14 and not in Nipponbare, while only two extra genes were found in the Nipponbare segment. Of the nine extra genes in CG14, five were enclosed into pack-MULE elements (17A24.25, 49I08.4, 49I08.45, 49I08.9 and 49I08.10, Supplemental Table IV-5). No putative protein function was found for these extra genes. In Nipponbare the two extra genes (Os06g04710 and Os06g05470) were classified as expressed proteins. This result suggests that most of the disruptions of microcollinearity in the *S_I* regions may be produced by gene movement mechanisms, such as the transposition of pack-MULE elements. These disruptions were mainly distributed into four large segments along the *S_I* regions (Figure IV-3). Detailed comparisons between CG14 and the two sub-species of *O. sativa* (ssp. *japonica* cv. Nipponbare and ssp. *indica* cv. 93-11) were carried out in these sites, in order to investigate the molecular mechanisms involved.

The comparison at the 93E18.3 and 93E18.5 CG14 genes region (Figure IV-3, box A), clearly showed a large insertion of 13,123 bp of sequence in CG14 relative to Nipponbare. This extra segment carries a pseudogene (93E18.3) and a gene coding for a putative protein (93E18.5), absent in *O. sativa* (Nipponbare and 93-11; data not shown). However, it was not possible to identify the mechanism that originated this insertion, since its extremities have no similarity with known TEs, and no traces of short duplications were clearly visible at the insertion site in CG14.

Comparisons between orthologous sequences around 66E18.3 (Putative protein) and 66E18.4 (Putative F-box protein) genes (Figure IV-3, box B) indicated that both CG14 and Nipponbare regions have undergone a large number of changes since the loci have diverged from a common ancestor. Here microcollinearity is perturbed by an order rearrangement involving 66E18.3/66E18.4 genes, compared to the positions of the orthologous genes from *O. sativa* (ssp. *japonica* and *indica*) (Figure IV-4). In addition to the inversion, a large number of TE insertions (two MITEs and two retrotransposons in CG14, and a large helitron and a

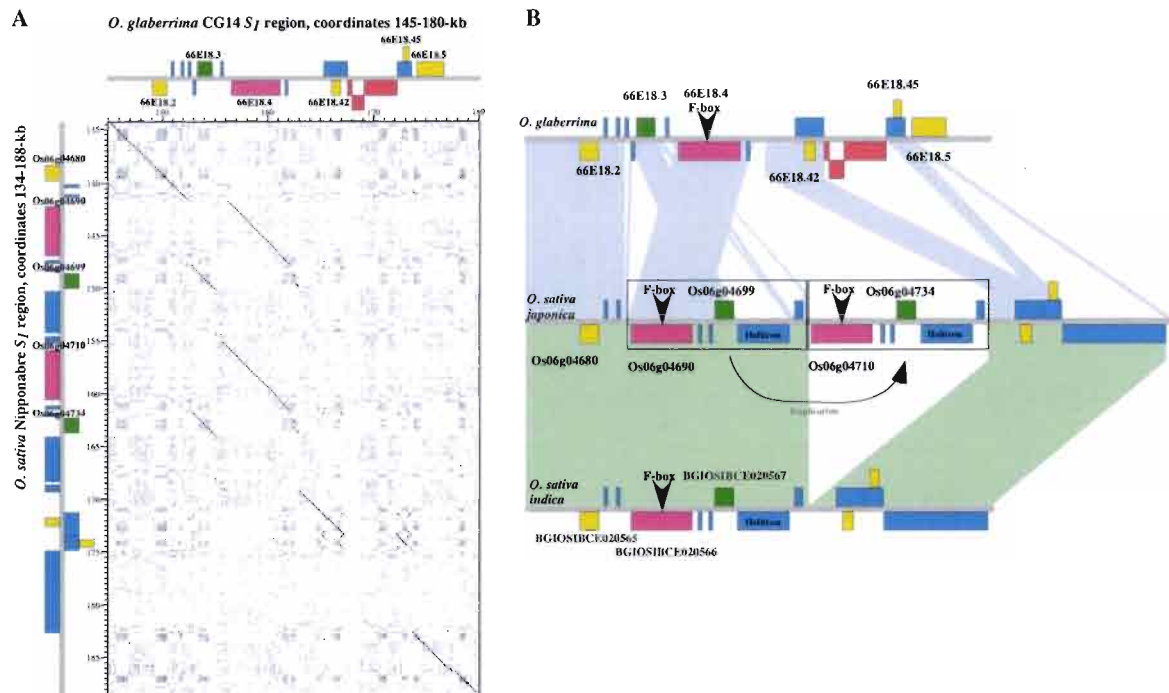


Figure IV-4. Comparisons of the orthologous sequences around the 66E18.3/66E18.4 CG14 genes. **A.** Dot plot comparison between the genomic region of genes 66E18.3 and 66E18.4 from CG14 (horizontal axis; coordinates 145-180 kb) against the *O. sativa* ssp. japonica cv. Nipponbare orthologous sequence (vertical axis; coordinates 2,034,806-2,088,806-bp on chromosome 6). **B.** Schematic representation of the comparison between the genomic region of genes 66E18.3 and 66E18.4 and their orthologous genes in *O. sativa* ssp. japonica cv. Nipponbare and *O. sativa* ssp. indica cv. 93-11. Colored backgrounds link orthologous regions. Boxes symbolize the positions of transposons and helitrons elements (light blue), Retrotransposons (red), 66E18.3 (Putative protein) gene and its orthologs (green), 66E18.4 (Putative F-box protein) genes and its orthologs (purple), and other genes (yellow).

transposon in Nipponbare and 93-11) were detected. Furthermore, in Nipponbare, a block of approximately 13 kb comprising genes Os06g04690, Os06g04690 and a Helitron has been duplicated in tandem orientation (Figure IV-5 B). This large duplication in Nipponbare is not conserved neither in CG14 nor 93-11. Unfortunately, the mechanisms at the origin of the gene inversion between 66E18.3 and 66E18.4 remain unknown, thus no evolutionary model could be developed. However it appears to be clear that after the divergence between *O. glaberrima* and *O. sativa*, numerous TE have been inserted up- and downstream the orthologous genes; and that after the divergence between *O. sativa* subspecies, a large duplication occurred in Nipponbare relative to 93-11. Altogether, these events indicate that this region may represent an intense spot of evolution between *O. glaberrima* and *O. sativa*.

At the 49I08.8 / 49I08.11 genes region (Figure IV-3, box C), localized within the *S₁* locus, the microcollinearity was perturbed by the presence of a total of six extra genes in *O. glaberrima* compared to *O. sativa* (Garavito *et al.*, 2009). Most of them (49I08.4; 49I08.45; 49I08.75; 49I08.9 and 49I08.10) appear to be enclosed within pack-MULE elements, suggesting that massive re-localization of pack-MULE elements in the *O. glaberrima* region may be here the mechanisms of the microcollinearity perturbation. (Supplemental Table IV-5).

Dot plot alignment of the CG14 and Nipponbare orthologous sequences around genes 88O22.2 / 88O22.7 showed a large inversion of more than 45 kb (Figure IV-3, box D). This inversion involved four different coding genes (88O22.3, 4, 5 and 88O22.6) in CG14 perturbs gene orders and orientations. A detailed comparison between orthologous sequences was carried out in order to identify breakpoints of the chromosomal inversion and to investigate the process responsible of such rearrangement. Close analysis indicated that the distal and proximal inversion breakpoints contain gene duplications in both species (respectively 88O22.2 / 88O22.7 genes and Os06g05209 / Os06g05272 genes). These genes, coding for Pectate lyase proteins, belong to a locally duplicated gene family composed of three gene copies (Family V, Table IV-1). Duplicated genes at the edge of the inversion are nearly identical with the exception of the first 36 extra-nucleotides at 5' end of 88O22.7 and Os06g05209 genes, resulting in twelve extra amino-acids for each gene (purple boxes and arrowheads, Figure IV-5 A). All genes at the inversion breakpoints in CG14 and Nipponbare appear intact and seem putatively functional (even after switching their upstream segment), since their coding regions are identical. A model for the inversion process is depicted in Figure IV-5 B. In the ancestral fragment (here the structure of the fragment is identical to

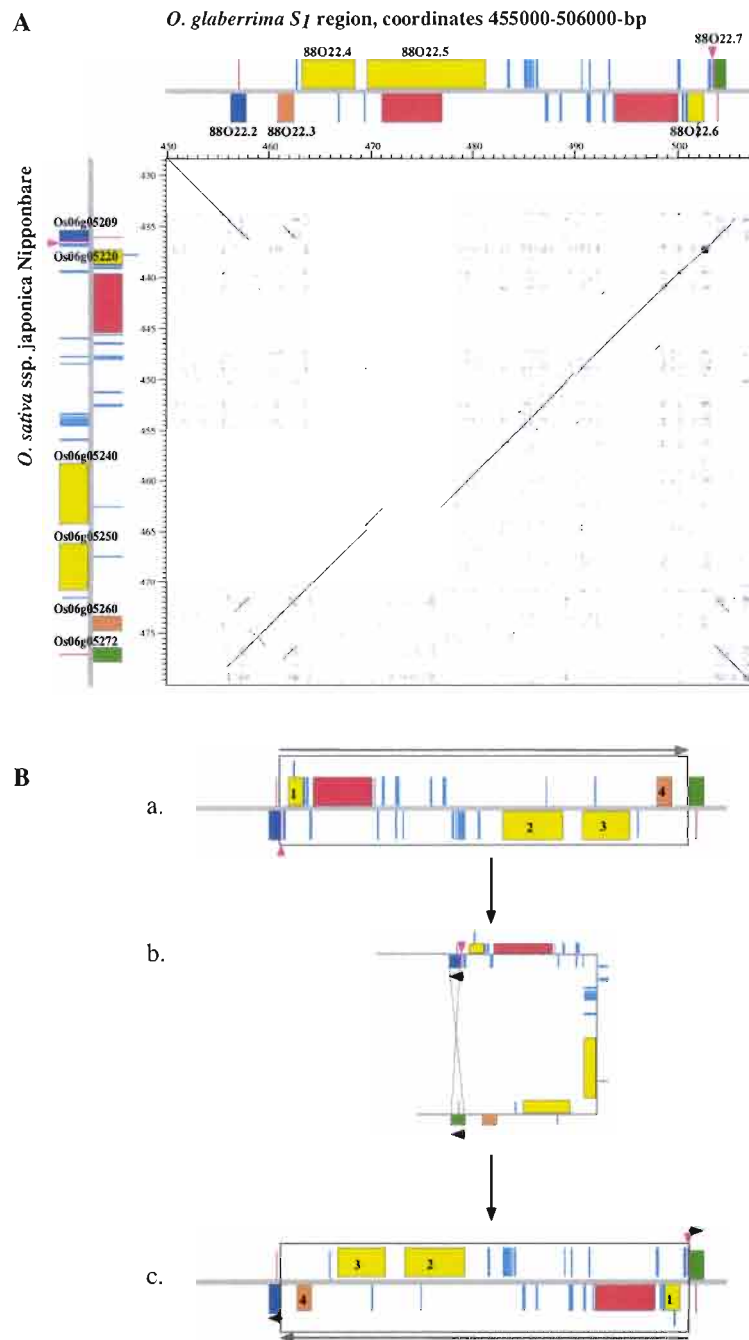


Figure IV-5. Comparisons of the orthologous sequences around the 88O22.2/88O22.7 CG14 genes. **A.** Dot plot comparison and structures between the genomic region of genes 88O22.3/88O22.7 from CG14 and their orthologous in Nipponbare. **B.** A model for the generation of the chromosomal inversion between *O. glaberrima* and *O. sativa*. **a.** Hypothetical ancestral segment (identical to the organization in the Nipponbare genome); **b.** Breakpoints occur by homologous recombination between two homologous genes coding for Pectate lyase proteins. **c.** Segment in *O. glaberrima*. Boxes symbolize the positions of Pectate lyase duplicated genes (blue, green and orange), transposons elements (light blue), Retrotransposons (red), and other genes (yellow). Purple boxes and arrowheads indicate the positions of twelve extra amino acids between duplicated Pectate lyase genes.

those of Nipponbare), two homologous Pectate lyase genes in opposite orientations flanked an internal region of 45 kb (Blue and green boxes, Figure IV-5 B). A mechanism of homologous recombination between inverted Pectate lyase genes occurred, leading to an inversion of the internal region and the exchange of upstream regions of Pectate lyase genes. Unfortunately the poor quality of the sequence assembly for 93-11 did not allow us to confirm the presence of the chromosomal inversion between *O. glaberrima* and *O. sativa ssp. indica*. Nevertheless, the comparisons between the genetic maps obtained from our *indica* and *japonica* populations (459 and 320 plants respectively) after reaching a high marker saturation around the site of the structural variation, showed no genetic recombination between inversion breakpoints for both *O. sativa* subspecies (Supplemental Figure III-2 and 3). As chromosomal inversions are known to suppress genetic recombination between normal and inverted chromosomal segments, the inversion seemed to be present also in *indica*.

Transposable elements participated to the dynamic evolution of S_I regions

Beside alteration of the order and orientation of genes through re-localization of pack-MULE elements, comparative analysis between Nipponbare and CG14 S_I regions reveals alterations of the S_I structure due to differential insertion or deletion of a multitude of transposable elements. In overall, more TE insertions occurred in the CG14 segment compared to Nipponbare (Supplemental Table IV-6). However, the size contribution of TE insertions is greater in Nipponbare compare to CG14. Since the divergence of the two species, more than 117 kb of TE (13% of the segment) was inserted in Nipponbare as against 72 kb (9%) in CG14. Size differences were mainly due to the insertion of long full-length LTR retrotransposons (Supplemental Table IV-6). The size difference between the S_I regions are in agreement with the genome size difference between *O. sativa* (425 Mb) and *O. glaberrima* (357 Mb) (Martinez *et al.*, 1994). Most of the TE appears randomly inserted along the S_I regions. One of the most striking exception of random insertion is a TE accumulation that occurred in the S_I locus (Figure IV-3). Here, the CG14 segment undergone a 1.5 X sequence size increase due to the local accumulation of TE in the neighborhood of the S_I candidate gene (49I08.11) (Garavito *et al.*, 2009). Beside the S_I locus, successive insertion of TE occurred specifically in Nipponbare, downstream the chromosomal inversion.

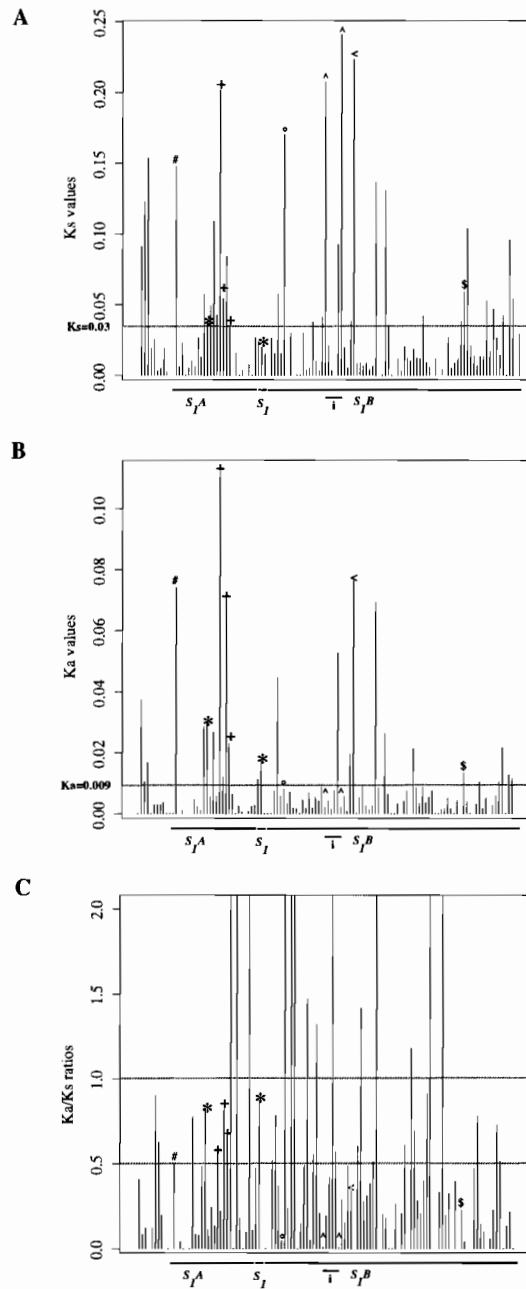


Figure IV-6. Representation of Ka and Ks values between 120 *O. sativa* and *O. glaberrima* orthologous genes in the S_I regions. **A.** Representation of Ks values. The horizontal line is the mean value of Ks ($K_s = 0.03$) for the 120 analyzed orthologous genes. **B.** Representation of Ka values. **C.** Representation of Ka/Ks values. The horizontal line is the mean value of Ka ($K_a = 0.09$) for the 120 analyzed orthologous genes. Lines below graphics represent analyzed genes located into S_I^A , S_I , S_I^B loci, and those within the chromosomal inversion (i). Symbols represent different genes or gene families as follows in *O. glaberrima*: (#) 93E08.17 (Putative protein), (*) 66E18.4 and 49I08.11 (Fbox proteins), (+) 66E18.8, 17A24.2 and 17A24.3 (LRR proteins), (°) 49I08.18 (Putative Serine/threonine protein kinase), (^) 88O22.2 and 88O22.7 (Pectate lyase gene located at breakpoint inversion), (<) 88O22.11 (Putative transferase), and (\$) 41E07.10 (Putative ubiquitin protein).

Gene divergences in S_I loci between *O. sativa* and *O. glaberrima*

Of the 143 annotated genes in the CG14 S_I regions, 120 were used for pairwise comparisons and divergence analysis with their respective Nipponbare orthologous genes. Twenty-three CG14 genes were not analyzed due to the absence of orthologous genes, deep gene structure differences between orthologous genes, or because one of the two orthologous genes was annotated as a pseudogene. Similar analysis were carried out as a control, using two other published loci in CG14: the ADH (Ammiraju *et al.*, 2008) and MOC1 (Lu *et al.*, 2009) loci. The mean level of rates of non-synonymous substitutions (K_a) and synonymous substitutions (K_s) are respectively 0.009 and 0.030 across the S_I regions (Supplemental Table IV-7). Similar mean level of K_a and K_s were obtained from ADH and MOC1 loci (respectively $K_a=0.007$, $K_s=0.040$ and $K_a=0.004$, $K_s=0.020$, data not shown). These data suggest that S_I loci as well as the ADH and MOC1 loci are globally under identical evolution rate acting on protein-coding genes.

Figure IV-6 shows K_s and K_a values for each of the 120 studied genes covering the 813 kb of the CG14 S_I loci (Supplemental Table IV-7). A glance of Figure IV-6 A shows a disparity of K_s divergence with several high peak values, significantly higher than both background and the mean value, suggesting that these genes correspond to regions of high nucleotide divergence in the S_I regions. We investigated whether high gene divergence may be correlated to structural variations such as transposable element abundance, duplications and chromosomal inversions. No clear correlation was found between highly divergent genes and transposable element abundance as illustrated by the S_I locus, where a clear accumulation of TE was observed in CG14 compared to Nipponbare (Garavito *et al.*, 2009), with no significant effect on gene divergence. In the region of a segmental duplication involving F-box genes in Nipponbare relative to CG14 (Figure IV-5), K_s clearly increase compared to background values (Figure IV-6 A). These genes are clustered into a block of nine genes, including three putative resistance LRR genes. Within the chromosomal inversion, no peak of high divergence was associated to genes at the exception of the hypothetical gene 88O22.6. However, duplicated Pectate lyase genes located at the inversion breakpoints show very high K_s values.

In addition to K_s , K_a values (Figure IV-6 B) showed disparities with several high peaks along S_I regions, indicating high level of non-synonymous substitutions. A clear correlation between high K_a and high K_s values was established in the region of a segmental duplication involving the F-box genes (located in the $S_I A$ locus). Here, six genes showing high K_s values

also showed high Ka values, suggesting a local, accelerated evolution driven by ecological adaptations (LRR genes).

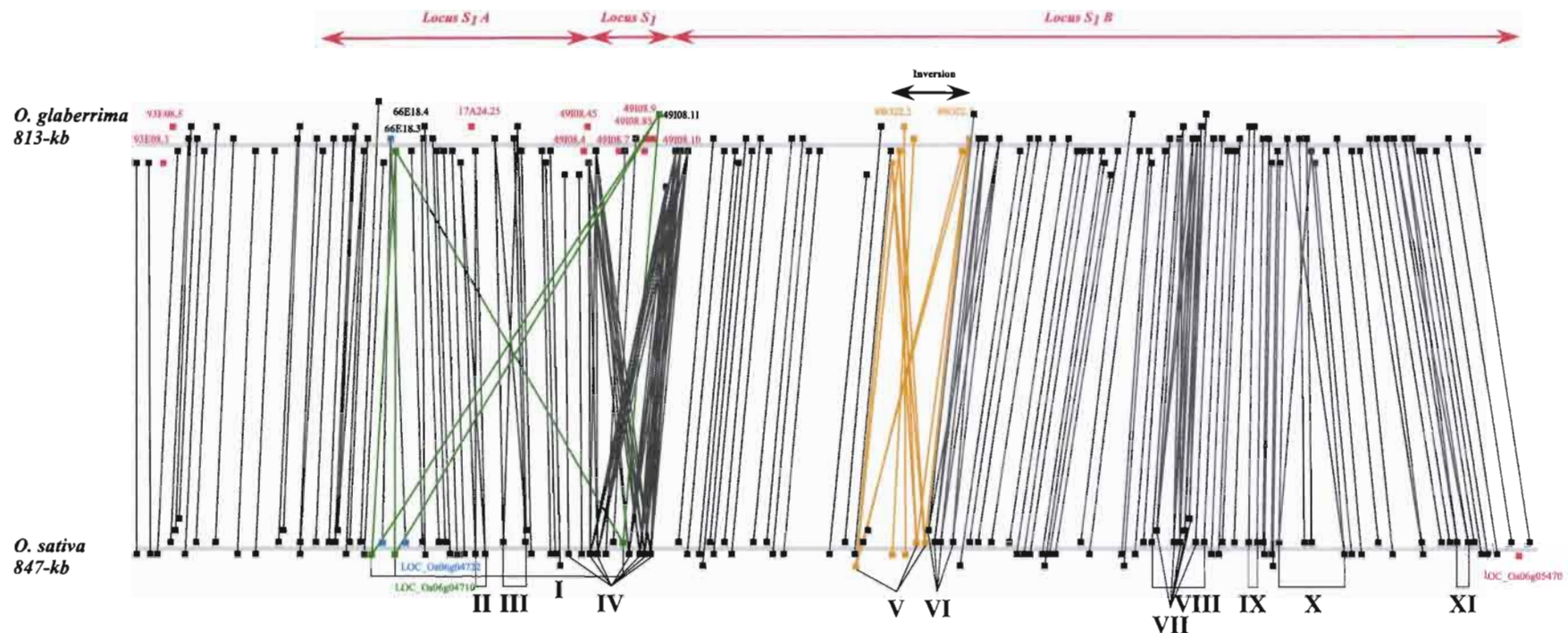
The Ka/Ks ratio was calculated to characterize the evolution of protein-coding sequences in the *S_I* loci (Figure IV-6 C and Supplemental Table IV-7). More than 80 % of the genes were shown to undergo under strong purifying selection (Supplemental Figure IV-4), seven genes to have a neutral evolution, and those whose Ks values are low to present the highest values.

Pattern of divergence of *S_IA* and *S_I* candidate gene regions

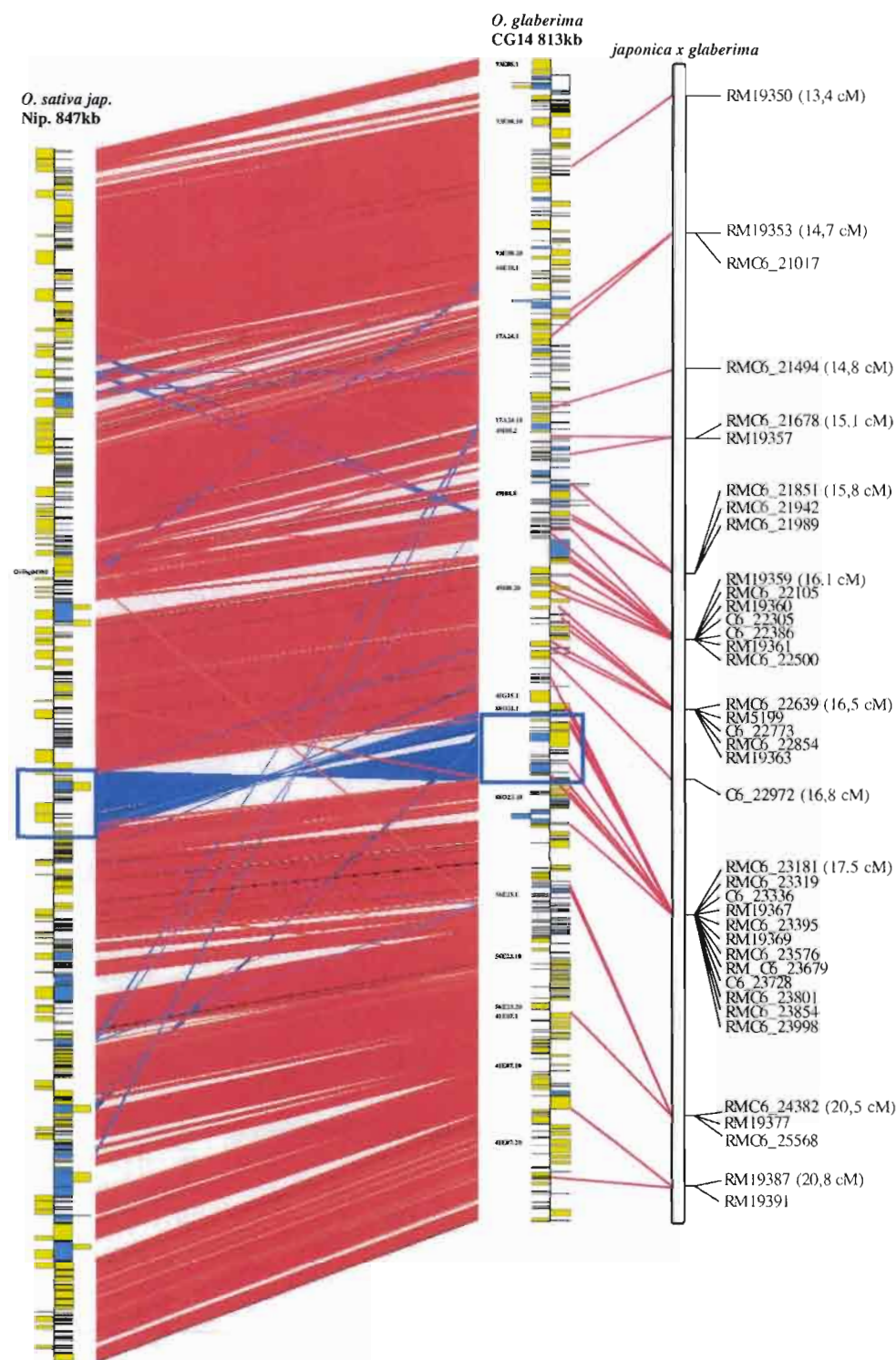
The gene 49I08.11 coding for a putative F-box-LRR protein has been proposed as a possible candidate gene for the *S_I* locus on the basis of the protein function, gene expression and evolution (Garavito *et al.*, 2009). Sequencing of the 813 kb of the *S_I* regions in *O. glaberrima* revealed a duplicated copy of the 49I08.11 F-box gene (66E18.4) 154 kb apart, in the *S_IA* locus (Figure IV-2 and Supplemental Table IV-2). The two duplicated F-box genes exhibited high overall sequence similarities (90.7% of nucleotide identities). Both *O. glaberrima* F-box genes were found to be conserved in orthologous *O. sativa* region. Altogether these data suggest the duplicated F-box 66E18.4 gene as a valuable candidate gene for *S_IA*.

To study evolution of the F-box gene family, detailed gene comparisons were performed at the *S_IA* and *S_I* loci. In the *S_I* locus, nucleotide and amino acid alignments show significant sequence variations between orthologous F-box genes in *O. glaberrima* and *O. sativa* (Supplemental Figure IV-4, 6 and 7). Ka/Ks rates, calculated for the *S_I* orthologous F-box genes, suggest a positive selection or relaxed selective constraint (Ka/Ks=0.87) (Garavito *et al.*, 2009), while the majority of genes in the *S_I* regions exhibits a purifying evolution (Supplemental Table IV-7). Besides coding region evolution, comparisons clearly established a different TE composition between CG14 and Nipponbare. The CG14 49I08.11 F-box gene appears to be embedded in an accumulation of TE that reshaped the upstream and downstream regions of the gene. Furthermore The 49I08.11 gene structure evolved through the insertion of a non-autonomous transposon nested into the fourth intron of the gene (Garavito *et al.*, 2009). In *S_IA* locus, the unique CG14 66E18.4 F-box gene is conserved with the two tandemly-duplicated genes in *O. sativa* (Os06g0469 and Os06g04710). While the structure of the Os06g04690 gene appears conserved with 66E18.4, the 5' part of the Os06g04710 gene exhibits several frame-shifts, resulting in a shorter predicted protein, lacking the F-box domain at the N-terminal region. Beside this variation, multiple alignments between *S_IA* F-box show numerous polymorphisms at the nucleotide and amino-acid levels

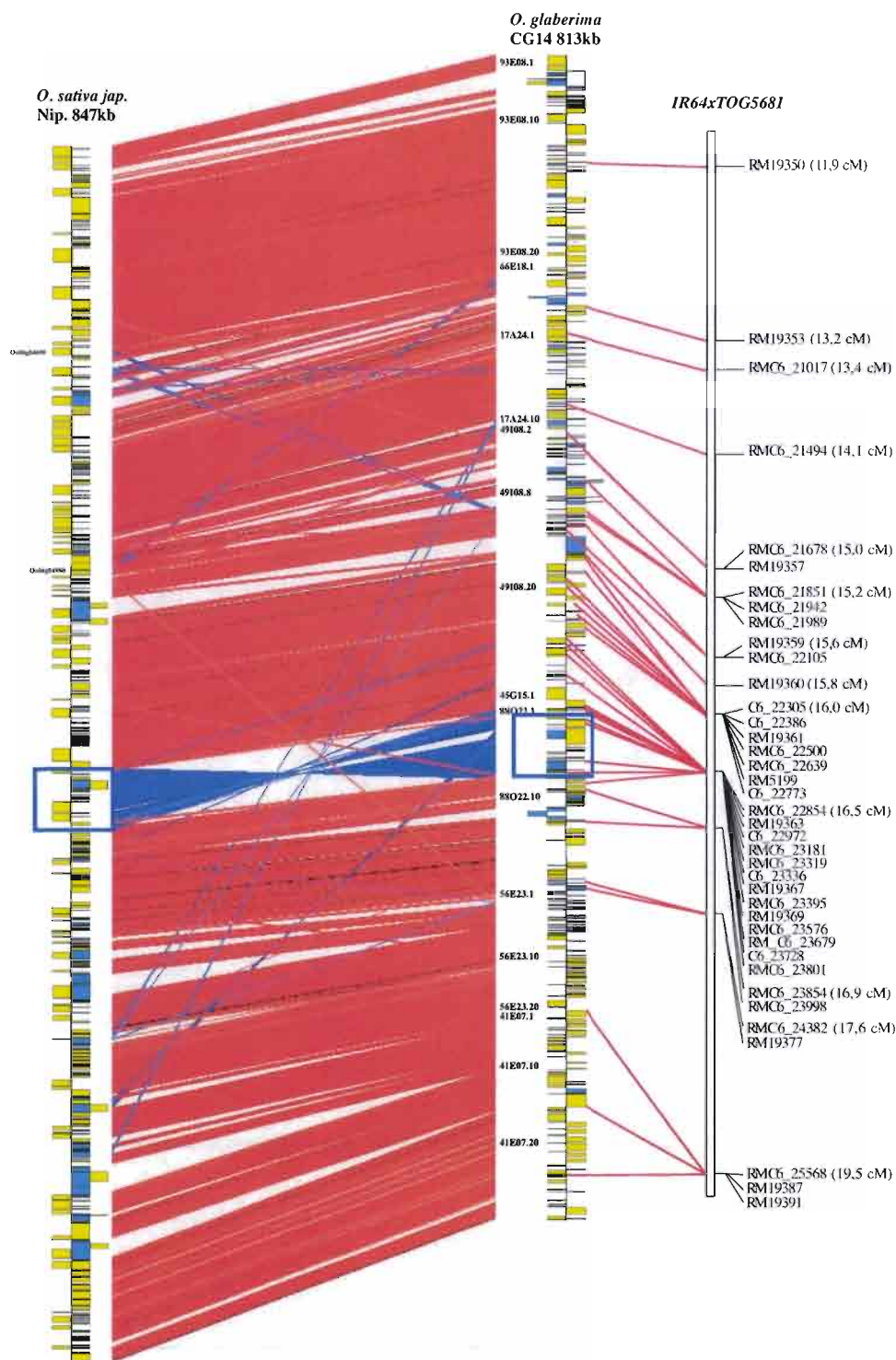
(Supplemental Figure IV-4 and 5), which shaped the corresponding phylogenetic tree (Supplemental Figure IV-6). Ka/Ks rates, calculated for these coding regions, suggest a similar evolution ($Ka/Ks=0.81$) to *S_l* F-box genes. In Nipponbare *O. sativa*, F-box gene evolved in the completely reshaped region compared to CG14 (Figure IV-4), through the impact of successive mechanisms such as gene movement, insertion of transposable elements and segmental duplication.



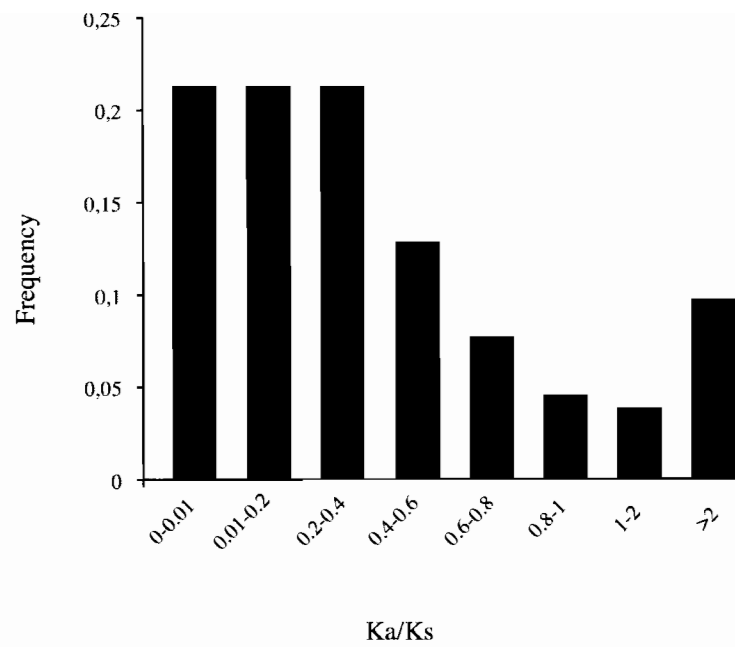
Supplemental Figure IV-1. Schematic representation of microcollinearity relationships between *O. glaberrima* cv. GC14 and *O. sativa* cv. Nipponbare S_1 regions. Colored boxes indicate positions and orientation of non-TE genes along axes representing the CG14 (upper segment) and Nipponbare (lower segment) S_1 regions. Colored lines linking boxes symbolize high identity relationships between one, or several genes from CG14 and Nipponbare. Red boxes indicate genes lacking orthologs. Blue and green boxes represent the positions of duplicated genes in Nipponbare compared to CG14. Orange boxes indicate the positions of inverted genes contained in the inversion. The positions of S_1A , S_1 and S_1B loci are indicated along the horizontal axis of CG14. Identified gene families in CG14 as classified in **Table IV-1** are indicated below the diagram.



Supplemental Figure IV-2. Comparison between structural variations identified between orthologous S_1 regions in CG 14 and Nipponbare, and the genetic map of the S_1 locus between *japonica* and *O. glaberrima*. Structural variations between orthologous S_1 regions (left) compared with the genetic map obtained from the *O. sativa* ssp. *japonica* x *O. glaberrima* crosses as previously described (Garavito *et al.*, 2009). Marker positions in the CG14 physical map are shown. Blue frames on the physical maps of *O. glaberrima* cv. CG 14 and *O. sativa* ssp *japonica* cv. Nipponbare indicate the position of the chromosomal inversion.



Supplemental Figure IV-3. Comparison between structural variations identified between orthologous S_1 regions in CG 14 and Nipponbare, and the genetic map of the S_1 locus between *indica* and *O. glaberrima*. Structural variations between orthologous S_1 regions (left) compared with the genetic map obtained from the *O. sativa* ssp. *indica* cv. IR64 x *O. glaberrima* cv. TOG5681 cross as previously described (Garavito *et al.*, 2009). Marker positions in the CG14 physical map are shown. Blue frames on the physical maps of *O. glaberrima* cv. CG 14 and *O. sativa* ssp. *japonica* cv. Nipponbare indicate the position of the chromosomal inversion.

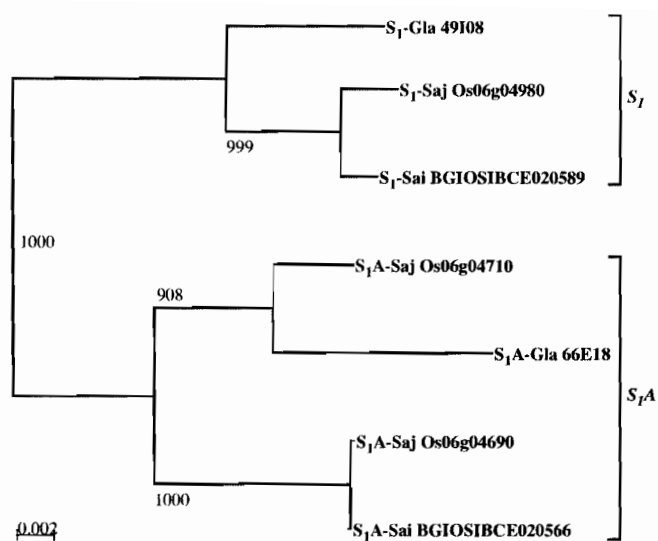


Supplemental Figure IV-4. Frequencies of the Ka/Ks values in the S_I regions.

S1A-Gla 66E18	GTACCTTGCTACTGCAACGGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGTGAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1200
S1A-Saj Os06g04710	GTACCTTGCTACTGCAACGGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGTGAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1201
S1A-Sai BGIOSIBCE020566	GTACCTTGCTACTGCAAGCGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGCTGAAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1202
S1A-Saj Os06g04690	GTACCTTGCTACTGCAAGCGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGCTGAAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1203
S1-Sai BGIOSIBCE020589	GTACCTTGCTACTGCAAGCGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGCTGAAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1204
S1-Saj Os06g04980	GTACCTTGCTACTGCAAGCGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGCTGAAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1205
S1-Gla 49108	GTACCTTGCTACTGCAAGCGCTCATCTCCCTGAAGGTGTTGGGTGTGACAAGCTGAAACCTGATAGAAATGAAGCTCCTAATGTCTCCATTTTGCATTTCAAGGGGACAAACCGAACTGAAACTTG	1206
S1A-Gla 66E18	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1207
S1A-Saj Os06g04710	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1208
S1A-Sai BGIOSIBCE020566	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1209
S1A-Saj Os06g04690	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1210
S1-Sai BGIOSIBCE020589	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1211
S1-Saj Os06g04980	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1212
S1-Gla 49108	GAGAACTTTGCAAAATAAGAGCCTATGCATGGTCCGCTCTGGTATGTTTATCATGCTGCTGCTGAACCTGCATCCATCATGCCAAATCTTGAATCTCTTGCCTAACAGTATGCAAGAGAGCGGCTTG	1213
S1A-Gla 66E18	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1214
S1A-Saj Os06g04710	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1215
S1A-Sai BGIOSIBCE020566	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1216
S1A-Saj Os06g04690	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1217
S1-Sai BGIOSIBCE020589	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1218
S1-Saj Os06g04980	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1219
S1-Gla 49108	TGCACCGAAGCTGTGTAGCAAAATCCTCTGCTCAGGCACCTGAGCATTGCGCTATTGGATTTTCCGAGCCTATGACTATTGTCCCTGGCTTCTTATGTTTATGCTGCACCTCTCTGGAGACTTTT	1220
S1A-Gla 66E18	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1221
S1A-Saj Os06g04710	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1222
S1A-Sai BGIOSIBCE020566	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1223
S1A-Saj Os06g04690	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1224
S1-Sai BGIOSIBCE020589	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1225
S1-Saj Os06g04980	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1226
S1-Gla 49108	GACTTGAATGTAATGCAAGCGAAGCTGCAGAAATGTTCAATTTTGCACATCCTGCAGATCTGAGATCAATACGAGAAGAGCAGCATCACAGCCTTAAAGGTGTGACGGTACATCATCTCTGTAA	1227
S1A-Gla 66E18	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1228
S1A-Saj Os06g04710	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1229
S1A-Sai BGIOSIBCE020566	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1230
S1A-Saj Os06g04690	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1231
S1-Sai BGIOSIBCE020589	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1232
S1-Saj Os06g04980	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1233
S1-Gla 49108	AGAGCTTGGTTGAGCTGACATGCCATATCCTTGAGAGTACAACTTCCCTTGAGTGCTTGACATTGGATGCTTCTCAGACTGGTTTAAAGTGTACTACTCCTGGCAGCAAAATCGGCAAAATGCCCTCCACT	1234
S1A-Gla 66E18	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1235
S1A-Saj Os06g04710	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1236
S1A-Sai BGIOSIBCE020566	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1237
S1A-Saj Os06g04690	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1238
S1-Sai BGIOSIBCE020589	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1239
S1-Saj Os06g04980	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1240
S1-Gla 49108	AGATAGGGACATTATCATGGAAGGTACAGAGGGGCTTGGCTATCAGAAGATACATCCAGCCTAGGGTACCTTCCACAGTGAAGCTGAAGTGTGGAAGCTGCAAGTGCATTCGACTGAAGCTTAG	1241

Supplemental Figure IV-5. Nucleotide sequence alignment of F-box genes in *S₁* and *S_{1A}* loci. Alignment between nucleotide sequences of F-box genes located in the *S₁* locus from *O. glaberrima* (S1-Gla 49108), *O. sativa* ssp. *japonica* (S1-Saj Os06g04980) and *O. sativa* ssp. *indica* (S1-Sai BGIOSIBCE020589) against the F-box genes in the *S_{1A}* locus from *O. glaberrima* (S1A-Gla 66E18), *O. sativa* ssp. *japonica*, (S1-Saj Os06g04710 and S1-Saj Os06g04690) and *O. sativa* ssp. *indica*, (S1-Sai BGIOSIBCE020566).

Supplemental Figure IV 5. Continuation.



Supplemental Figure IV-7. Phylogenetic tree of the F-box genes located in the $S_I A$ - S_I loci region. Tree constructed after the nucleotide alignment with ClustalX. Bootstrap numbers are shown.

Supplemental Table IV-1. New molecular markers located in the *S_I* regions.

Molecular marker	Position Chr. 6 Nipponbare	Primer Forward	Primer Reverse	Annealing (°C)	MgCl₂ (mM)
C6_19742	1 974 208	TTTCTCCTTTGGACGGAAC	TTTCTGCATGTTTGCCATTG	61	1,5
C6_19918	1 991 867	GGCAACAACAGAGGTTTGTG	TGACCTAAATGCCCTTCTTCA	61	1,5
C6_20017	2 001 715	ACTCGTGACCAACCCACCTA	GGGTCCCATCCTATCCAAG	61	1,5
RMC6_20107	2 010 723	TGATTAAGTGGTGATTGGTCA	AATGAGCATGATCACAGAGGA	59	1,5
RMC6_20199	2 019 980	AAATCAAGGTCGGCTCATTC	TCTCGCAGAGATCAAATTGC	59	1,5
RMC6_20349	2 034 961	TAACAGGGGGTGCTATGGAC	GGTCGAAAACGGTAGCACTT	59	1,5
C6_20417	2 041 714	TACTACCCGACCACCTTCCT	AAACTTGTCTTACGCTCTGTTT	61	1,5
C6_20707	2 070 780	TCTGCAACAAGTCTGCCACT	CCTTGATCCCAAAATTGGTT	61	1,5
C6_20928	2 092 855	ATTGGATCAACCACATGCTC	TCTGGTTGAGTTCACTGTTTAT	61	1,5
RMC6_21017	2 101 746	TGAACCTGCACTCATTGTAAACC	GAAGTGCGCATTACAAGCTC	59	1,5
RMC6_21494	2 149 463	CTGCTGGTGGACACATTTTG	TCATTCCAATGTCCAAACGA	59	1,5
C6_22972	2 297 236	GAGCTTGGCGTAGCTGATTC	TTCTTCCCATCTTCCATTG	61	2,0
RMC6_23181	2 318 116	CATCGTGTGCCCTTCAGTTA	ACTTTAAGCCCCATTGGGA	59	1,5
RMC6_23319	2 331 906	TGAATGCAGCATGTGTTCTT	TCATTCTTTGCCCATTTGTTT	59	1,5
C6_23336	2 333 657	GGCTGAATTGGGTAAGGTTG	CCCATTGGACTGCTTGTCT	61	1,5
RMC6_23395	2 339 577	CTGAGATTTTCTGCGTGCAA	GCGGTCATGATCAGGAAGAT	59	1,5
RMC6_23576	2 357 660	GTCCAAGCTCACAGCCTTCT	GCATGTTTGTGATGGGTGTTT	59	1,5
RMC6_23679	2 367 997	TCAGGAATGGGGCATAAAAC	TCCTTTTCAACTCGGTAACCA	59	1,5
RMC6_23801	2 380 193	CCGTGTGGCAGTTTATGAGG	GAACACGCGGAGGAACAC	59	1,5
RMC6_23854	2 385 488	TTTTCCATGGTTGTGGGTAA	GATGCCACGATCATGCTACA	59	1,5
RMC6_23998	2 399 814	TTTCAAACAACATAGTTTTCGTCA	TGGCTTCTCTCGTGGAGTTT	59	1,5
RMC6_24382	2 438 297	CCAATCCGAATAGACCCGTA	CCTCCAGGTAGTTCGGTCAA	59	1,5
C6_27332	2 733 214	GACAGTGTAAGGGACATTGG	CAGCCTAGAATTGGGTGGA	61	1,5
C6_27423	2 742 310	TCCGACGATATCGCTTACTC	GGGGCAGGGTTCACAATC	61	1,5

Marker names with the “RMC6” and “C6” prefix correspond to microsatellites and InDels, respectively.

Supplemental Table IV-2. List of identified genes in the 813-kb of the *O. glaberrima* var. CG14 *S*₁ region.

Locus	Gene Name	Putative Nipponbare orthologous gene	Putative Function	Protein domain	Best BLASTX homology (Swiss prot)	Best BLASTN homology
-	OG-BBa0093E08.1 ^a	LOC_Os06g04450	Putative Sec1-like protein	pfam00995, Sec1, Sec1 family	Q5VNU3 <i>O. sativa</i> Probable protein transport Sec1b (0.0)	AK100230 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.2	LOC_Os06g04460	Putative protein	cd03407, Band_7_4, A subgroup of the band 7 domain of flotillin (reggie) like proteins.	P16148 <i>Lupinus polyphyllus</i> Protein PPLZ12 (2e-69)	AK102678 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.3*	no orthologous gene	Putative protein	/	/	AK120604 <i>O. sativa</i> flcDNA (5e-41)
-	OG-BBa0093E08.5	no orthologous gene	Putative protein	pfam03478, Protein of unknown function (DUF295)	/	AP004234 <i>O. sativa</i> chr. 6 (0.0)
-	OG-BBa0093E08.6	LOC_Os06g04470	Putative protein	/	/	AK073180 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.7	LOC_Os06g04480	Putative protein	/	Q56XJ7 <i>A. thaliana</i> protein At4g22760 (2e-09)	AK109132 <i>O. sativa</i> flcDNA (3e-167)
-	OG-BBa0093E08.8	LOC_Os06g04490	Putative protein	/	/	AK288873 <i>O. sativa</i> flcDNA (1e-166)
-	OG-BBa0093E08.9	LOC_Os06g04500	Putative cornichon protein	pfam03311, Cornichon protein	P53173 <i>S. cerevisiae</i> ER-derived vesicles protein ERV14 (5e-20)	CT829434 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.10	LOC_Os06g04510	Putative Enolase	cd03313, Enolase; PRK00077, phosphopyruvate hydratase	P26301 <i>Z. mays</i> Enolase 1	AK069316 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.11	LOC_Os06g04520	Putative protein	/	Q3UMB9 <i>M. musculus</i> UPF0681 protein KIAA1033 (2e-37)	AK065081 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.12	LOC_Os06g04530	Putative protease	pfam00574, CLP_protease	Q9XJ36 <i>Arabidopsis</i> ATP-dependent Clp protease proteolytic subunit (6e-90)	AK104375 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.13	LOC_Os06g04540	Putative protein	cl00720, Domain of unknown function (DUF296)	Q9S7C9 <i>Arabidopsis</i> Putative DNA-binding protein ESCAROLA (4e-44)	AK107405 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.14	LOC_Os06g04560	Putative Armadillo protein	cd00106, Kinesin motor domain; cd00020, Armadillo/beta-catenin-like repeats	Q5VQ09 <i>O. sativa</i> Armadillo repeat-containing kinesin-like protein (0.0)	AK242714 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.145*	LOC_Os06g04530	Hypothetical protein	/	/	AP002071 <i>O. sativa</i> chr. 6 (5e-172)
-	OG-BBa0093E08.15	LOC_Os06g04580	Putative protein	cl02094, YEATS family	Q755P0 <i>Eremothecium gossypii</i> Protein AF-9 homolog (6e-31)	AK109053 <i>O. sativa</i> flcDNA (0.0)
-	OG-BBa0093E08.16	LOC_Os06g04590	Putative Auxin responsive protein	pfam02519, Auxin responsive protein	P33083 <i>G. max</i> Auxine induced protein 6b (1e-7)	AP003708 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa0093E08.17	LOC_Os06g04600	Putative protein	cl00935, Brix, Brix domain	Q5R631 <i>Pongo abelii</i> Ribosome production factor 1 (2e-73)	AK072896 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa0093E08.18	LOC_Os06g04610	Putative Ribosome Binding factor	cl00542, RBFA, Ribosome-binding factor A	Q65693 <i>A. thaliana</i> Probable ribosome-binding factor A (4e-56)	AK099752 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa0093E08.19	LOC_Os06g04620	Putative Epimerase	pfam01370, NAD dependent epimerase/dehydratase family	P93031 <i>A. thaliana</i> GDP-mannose 4,6 dehydratase 2 (4e-157)	AK073547 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa0093E08.195	LOC_Os06g04630	Putative protein	cl00321, AAT_I, Aspartate aminotransferase (AAT) superfamily	Q07051, <i>Aeromonas jandei</i> L-allo-threonine aldolase (e-7)	AP003708 <i>O. sativa</i> chr. 6 (1e-135)
<i>S₁A</i>	OG-BBa0093E08.20	LOC_Os06g04640	Putative transcription factor	cl02556, Bromodomain	Q9FT54 <i>A. thaliana</i> Transcription factor GTE6	AK250146 <i>H. vulgare</i> cDNA (0.0)
<i>S₁A</i>	OG-BBa0093E08.21	LOC_Os06g04650	Putative methionine sulfoxide reductase	cl00366, Peptide methionine sulfoxide reductase	P08761 <i>D. melanogaster</i> Peptide methionine sulfoxide reductase (7e-30)	CT832709 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa0093E08.22	LOC_Os06g04660	Putative protein	cl01206, 2OG-Fe(II) oxygenase superfamily	Q6P6C2 <i>H. sapiens</i> Alkylated DNA repair protein alkB homolog 5 (5e-12)	AK073705 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.1	LOC_Os06g04670	Putative CAF1 ribonuclease	cl09229, CAF1 family ribonuclease	Q74856 <i>S. pombe</i> Poly(A) ribonuclease pop2 (4e-27)	AP003708 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa66E18.2	LOC_Os06g04680	Putative protein	pfam03141, DUF248, Putative methyltransferase	/	AK069536 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.3	LOC_Os06g04699	Putative protein	/	/	AK105368 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.4	LOC_Os06g04690	Putative F-box protein	/	Q9SMT9 <i>A. thaliana</i> FBD-associated F-box protein (5e-4)	AK073027 <i>O. sativ</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.42*	LOC_Os06g04760	Hypothetical protein	/	/	AP003708 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa66E18.45	no orthologous gene	Hypothetical protein	/	/	AP003708 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa66E18.5	LOC_Os06g04780	Putative protein	pfam09747, Coiled-coil domain containing protein (DUF2052)	Q96F63 <i>H. sapiens</i> Coiled-coil domain-containing protein 9 (4e-16)	AK106220 <i>O. sativa</i> flcDNA (0.0)

Supplemental Table IV 2. Continuation

Locus	Gene Name	Putative Nipponbare orthologous gene	Putative Function	Protein domain	Best BLASTX homology (Swiss prot)	Best BLASTN homology
<i>S₁A</i>	OG-BBa66E18.6	LOC_Os06g04790	Putative HAD phosphatase	cl00375, Haloacid Dehalogenase-like Hydrolases	P27061 <i>S. lycopersicum</i> Acid phosphatase (6e-87)	AK058604 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.7	LOC_Os06g04800	Putative proteasome subunit	cd03762, proteasome beta type-6 subunit.	Q8LD27 <i>A. thaliana</i> Proteasome subunit beta type-6 (5e-97)	AK104152 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa66E18.8	LOC_Os06g04810	Putative LRR protein	pfam08263, Leucine rich repeat N-terminal domain	Q9LYN8 <i>A. thaliana</i> Leucine-rich repeat receptor protein kinase EXS (2e-56)	AP003708 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa17A24.1	LOC_Os06g04820	Putative methyltransferase	cl01306, Leucine carboxyl methyltransferase	Q6P4Z6 <i>Rattus norvegicus</i> Leucine carboxyl methyltransferase 1 (7e-43)	AK063371 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.2	LOC_Os06g04830	Putative LRR protein	COG4886, Leucine-rich repeat (LRR) protein	Q9FL28 <i>A. thaliana</i> LRR receptor-like serine/threonine-protein kinase (4e-72)	AK287701 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.25*	no orthologous gene	Hypothetical protein	/	/	AP004329 <i>O. sativa</i> chr. 6 (0.0)
<i>S₁A</i>	OG-BBa17A24.3	LOC_Os06g04840	Putative LRR protein	pfam08263, Leucine rich repeat N-terminal domain	Q9FL28 <i>A. thaliana</i> LRR receptor-like serine/threonine-protein kinase (2e-69)	AK100368 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.4	LOC_Os06g04850	Putative protein	cd00086, homeodomain	Q5VPE5 <i>O. sativa</i> Homeobox-leucine zipper protein HOX28 (3e-75)	AK059116 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.5	LOC_Os06g04860	Hypothetical protein	/	/	AP003708 <i>O. sativa</i> chr. 6 (2e-45)
<i>S₁A</i>	OG-BBa17A24.6	LOC_Os06g04870	Putative protein	cd00086, homeodomain,	Q5VPE3 <i>O. sativa</i> Homeobox-leucine zipper protein HOX2 (5e-96)	AK105150 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.7	LOC_Os06g04880	Putative serine/Threonine protein kinase	cd00180, Serine/Threonine protein kinases, catalytic domain; cd01989, The N-terminal domain of Eukaryotic Serine Threonine kinases	Q9LU47 <i>A. thaliana</i> Putative U-box domain-containing protein 53 (3e-164)	AK065683 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa17A24.8	LOC_Os06g04889	Putative protein	/	/	AK058823 <i>O. sativa</i> flcDNA (1e-99)
<i>S₁A</i>	OG-BBa17A24.9	LOC_Os06g04900	Putative protein	cl09119, Major Facilitator Superfamily (MFS)	Q07423 <i>Ricin</i> communis Hexose carrier protein HEX6 (9e-147)	NM_001063291 <i>O. sativa</i> Os06g0141000 (0.0)
<i>S₁A</i>	OG-BBa17A24.10	LOC_Os06g04910	Putative NADH pyrophosphatase protein	pfam09296, NUDIX-like; cd03429, NADH pyrophosphatase	Q94A82 <i>A. thaliana</i> NADH pyrophosphatase NUDT19 (8e-69)	NM_001063292 <i>O. sativa</i> Os06g0141100 (0.0)
<i>S₁A</i>	OG-BBa0049I08.2	LOC_Os06g04920	Putative Zinc finger protein	pfam00641, Zn-finger	O13801 <i>S. pombe</i> RNA-binding protein (2e-9)	CT830197 <i>O. sativa</i> flcDNA (0.0)
<i>S₁A</i>	OG-BBa0049I08.3	LOC_Os06g04930	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (2e-11)	AK121184 <i>O. sativa</i> flcDNA (2e-131)
<i>S₁A</i>	OG-BBa0049I08.4	no orthologous gene	Hypothetical protein	cd06163 zinc metalloproteases	P73714 <i>Synechocystis</i> (1e-9)	AC136219 <i>O. sativa</i> BAC (0.0)
<i>S₁A</i>	OG-BBa0049I08.45*	no orthologous gene	Hypothetical protein	/	/	AC109601 <i>O. sativa</i> BAC (8e-41)
<i>S₁</i>	OG-BBa0049I08.5	LOC_Os06g04940	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>Glycine max</i> Early nodulin 93 (8e-20)	AK122162 <i>O. sativa</i> flcDNA (2e-148)
<i>S₁</i>	OG-BBa0049I08.6	LOC_Os06g04950	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (4e-14)	AK121791 <i>O. sativa</i> flcDNA (1e-157)
<i>S₁</i>	OG-BBa0049I08.7	no orthologous gene	Hypothetical protein	/	/	/
<i>S₁</i>	OG-BBa0049I08.75	no orthologous gene	Hypothetical protein	/	/	AK107277 <i>O. sativa</i> flcDNA (1e-115)
<i>S₁</i>	OG-BBa0049I08.8	LOC_Os06g04970	Putative protein	/	Q9SH88 <i>A. thaliana</i> Ribosome biogenesis regulatory protein (3e-48)	NM_001063299 <i>O. sativa</i> Os06g0142000 (0.0)
<i>S₁</i>	OG-BBa0049I08.85*	no orthologous gene	Hypothetical protein	/	/	AP002838 <i>O. sativa</i> BAC (2e-156)
<i>S₁</i>	OG-BBa0049I08.9*	no orthologous gene	Hypothetical protein	/	/	AP004090 <i>O. sativa</i> BAC (8e-69)
<i>S₁</i>	OG-BBa0049I08.10*	no orthologous gene	Hypothetical protein	/	P47927 <i>A. thaliana</i> Floral homeotic protein APETALA 2 (4e-8)	AC121363 <i>O. sativa</i> BAC (3e-128)
<i>S₁</i>	OG-BBa0049I08.11	LOC_Os06g04980	Putative F-box protein	/	Q9FJT2 <i>A. thaliana</i> F-box/FBD/LRR-repeat (5e-4)	AK073027 <i>O. sativa</i> flcDNA (0.0)

Supplemental Table IV 2. Continuation

Locus	Gene Name	Putative Nipponbare orthologous gene	Putative Function	Protein domain	Best BLASTX homology (Swiss prot)	Best BLASTN homology
<i>S_i</i>	OG-BBa0049108.12	LOC_Os06g04990	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	AK122162 <i>O. sativa</i> flcDNA (2e-175)
<i>S_iB</i>	OG-BBa0049108.13	LOC_Os06g05000	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (1e-14)	CT837557 <i>O. sativa</i> flcDNA (4e-177)
<i>S_iB</i>	OG-BBa0049108.14	LOC_Os06g05010	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	CT827995 <i>O. sativa</i> flcDNA (1e-158)
<i>S_iB</i>	OG-BBa0049108.15	LOC_Os06g05020	Putative ENOD93 protein	pfam03386, Early nodulin 93 ENOD93 protein	Q02921 <i>G. max</i> Early nodulin 93 (1e-17)	CT827995 <i>O. sativa</i> flcDNA (9e-173)
<i>S_iB</i>	OG-BBa0049108.16	LOC_Os06g05050 ^C	Putative Serine/Threonine protein kinases	cd00180, Serine/Threonine protein kinases	Q9LMN7 <i>A. thaliana</i> Wall-associated receptor kinase 5 (5e-130)	NM_001063304 <i>O. sativa</i> Os06g0142500 (0.0)
<i>S_iB</i>	OG-BBa0049108.17	LOC_Os06g05060	Putative Early flowering protein	/	O82804 <i>A. thaliana</i> Protein EARLY FLOWERING 3 (1e-55)	NM_001063305 <i>O. sativa</i> Os06g0142600 (0.0)
<i>S_iB</i>	OG-BBa0049108.18	LOC_Os06g05070	Putative Serine/Threonine protein kinases	cd00180, Serine/Threonine protein kinases	O23081 <i>A. thaliana</i> Cysteine-rich receptor-like protein kinase 41 (2e-57)	AP000399 <i>O. sativa</i> BAC (0.0)
<i>S_iB</i>	OG-BBa0049108.19	LOC_Os06g05080	Putative cytochrome oxidase	cd00924, Cytochrome c oxidase subunit Vb	P00428 <i>Bos taurus</i> Cytochrome c oxidase subunit 5B, mitochondrial (6e-14)	CT834467 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0049108.20	LOC_Os06g05090	Putative methyltransferase protein	ed02440, S-adenosylmethionine-dependent methyltransferases	Q9SNQ2 <i>O. sativa</i> protein arginine N-methyltransferase (0.0)	AK103971 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0049108.21	LOC_Os06g05100	Putative 1-deoxy-D-xylulose-5-phosphate synthase	cd02007, Thiamine pyrophosphate (TPP) family, DXS subfamily, TPP-binding module	Q38854 <i>A. thaliana</i> 1-deoxy-D-xylulose-5-phosphate synthase, chloroplastic (0.0)	AK121920 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0049108.22	LOC_Os06g05110	Putative Superoxide dismutase	PRK10543, superoxide dismutase; pfam02777, Iron/manganese superoxide dismutases	P22302 <i>Nicotiana plumbaginifolia</i> Superoxide dismutase [Fe], chloroplastic (5e-60)	AK071301 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0049108.23	LOC_Os06g05120	Putative protein	/	/	AK061597 <i>O. sativa</i> flcDNA (9e-80)
<i>S_iB</i>	OG-BBa0049108.24	LOC_Os06g05130	Putative Myristoyl-acyl carrier protein thioesterase	ed00586, 4-hydroxybenzoyl-CoA thioesterase (4HBT)	Q9SQI3 <i>Gossypium hirsutum</i> Myristoyl-acyl carrier protein thioesterase, chloroplastic (6e-146)	AK120946 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0049108.25	LOC_Os06g05140	Putative PPR protein	pfam01535, PPR repeat	O49287 <i>A. thaliana</i> Putative pentatricopeptide repeat-containing protein (2e-93)	AP003487 <i>O. sativa</i> BAC (0.0)
<i>S_iB</i>	OG-BBa0049108.26	LOC_Os06g05150 ^C	Putative protein	pfam03371, PRP38 family	Q80SY5 <i>M. musculus</i> Pre-mRNA-splicing factor 38B (2e-40)	AP003487 <i>O. sativa</i> BAC (0.0)
<i>S_iB</i>	OG-BBa0049108.27	LOC_Os06g05160	Putative sulfate transporter	pfam00916, Sulfate transporter family	Q9LW86 <i>A. thaliana</i> Probable sulfate transporter (1e-141)	AK067270 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0045G15.1	LOC_Os06g05180	Putative Coatomer protein	pfam04053, Coatomer WD associated region; cd00200, WD40	Q5VQ78 <i>O. sativa</i> Coatomer subunit beta'-1	AK111584 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0045G15.2	LOC_Os06g05190	Putative BRCA1 protein	cl00038, Breast Cancer Suppressor Protein (BRCA1), carboxy-terminal domain.	Q60596 <i>M. musculus</i> DNA repair protein XRCC1 (2e-23)	AB292807 <i>O. sativa</i> xrec1 mRNA (0.0)
<i>S_iB</i>	OG-BBa0088O22.1	LOC_Os06g05200	Putative Zinc finger protein	cd00162, RING-finger	/	AP003487 <i>O. sativa</i> chr. 6 (0.0)
<i>S_iB</i>	OG-BBa0088O22.2	LOC_Os06g05209	Putative pectate lyase protein	cl01593, Pectate lyase	P40973 <i>Lilium longiflorum</i> Pectate lyase (1e-126)	AK100191 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0088O22.3	LOC_Os06g05260	Putative pectate lyase protein	cl01593, Pectate lyase	P40973 <i>Lilium longiflorum</i> Pectate lyase (1e-142)	AK070509 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0088O22.4	LOC_Os06g05250	Putative GTP-binding protein	PRK05433, GTP-binding protein LepA; cd03699, lepA_II; ed03709, lepA_C	Q8N442 <i>H. sapiens</i> GTP-binding protein GUF1 (0.0)	AK103630 <i>O. sativa</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0088O22.5	LOC_Os06g05240	Putative carboxypeptidase	cl10475, Zinc carboxypeptidase	Q9JHW1 <i>Ratus norvegicus</i> Carboxypeptidase D (5e-69)	BT038203 <i>Z. mays</i> flcDNA (0.0)
<i>S_iB</i>	OG-BBa0088O22.6	LOC_Os06g05220	Hypothetical protein	pfam03080, <i>Arabidopsis</i> proteins of unknown function	/	AP003487 <i>O. sativa</i> chr. 6 (6e-96)
<i>S_iB</i>	OG-BBa0088O22.7	LOC_Os06g05272	Putative pectate lyase protein	cl01593, Pectate lyase	P40973 <i>Lilium longiflorum</i> Pectate lyase (2e-134)	AK100191 <i>O. sativa</i> flcDNA (0.0)

Supplemental Table IV 2. Continuation

Locus	Gene Name	Putative Nipponbare orthologous gene	Putative Function	Protein domain	Best BLASTX homology (Swiss prot)	Best BLASTN homology
<i>S₇B</i>	OG-BBa0088O22.8	LOC_Os06g05284	Putative Transferase	pfam02458, Transferase family	Q9ZWR8 <i>Gentiana triflora</i> Anthocyanin 5-aromatic acyltransferase (2c-16)	AP003487 <i>O. sativa</i> chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0088O22.9	LOC_Os06g05300	Putative Transferase	pfam02458, Transferase	Q9ZWR8 <i>Gentiana triflora</i> Anthocyanin 5-aromatic acyltransferase (4c-14)	AP003487 <i>O. sativa</i> chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0088O22.10	LOC_Os06g05310	Putative Transferase	pfam02458, Transferase	Q9ZWR8 <i>Gentiana triflora</i> Anthocyanin 5-aromatic acyltransferase (2c-16)	AP003487 <i>O. sativa</i> chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0088O22.11	LOC_Os06g05320	Putative Transferase	pfam02458, Transferase	/	AK106249 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0088O22.12	LOC_Os06g05340	Putative AP2 proyein	cd00018, AP2	Q6XSY6 <i>A. thaliana</i> Ethylene-responsive transcription factor WR11 (3c-74)	AP003487 <i>O. sativa</i> chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0088O22.13	LOC_Os06g05350	Putative Whirly transcription factor protein	pfam08536, Whirly transcription factor	/	CT836194 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0088O22.14	LOC_Os06g05359	Putative protein	cl09099, P-loop containing Nucleoside Triphosphate Hydrolases	Q9STE7 <i>A. thaliana</i> Putative disease resistance RPP13-like protein (5c-75)	AK120452 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0088O22.15	LOC_Os06g05368	Putative protein	/	/	NM_001063327 <i>O. sativa</i> Os06g0146200 (1c-151)
<i>S₇B</i>	OG-BBa0088O22.16	LOC_Os06g05380	Putative WRKY protein	pfam03106, WRKY DNA -binding domain	Q9LXG8 <i>A. thaliana</i> Probable WRKY transcription factor 72 (4c-53)	BK005076 <i>O. sativa</i> WRKY transcription factor (0.0)
<i>S₇B</i>	OG-BBa0088O22.17	LOC_Os06g05390	Putative protein	pfam09353, Domain of unknown function (DUF1995)	/	AK101052 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0088O22.18	LOC_Os06g05400	Putative Iron-sulphur protein	cl00400, Iron-sulphur cluster biosynthesis	Q9XIK3 <i>A. thaliana</i> Iron-sulfur assembly protein IscA (5c-47)	CT832410 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0088O22.19	LOC_Os06g05410	Hypothetical protein	/	/	AP003767 <i>O. sativa</i> chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.1	LOC_Os06g05420	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (3c-173)
<i>S₇B</i>	OG-BBa0056F23.2	LOC_Os06g05424	Putative protein	/	/	AK071352 <i>O. sativa</i> flcDNA (4c-138)
<i>S₇B</i>	OG-BBa0056F23.3	LOC_Os06g05430	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (3c-169)
<i>S₇B</i>	OG-BBa0056F23.4	LOC_Os06g05440	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (3c-178)
<i>S₇B</i>	OG-BBa0056F23.5	LOC_Os06g05470	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.6	LOC_Os06g05480	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.7	LOC_Os06g05510	Putative protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.8	LOC_Os06g05520	Putative protein kinase	cd00180, Serine/Threonine protein kinases, catalytic domain	Q9S7U9 <i>A. thaliana</i> Mitogen-activated protein kinase kinase 2 (7c-124)	AK111598 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.9	LOC_Os06g05530	Putative protein	/	/	CT828147 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.10	LOC_Os06g05540	Hypothetical protein	/	/	AP003767 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.11	LOC_Os06g05550	GDSL esterase/lipase protein	cd01837, SGNH_plant_lipase_like	Q0WP19 <i>A. thaliana</i> GDSL esterase/lipase (2c-95)	CT831308 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.12	LOC_Os06g05560	Putative protein	pfam03478, Protein of unknown function (DUF295)	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.13	LOC_Os06g05580	Putative F-box protein	pfam03478, Protein of unknown function (DUF295)	O49315 <i>A. thaliana</i> Putative F-box protcin (2c-6)	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.14*	LOC_Os06g05590	Putative F-box protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.15	LOC_Os06g05600	Putative F-box protein	pfam03478, Protein of unknown function (DUF295); cl02535, F-box, F-box domain	/	AK287605 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.16	LOC_Os06g05610	Putative F-box protein	pfam03478, Protein of unknown function (DUF295); cl02535, F-box, F-box domain	/	AK107542 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.17*	LOC_Os06g05620	Putative F-box protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₇B</i>	OG-BBa0056F23.18	LOC_Os06g05630	GDSL esterase/lipase protein	cd01837, SGNH_plant_lipase_like	Q0WP19 <i>A. thaliana</i> GDSL esterase/lipase (2c-95)	AK099064 <i>O. sativa</i> flcDNA (0.0)
<i>S₇B</i>	OG-BBa0056F23.19	LOC_Os06g05640	Putative protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)

Supplemental Table IV 2. Continuation

Locus	Gene Name	Putative Nipponbare orthologous gene	Putative Function	Protein domain	Best BLASTX homology (Swiss prot)	Best BLASTN homology
<i>S₁B</i>	OG-BBa0056F23.20	LOC_Os06g05650	Putative protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (3e-93)
<i>S₁B</i>	OG-BBa0041E07.1	LOC_Os06g05660	Putative Nucleosome protein	pfam00956, Nucleosome assembly protein (NAP)	Q4U0Y4 <i>Xenopus laevis</i> Nucleosome assembly protein 1-like (6e-38)	AK071177 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.2	LOC_Os06g05670	Putative protein	/	/	AK102827 <i>O. sativa</i> flcDNA (3e-179)
<i>S₁B</i>	OG-BBa0041E07.3	LOC_Os06g05690	putative Cystein synthase protein	COG0031, CysK, Cysteine synthase	P32260 <i>Spinacia oleracea</i> Cysteine synthase (1e-91)	AK243242 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.4	LOC_Os06g05700	putative Cystein synthase protein	COG0031, CysK, Cysteine synthase	Q9XE48 <i>O. sativa</i> Cysteine synthase (9e-89)	AK072457 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.5	LOC_Os06g05710	Putative protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₁B</i>	OG-BBa0041E07.6	LOC_Os06g05720	Putative protein	/	/	AK068890 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.7	LOC_Os06g05730	Putative protein	/	/	AK072773 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.8	LOC_Os06g05740	Putative protein	/	/	AK287795 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.9	LOC_Os06g05750	Putative transferase protein	pfam02458, Transferase family	O24645 <i>Dianthus caryophyllus</i> benzoyltransferase protein (2e-12)	AK063346 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.10	LOC_Os06g05760	Putative ubiquitin protein	cd01769, UBL	O14399 <i>S. pombe</i> Ubiquitin-like protein (e-5)	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₁B</i>	OG-BBa0041E07.11*	LOC_Os06g05770	Hypothetical protein	/	/	/
<i>S₁B</i>	OG-BBa0041E07.12	LOC_Os06g05790	Putative transferase protein	pfam02458, Transferase family	/	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₁B</i>	OG-BBa0041E07.13	LOC_Os06g05800	Putative protein	cd00590, RRM (RNA recognition motif)	Q9WVB0 <i>M. musculus</i> RNA-binding protein with multiple splicing (6e-12)	AK068636 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.14	LOC_Os06g05804	Putative trafficking protein	cl02131, Sedlin_N	Q54UU1 <i>Dictyostelium discoideum</i> Trafficking protein particle complex (3e-31)	AK241792 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.15	LOC_Os06g05820	Putative protein	cl09099, P-loop NTPase; PRK11823, DNA repair protein RadA	P37572 <i>Bacillus subtilis</i> DNA repair protein radA homolog (2e-92)	AB111516 <i>O. sativa</i> OsRadA mRNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.16	LOC_Os06g05830	Putative protein kinase	ed00180, S_TKc, Serine/Threonine protein kinases; pfam00069, Pkinase	O24585 <i>Z. mays</i> Putative receptor protein kinase CRINKLY4 (4e-48)	AK106925 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.17	LOC_Os06g05860	Putative phosphofructokinase protein	cl00204, PFK, Phosphofructokinase	Q94AA4 <i>A. thaliana</i> 6-phosphofructokinase 3 (0.0)	BT063863 <i>Z. mays</i> flcDNA mRNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.18	LOC_Os06g05870	Putative protein phosphatase	cl00053, PTPc Protein tyrosine phosphatases	P0C089 <i>Ratus norvegicus</i> Protein-tyrosine phosphatase mitochondrial 1 (1e-25)	BT068138 <i>Z. mays</i> flcDNA clone (4e-176)
<i>S₁B</i>	OG-BBa0041E07.19	LOC_Os06g05880	Putative profilin protein	cd00148, Profilin binds actin monomers	Q5VMJ3 <i>O. sativa</i> Profilin LP04 (6e-75)	CT833929 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.20	LOC_Os06g05890	Putative zinc-finger protein	cd00021, BBOX, B-Box-type zinc finger; zinc binding domain (CHC3H2)	Q9SYM2 <i>A. thaliana</i> Putative salt tolerance-like protein (1e-43)	AK104083 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.21	LOC_Os06g05900	Putative Methylase protein	COG2227, 2-polyphenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol methylase	O49354 <i>A. thaliana</i> Hexaprenyldihydroxybenzoate methyltransferase (4e-109)	AK064315 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.22	LOC_Os06g05910	Putative methyltransferases	cd02440, S-adenosylmethionine-dependent methyltransferases	O49354 <i>A. thaliana</i> Hexaprenyldihydroxybenzoate methyltransferase (4e-49)	AK064640 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.23	LOC_Os06g05920	Putative PRR protein	pfam01535, PPR repeat	Q9F1X3 <i>A. thaliana</i> Pentatricopeptide repeat-containing protein (4e-170)	AP008212 <i>O. sativa</i> , chr. 6 (0.0)
<i>S₁B</i>	OG-BBa0041E07.24	LOC_Os06g05930	Putative protein	/	/	CT828676 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.25	LOC_Os06g05940	Putative protein	cl02724, DUF821	A0NDG6 <i>Anopheles gambiae</i> CAP10 family protein AGAP004267 (2e-18)	AK109934 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.26	LOC_Os06g05950	Putative protein	/	/	AK099578 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.27	LOC_Os06g05960	Putative protein	/	/	AP008212 <i>O. sativa</i> , chr. 6 (3e-153)
<i>S₁B</i>	OG-BBa0041E07.28	LOC_Os06g05980	Putative protein	cl01037, Integral membrane protein DUF6	Q10354 <i>S. pombe</i> Uncharacterized transporter (8e-48)	AK072625 <i>O. sativa</i> flcDNA (0.0)
<i>S₁B</i>	OG-BBa0041E07.29*	LOC_Os06g05990	/	/	/	/

p: partial gene; *: Pseudogene.

Supplemental Table IV-3. List of the different types of TE found in the *O. glaberrima* *S*₁ region.

Class	Sub class	Types	Numbers
I	Non LTR retrotransposons	LINEs	5
		SINEs	13
	LTR retrotransposons	Solo LTR	11
		Complete	4
		partial	4
II	MITEs	Tourist	5
		Stowaway	65
		Other	128
	Transposons	-	138

Supplemental Table IV-4. Evaluation of the insertion time of intact LTR retrotransposons in the *O. glaberrima* and *O. sativa* *S*₁ region.

Name of complete LTR retrotransposons	Coordinates in <i>O. glaberrima</i> (bp)	Coordinates in <i>O. sativa</i> (bp)	LTR lenght (bp)	LTR % of nucleotide identity	Number of synonymous substitution/site/year	Insertion time in glaberrima (MYA)	Insertion time in sativa (MYA)
Orstert00200220	335483-346812	/	439-439	97,7	0,0233 (S.E. 0,007)	0,896 (0,269)	/
SC3	470958-476663	/	366/366	100	0	0	/
BAJIE	525043-533430	496332-497213 (solo)	765/765	98,3	0,0172 (S.E. 0,005)	0,661 (0,192)	/
SZ-59	492532-498164	441302-446946	344-360/353-361	93,6/97,8	0,0207 (S.E. 0,008)/ 0,0225 (S.E. 0,008)	0,796 (0,307)	0,865 (0,307)
RETRO2	/	314886-327785	1089-1089	98,5	0,0121 (S.E. 0,003)	/	0,465 (0,115)
RIRE5	/	561516-567661	755-899	84	0,0342 (S.E. 0,007)	/	1,315 (0,269)
OSR4	/	666273-671925	330-350	94,3	0,006 (S.E. 0,004)	/	0,230 (0,153)
Orstert00t0095	/	763051-775978	1101-1110	99,2	0,0036 (S.E. 0,002)	/	0,138 (0,076)

The rate of synonymous substitution/site/year used for calculation was 1.3×10^{-8} (Ma and Bennetzen, 2004).

Supplemental Table IV-5. List of identified pack-MULEs and enclosed genes in the *O. glaberrima* cv. CG14 *S_I* region.

Name of transposon	Coordinates (bp)	Enclosed gene	Putative Function	Protein domain	Best BLASTX homology (swiss prot)	Best BLASTN homology
Pack-MULE#1	96761-98626	OG-BBa0093E08.145	Hypothetical protein; Pseudogene	/	/	AP002071 <i>O. sativa</i> chr. 6 (5e-172)
Pack-MULE#2	130291-132132	OG-BBa0093E08.195	Putative protein	cl00321, AAT_I, Aspartate aminotransferase (AAT) superfamily	O07051, Acromonas jandei L-allo-threonine aldolase (e-7)	AP003708 <i>O. sativa</i> chr. 6 (1e-135)
Pack-MULE#3	165290-167410	OG-BBa0066E18.42	Hypothetical protein; Pseudogene	/	/	AP003708 <i>O. sativa</i> chr. 6 (0.0)
Pack-MULE#4	172505-173829	OG-BBa0066E18.45	Hypothetical protein	/	/	AP003708 <i>O. sativa</i> chr. 6 (0.0)
Pack-MULE#5	201668-203440	OG-BBa17A24.25	Hypothetical protein; Pseudogene	/	/	AP004329 <i>O. sativa</i> chr. 6 (0.0)
Pack-MULE#6	270384-271956	OG-BBa0049I08.4	Hypothetical protein	cd06163 zinc metalloproteases	P73714 Synechocystis (1e-9)	AC136219 <i>O. sativa</i> BAC (0.0)
Pack-MULE#7	272132-273102	OG-BBa0049I08.45	Hypothetical protein; Pseudogene	/	/	AC109601 <i>O. sativa</i> BAC (8e-41)
Pack-MULE#8	294743-296002	OG-BBa0049I08.75	Hypothetical protein	/	/	Ak107277 <i>O. sativa</i> fleDNA (1e-115)
Pack-MULE#9	307632-308718	OG-BBa0049I08.9	Hypothetical protein; Pseudogene	/	/	AP004090 <i>O. sativa</i> BAC (8e-69)
Pack-MULE#10	310525-311981	OG-BBa0049I08.10	Hypothetical protein; Pseudogene	/	P47927 <i>A. thaliana</i> Floral homeotic protein APETALA 2 (4e-8)	AC121363 <i>O. sativa</i> BAC (3e-128)

Supplemental Table IV-6. List and size contribution of TE found differentially inserted between the *O. glaberrima* and *O. sativa* *S_I* regions.

Class	Types	<i>O. glaberrima</i> (Number)	<i>O. glaberrima</i> (Size, bp)	<i>O. sativa</i> (Number)	<i>O. sativa</i> (Size, bp)
I	SINEs	5	1274	0	0
	LTR retrotransposons	14	36124	17	77276
II	MITEs	39	9425	28	6008
	Transposons	35	24183	26	26059
III	Helitron	0	0	2	8147
Other	Other	2	1946	0	0
Total		95	72952	73	117490

Supplemental Table IV-7. Sequence comparison between orthologous coding sequences in the *O. glaberrima* cv. CG14 and *O. sativa* cv. Nipponbare *S₁* regions.

Locus	<i>O. glaberrima</i> Gene Name	Putative Function	Putative <i>O. sativa</i> (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	Ka (dN)	Ks (dS)	Ka/Ks
-	OG-BBa0093E08.1 ^b	Putative Scel-like protein	LOC_Os06g04450	NA	NA	NA	NA	NA	NA	NA
-	OG-BBa0093E08.2	Putative protein	LOC_Os06g04460	100	100	100	867	0.0000	0.0000	0.3504
-	OG-BBa0093E08.3*	Putative protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
-	OG-BBa0093E08.5	Putative protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
-	OG-BBa0093E08.6	Putative protein	LOC_Os06g04470	94.4	92.2	95.1	1233	0.0372	0.0914	0.4065
-	OG-BBa0093E08.7	Putative protein	LOC_Os06g04480	95	95.2	96.8	378	0.0104	0.1226	0.0845
-	OG-BBa0093E08.8	Putative protein	LOC_Os06g04490	87.1	85.5	86.2	455	0.0167	0.1534	0.1092
-	OG-BBa0093E08.9	Putative cornichon protein	LOC_Os06g04500	99.5	100	100	408	0.0000	0.0190	0.0010
-	OG-BBa0093E08.10	Putative Enolase	LOC_Os06g04510	99.2	99.3	99.3	1341	0.0030	0.0252	0.120
-	OG-BBa0093E08.11	Putative protein	LOC_Os06g04520	81.9	81.6	81.6	1995	0.003	0.0033	0.9003
-	OG-BBa0093E08.12	Putative protease	LOC_Os06g04530	99.7	99.3	99.3	894	0.0030	0.0048	0.6253
-	OG-BBa0093E08.13	Putative protein	LOC_Os06g04540	99.1	98.8	98.8	990	0.0037	0.0189	0.195
-	OG-BBa0093E08.14	Putative Armadillo protein	LOC_Os06g04560	99.9	100	100	2676	0.0000	0.0024	0.001
-	OG-BBa0093E08.145*	Hypothetical protein	LOC_Os06g04530	NA	NA	NA	NA	NA	NA	NA
-	OG-BBa0093E08.15	Putative protein	LOC_Os06g04580	100	100	100	951	0.0000	0.0000	0.4546
-	OG-BBa0093E08.16	Putative Auxin responsive protein	LOC_Os06g04590	100	100	100	360	0.0000	0.0000	0.001
<i>S_{1A}</i>	OG-BBa0093E08.17	Putative protein	LOC_Os06g04600	96.7	83.5	85.2	1216	0.0739	0.1474	0.5011
<i>S_{1A}</i>	OG-BBa0093E08.18	Putative Ribosome Binding factor	LOC_Os06g04610	99.8	100	100	630	0.0000	0.0059	0.0010
<i>S_{1A}</i>	OG-BBa0093E08.19	Putative Epimerase	LOC_Os06g04620	99.7	99.7	100	1128	0.001	0.023	0.0420
<i>S_{1A}</i>	OG-BBa0093E08.195	Putative protein	LOC_Os06g04630	100	100	100	498	0.0000	0.0000	0.2965
<i>S_{1A}</i>	OG-BBa0093E08.20	Putative transcription factor	LOC_Os06g04640	99.8	100	100	1083	0.0000	0.0053	0.0010
<i>S_{1A}</i>	OG-BBa0093E08.21	Putative methionine sulfoxide reductase	LOC_Os06g04650	99.7	100	100	765	0.0000	0.0102	0.0010
<i>S_{1A}</i>	OG-BBa0093E08.22	Putative protein	LOC_Os06g04660	99.5	99	99.2	1851	0.0046	0.0059	0.777
<i>S_{1A}</i>	OG-BBa66E18.1	Putative CAF1 ribonuclease	LOC_Os06g04670	99	98.8	99.1	1008	0.0022	0.0261	0.0845
<i>S_{1A}</i>	OG-BBa66E18.2	Putative protein	LOC_Os06g04680	99.2	99.2	99.5	1152	0.001	0.013	0.0783
<i>S_{1A}</i>	OG-BBa66E18.3	Putative protein	LOC_Os06g04699	96.9	94	96	1053	0.0276	0.0571	0.4831
<i>S_{1A}</i>	OG-BBa66E18.4	Putative F-box protein	LOC_Os06g04690	96.9	94	96.9	1560	0.0299	0.0369	0.8120
<i>S_{1A}</i>	OG-BBa66E18.42*	Hypothetical protein	LOC_Os06g04760	NA	NA	NA	NA	NA	NA	NA
<i>S_{1A}</i>	OG-BBa66E18.45	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S_{1A}</i>	OG-BBa66E18.5	Putative protein	LOC_Os06g04780	98.2	98.1	98.4	936	0.0054	0.0487	0.1109
<i>S_{1A}</i>	OG-BBa66E18.6	Putative HAD phosphatase	LOC_Os06g04790	96.3	94.9	98	882	0.0266	0.1089	0.244
<i>S_{1A}</i>	OG-BBa66E18.7	Putative proteasome subunit	LOC_Os06g04800	98.5	98.8	99.6	741	0.0056	0.0427	0.1318
<i>S_{1A}</i>	OG-BBa66E18.8	Putative LRR protein	LOC_Os06g04810	86.9	79.8	86.3	2892	0.1114	0.2014	0.5528
<i>S_{1A}</i>	OG-BBa17A24.1	Putative methyltransferase	LOC_Os06g04820	97.5	97.1	98	1026	0.0119	0.054	0.2194
<i>S_{1A}</i>	OG-BBa17A24.2	Putative LRR protein	LOC_Os06g04830	88	82.7	87.2	2955	0.0696	0.0839	0.8295
<i>S_{1A}</i>	OG-BBa17A24.25*	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S_{1A}</i>	OG-BBa17A24.3	Putative LRR protein	LOC_Os06g04840	95	92.4	93.2	2754	0.0229	0.0351	0.6528
<i>S_{1A}</i>	OG-BBa17A24.4	Putative protein	LOC_Os06g04850	98.3	97.3	98.1	780	0.0063	0.0001	99
<i>S_{1A}</i>	OG-BBa17A24.5	Hypothetical protein	LOC_Os06g04860	99.5	100	100	195	0.0000	0.0156	0.001
<i>S_{1A}</i>	OG-BBa17A24.6	Putative protein	LOC_Os06g04870	99.5	99	99.4	930	0.0025	0.0000	99
<i>S_{1A}</i>	OG-BBa17A24.7	Putative serine/Threonine protein kinase	LOC_Os06g04880	99.9	99.9	99.9	2421	0.0006	0.0032	0.1766
<i>S_{1A}</i>	OG-BBa17A24.8	Putative protein	LOC_Os06g04889	100	100	100	207	0.0000	0.0000	0.0761
<i>S_{1A}</i>	OG-BBa17A24.9	Putative protein	LOC_Os06g04900	99.9	99.8	99.8	1563	0.0007	0.0072	0.0983
<i>S_{1A}</i>	OG-BBa17A24.10	Putative NADH pyrophosphatase protein	LOC_Os06g04910	99.3	99	99.3	1224	0.0025	0.0000	99
<i>S_{1A}</i>	OG-BBa0049I08.2	Putative Zinc finger protein	LOC_Os06g04920	99.3	99.3	100	438	0.0028	0.0266	0.1063
<i>S_{1A}</i>	OG-BBa0049I08.3	Putative ENOD93 protein	LOC_Os06g04930	61.3	60.9	62.2	NA	NA	NA	NA
<i>S_{1A}</i>	OG-BBa0049I08.4	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S_{1A}</i>	OG-BBa0049I08.45*	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049I08.5	Putative ENOD93 protein	LOC_Os06g04940	65	54.1	59.5	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049I08.6	Putative ENOD93 protein	LOC_Os06g04950	89.1	76.8	78.4	376	NA	NA	NA
<i>S₁</i>	OG-BBa0049I08.7	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049I08.75	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049I08.8	Putative F-box protein	LOC_Os06g04970	98.6	97.7	98.4	2073	0.0112	0.0237	0.4728

Supplemental Table IV 7. Continuation

Locus	<i>O. glaberrima</i> Gene Name	Putative Function	Putative <i>O. sativa</i> (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	Ka (dN)	Ks (dS)	Ka/Ks
<i>S₁</i>	OG-BBa0049108.85*	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049108.9*	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049108.10*	Hypothetical protein	no orthologous gene	NA	NA	NA	NA	NA	NA	NA
<i>S₁</i>	OG-BBa0049108.11	Putative protein	LOC_Os06g04980	98.1	96.5	97.7	1542	0.0173	0.0198	0.8715
<i>S₁</i>	OG-BBa0049108.12	Putative ENOD93 protein	LOC_Os06g04990	99.7	100	100	348	0.0000	0.0142	0.001
<i>S₁B</i>	OG-BBa0049108.13	Putative ENOD93 protein	LOC_Os06g05000	100	100	100	348	0.0000	0.0000	99
<i>S₁B</i>	OG-BBa0049108.14	Putative ENOD93 protein	LOC_Os06g05010	99.4	100	100	351	0.000025	0.0260	0.001
<i>S₁B</i>	OG-BBa0049108.15	Putative ENOD93 protein	LOC_Os06g05020	99.1	98.3	99.1	348	0.0074	0.0143	0.5175
<i>S₁B</i>	OG-BBa0049108.16	Putative Serine/Threonine protein kinases	LOC_Os06g05050 ^c	92	87.8	92.1	2291	0.0447	0.0571	0.7828
<i>S₁B</i>	OG-BBa0049108.17	Putative Early flowering protein	LOC_Os06g05060	99.2	98.8	99.1	2283	0.0057	0.0155	0.3669
<i>S₁B</i>	OG-BBa0049108.18	Putative Serine/Threonine protein kinases	LOC_Os06g05070	97.6	97.6	98.5	1386	0.0081	0.1700	0.0465
<i>S₁B</i>	OG-BBa0049108.19	Putative cytochrome oxidase	LOC_Os06g05080	99.8	99.4	99.4	465	0.0032	0.0000	99
<i>S₁B</i>	OG-BBa0049108.20	Putative methyltransferase protein	LOC_Os06g05090	98.8	98.4	99	1143	0.0071	0.0299	0.2366
<i>S₁B</i>	OG-BBa0049108.21	Putative 1-deoxy-D-xylulose-5-phosphate synthase	LOC_Os06g05100	99.9	99.6	99.7	2169	0.0018	0.0000	99
<i>S₁B</i>	OG-BBa0049108.22	Putative Superoxide dismutase	LOC_Os06g05110	99.9	99.6	100	768	0.0016	0.0000	99
<i>S₁B</i>	OG-BBa0049108.23	Putative protein	LOC_Os06g05120	NA	NA	NA	NA	NA	NA	NA
<i>S₁B</i>	OG-BBa0049108.24	Putative Myristoyl-acyl carrier protein thioesterase	LOC_Os06g05130	97.4	99	99	1284	0.0044	0.0300	0.1457
<i>S₁B</i>	OG-BBa0049108.25	Putative PPR protein	LOC_Os06g05140	99.8	99.6	99.6	2292	0.0018	0.0037	0.4789
<i>S₁B</i>	OG-BBa0049108.26	Putative protein	LOC_Os06g05150 ^c	98	97.2	97.5	1296	0.0070	0.0048	1.471
<i>S₁B</i>	OG-BBa0049108.27	Putative sulfate transporter	LOC_Os06g05160 ^p	99.2	99.4	99.6	2013	0.0018	0.0373	0.0489
<i>S₁B</i>	OG-BBa0045G15.1	Putative Coatomer protein	LOC_Os06g05180	99.3	98.7	98.8	2736	0.0054	0.0098	0.5528
<i>S₁B</i>	OG-BBa0045G15.2	Putative BRCA1 protein	LOC_Os06g05190	99.6	99.1	99.7	1044	0.0042	0.0032	1.3209
<i>S₁B</i>	OG-BBa0088O22.1	Putative Zinc finger protein	LOC_Os06g05200	91.4	89.8	90.3	618	0.0086	0.0414	0.2088
<i>S₁B</i>	OG-BBa0088O22.2	Putative pectate lyase protein	LOC_Os06g05209	76.3	76.7	76.9	1362	0.0020	0.2075	0.0098
<i>S₁B</i>	OG-BBa0088O22.3	Putative pectate lyase protein	LOC_Os06g05260	99	98.4	98.4	1353	0.0041	0.0209	0.1942
<i>S₁B</i>	OG-BBa0088O22.4	Putative GTP-binding protein	LOC_Os06g05250	99.8	99.7	99.8	1992	0.0014	0.0035	0.4173
<i>S₁B</i>	OG-BBa0088O22.5	Putative carboxypeptidase	LOC_Os06g05240	99.4	98.3	98.3	1053	0.0076	0.0001	99
<i>S₁B</i>	OG-BBa0088O22.6	Hypothetical protein	LOC_Os06g05220	72.1	67.1	68.8	728	0.0526	0.0927	0.5674
<i>S₁B</i>	OG-BBa0088O22.7	Putative pectate lyase protein	LOC_Os06g05272	76.2	76.7	76.9	1362	0.0020	0.2409	0.0084
<i>S₁B</i>	OG-BBa0088O22.8	Putative Transferase	LOC_Os06g05284	99	98.4	98.8	1473	0.0057	0.0198	0.2873
<i>S₁B</i>	OG-BBa0088O22.9	Putative Transferase	LOC_Os06g05300	99.3	99.2	99.2	1479	0.0008	0.0053	0.1513
<i>S₁B</i>	OG-BBa0088O22.10	Putative Transferase	LOC_Os06g05310	96.9	95.5	97.1	1353	0.0185	0.0380	0.4872
<i>S₁B</i>	OG-BBa0088O22.11	Putative Transferase	LOC_Os06g05320	65.6	53.8	55.2	1458	0.0763	0.2235	0.3414
<i>S₁B</i>	OG-BBa0088O22.12	Putative AP2 proyein	LOC_Os06g05340	99.3	99.5	99.5	1200	0.0000	0.0083	0.0010
<i>S₁B</i>	OG-BBa0088O22.13	Putative Whirly transcription factor protein	LOC_Os06g05350	99.4	98.9	99.3	819	0.0052	0.0086	0.6043
<i>S₁B</i>	OG-BBa0088O22.14	Putative protein	LOC_Os06g05359	99.2	98.1	98.5	3081	0.0093	0.0065	1.4143
<i>S₁B</i>	OG-BBa0088O22.15	Putative protein	LOC_Os06g05368	72.7	55	59.7	432	NA	NA	NA
<i>S₁B</i>	OG-BBa0088O22.16	Putative WRKY protein	LOC_Os06g05380	99.6	99.4	99.4	1584	0.0025	0.0089	0.2758
<i>S₁B</i>	OG-BBa0088O22.17	Putative protein	LOC_Os06g05390	99.8	99.7	99.7	1023	0.0013	0.0042	0.3090
<i>S₁B</i>	OG-BBa0088O22.18	Putative Iron-sulphur protein	LOC_Os06g05400	99.6	99.5	100	561	0.0025	0.0066	0.3816
<i>S₁B</i>	OG-BBa0088O22.19	Hypothetical protein	LOC_Os06g05410	87.5	86.2	87.7	393	0.0693	0.1361	0.5093
<i>S₁B</i>	OG-BBa0056F23.1	Putative protein	LOC_Os06g05420	99.4	98.3	98.3	348	0.0083	0.0001	99
<i>S₁B</i>	OG-BBa0056F23.2	Putative protein	LOC_Os06g05424	100	100	100	276	0.0000	0.0000	0.0010
<i>S₁B</i>	OG-BBa0056F23.3	Putative protein	LOC_Os06g05430	95.5	94.1	95.8	357	0.0262	0.1303	0.2012
<i>S₁B</i>	OG-BBa0056F23.4	Putative protein	LOC_Os06g05440	99.2	98.3	99.2	360	0.0063	0.0355	0.1768
<i>S₁B</i>	OG-BBa0056F23.5	Putative protein	LOC_Os06g05470	100	100	100	459	0.0000	0.0000	0.0010
<i>S₁B</i>	OG-BBa0056F23.6	Putative protein	LOC_Os06g05480	100	100	100	423	0.0000	0.0000	0.0010
<i>S₁B</i>	OG-BBa0056F23.7	Putative protein	LOC_Os06g05510	96.7	96.5	96.5	423	0.0032	0.0119	0.2625
<i>S₁B</i>	OG-BBa0056F23.8	Putative protein kinase	LOC_Os06g05520	99.9	100	100	1059	0.0000	0.0027	0.0010
<i>S₁B</i>	OG-BBa0056F23.9	Putative protein	LOC_Os06g05530	99.1	99.1	99.6	699	0.0041	0.0200	0.2078

Supplemental Table IV 7. Continuation

Locus	<i>O. glaberrima</i> Gene Name	Putative Function	Putative <i>O. sativa</i> (Nipponbare) orthologous gene	% of nucleotide identity	% of protein identity	% of similarity	Length (bp)	Ka (dN)	Ks (dS)	Ka/Ks
<i>S₁B</i>	OG-BBa0056F23.10	Hypothetical protein	LOC_Os06g05540	99.1	98.5	98.5	582	0.0073	0.0121	0.6084
<i>S₁B</i>	OG-BBa0056F23.11	GDSL esterase/lipase protein	LOC_Os06g05550	99.8	100	100	1056	0.0000	0.0103	0.0010
<i>S₁B</i>	OG-BBa0056F23.12	Putative protein	LOC_Os06g05560	98.1	95.9	97.1	513	0.0213	0.0181	1.1748
<i>S₁B</i>	OG-BBa0056F23.13	Putative F-box protein	LOC_Os06g05580	99.1	98.2	98.2	1182	0.0084	0.0122	0.6896
<i>S₁B</i>	OG-BBa0056F23.14*	Putative F-box protein	LOC_Os06g05590	NA	NA	NA	NA	NA	NA	NA
<i>S₁B</i>	OG-BBa0056F23.15	Putative F-box protein	LOC_Os06g05600	99.5	99.2	99.5	1194	0.0033	0.0116	0.2812
<i>S₁B</i>	OG-BBa0056F23.16	Putative F-box protein	LOC_Os06g05610	98.4	98.1	98.9	1107	0.0086	0.0419	0.2053
<i>S₁B</i>	OG-BBa0056F23.17*	Putative F-box protein	LOC_Os06g05620	NA	NA	NA	NA	NA	NA	NA
<i>S₁B</i>	OG-BBa0056F23.18	GDSL esterase/lipase protein	LOC_Os06g05630	99.5	99.2	99.4	1083	0.0036	0.0090	0.3962
<i>S₁B</i>	OG-BBa0056F23.19	Putative protein	LOC_Os06g05640	99.5	98.9	99.2	1854	0.0053	0.0059	0.9122
<i>S₁B</i>	OG-BBa0056F23.20	Putative protein	LOC_Os06g05650	99.5	98.5	100	198	0.0075	0.0001	99
<i>S₁B</i>	OG-BBa0041E07.1	Putative Nucleosome protein	LOC_Os06g05660	99.8	100	100	1137	0.0000	0.0118	0.0010
<i>S₁B</i>	OG-BBa0041E07.2	Putative protein	LOC_Os06g05670	100	100	100	354	0.0000	0.0000	0.0010
<i>S₁B</i>	OG-BBa0041E07.3	putative Cystein synthase protein	LOC_Os06g05690	99.8	99.7	100	1020	0.0013	0.0040	0.3329
<i>S₁B</i>	OG-BBa0041E07.4	putative Cystein synthase protein	LOC_Os06g05700	99.8	99.7	100	1044	0.0015	0.0000	99
<i>S₁B</i>	OG-BBa0041E07.5	Putative protein	LOC_Os06g05710	98.5	98.3	98.3	717	0.0052	0.0266	0.1943
<i>S₁B</i>	OG-BBa0041E07.6	Putative protein	LOC_Os06g05720	99.8	99.7	100	858	0.0016	0.0048	0.3245
<i>S₁B</i>	OG-BBa0041E07.7	Putative protein	LOC_Os06g05730	98.5	98.7	98.7	672	0.0000	0.0089	0.0010
<i>S₁B</i>	OG-BBa0041E07.8	Putative protein	LOC_Os06g05740	99.5	99	99	921	0.0042	0.0105	0.3977
<i>S₁B</i>	OG-BBa0041E07.9	Putative transferase protein	LOC_Os06g05750	99.7	100	100	1425	0.0000	0.0378	0.0010
<i>S₁B</i>	OG-BBa0041E07.10	Putative ubiquitin protein	LOC_Os06g05760	99.7	97	98	915	0.0134	0.0586	0.2277
<i>S₁B</i>	OG-BBa0041E07.11*	Hypothetical protein	LOC_Os06g05770	NA	NA	NA	NA	NA	NA	NA
<i>S₁B</i>	OG-BBa0041E07.12	Putative transferase protein	LOC_Os06g05790	78.7	78.7	78.9	1803	0.0039	0.1038	0.0371
<i>S₁B</i>	OG-BBa0041E07.13	Putative protein	LOC_Os06g05800	99.6	100	100	903	0.0000	0.0206	0.0010
<i>S₁B</i>	OG-BBa0041E07.14	Putative trafficking protein	LOC_Os06g05804	99.5	100	100	432	0.0000	0.0135	0.0010
<i>S₁B</i>	OG-BBa0041E07.15	Putative protein	LOC_Os06g05820	99.6	99.4	99.5	1860	0.0032	0.0068	0.4742
<i>S₁B</i>	OG-BBa0041E07.16	Putative protein kinase	LOC_Os06g05830	98.8	97.7	97.7	1600	0.0103	0.0132	0.7807
<i>S₁B</i>	OG-BBa0041E07.17	Putative phosphofructokinase protein	LOC_Os06g05860	99.5	99.5	99.6	1680	0.0016	0.0109	0.1431
<i>S₁B</i>	OG-BBa0041E07.18	Putative protein phosphatase	LOC_Os06g05870	98.3	98.9	99.3	816	0.0051	0.0525	0.0978
<i>S₁B</i>	OG-BBa0041E07.19	Putative profilin protein	LOC_Os06g05880	99.7	100	100	396	0.0000	0.0167	0.0010
<i>S₁B</i>	OG-BBa0041E07.20	Putative zinc-finger protein	LOC_Os06g05890	98.5	99.4	100	1083	0.0027	0.0466	0.0569
<i>S₁B</i>	OG-BBa0041E07.21	Putative Methylase protein	LOC_Os06g05900	97.9	97.9	98.5	984	0.0059	0.0261	0.2250
<i>S₁B</i>	OG-BBa0041E07.22	Putative methyltransferases	LOC_Os06g05910	98.9	97.9	98.3	711	0.0104	0.0142	0.7281
<i>S₁B</i>	OG-BBa0041E07.23	Putative PRR protein	LOC_Os06g05920	92.6	92.9	93.2	2197	0.0216	0.0420	0.5148
<i>S₁B</i>	OG-BBa0041E07.24	Putative protein	LOC_Os06g05930	100	100	100	198	0.0000	0.0000	NA
<i>S₁B</i>	OG-BBa0041E07.25	Putative protein	LOC_Os06g05940	97.7	96.7	98	1650	0.0127	0.0958	0.1325
<i>S₁B</i>	OG-BBa0041E07.26	Putative protein	LOC_Os06g05950	97.6	97.2	98	1185	0.0115	0.0538	0.2139
<i>S₁B</i>	OG-BBa0041E07.27	Putative protein	LOC_Os06g05960	100	100	100	306	0.0000	0.0000	NA
<i>S₁B</i>	OG-BBa0041E07.28	Putative protein	LOC_Os06g05980	99.2	99.8	99.8	1296	0.0000	0.0292	0.0010
<i>S₁B</i>	OG-BBa0041E07.29 ^p	/	LOC_Os06g05990	NA	NA	NA	NA	NA	NA	NA

p: Partial gene; *: Pseudogene.

V

EXPLORATION DE LA REGION DU LOCUS S_I CHEZ

O. glaberrima* ET SON ANCETRE *O. barthii

Introduction

Il est admis qu'*O. glaberrima* fut domestiquée à partir de *O. barthii*, dans le delta du fleuve Niger – considéré comme le centre primaire d'origine –, peut-être 1000 ans avant J.C. (Portères, 1962; Murray, 2004). Ces deux espèces montrent une différenciation liée à leur écologie, ce qui permet de classer leurs accessions en deux écotypes principaux, un écotype d'eaux profondes et un écotype pluvial strict. Cette différenciation des deux espèces selon les écologies semble indiquer que le processus de domestication a permis de conserver une partie de la variabilité et des adaptations environnementales d'*O. barthii* dans les accessions d'*O. glaberrima* (Bezançon, 1994). Toutefois, comme chez *O. sativa*, la domestication a entraîné une réduction de la diversité allélique chez *O. glaberrima* par rapport à son ancêtre (Bezançon *et al.*, 1989).

O. sativa fut introduite en Afrique, principalement par les navigateurs européens au XVI^{ème} siècle (Bezançon, 1994). Malgré son statut d'espèce cultivée autochtone en Afrique, *O. glaberrima* a été progressivement reléguée au second plan après l'introduction d'*O. sativa* sur le continent, voyant ainsi sa diversité génétique menacée. Vers la moitié du XX^{ème} siècle, la culture du riz en Afrique était basée sur un système de parcelles cultivées par les villageois, en mélanges de variétés d'*O. sativa* et d'*O. glaberrima*, adaptées aux conditions et aux préférences du village (Pélissier, 1966). Actuellement, on distingue deux types de riziculture en Afrique : les zones avec une bonne maîtrise de l'eau et presque exclusivement cultivées en *O. sativa*, et les zones à inondation non contrôlée ou d'eaux profondes où on trouve encore des variétés d'*O. glaberrima* (Bezançon, 1994). Par ailleurs, et par le biais de ces cultures en mélange pendant plus de 500 ans et malgré la forte barrière de reproduction entre les deux espèces, on trouve des formes hybrides fertiles parmi les accessions prospectées en Afrique (Semon *et al.*, 2005; Barry *et al.*, 2007; Nuijten *et al.*, 2009).

Dans les chapitres précédents, nous avons présenté plusieurs résultats concernant la cartographie génétique fine et physique du locus S_I sur quatre populations interspécifiques, ainsi qu'une analyse et une comparaison détaillée de la région de S_I entre la séquence annotée disponible d'*O. sativa* ssp. *japonica* cv. Nipponbare (TIGR version 6) et la séquence de la variété d'écologie pluviale d'*O. glaberrima* CG14. Cependant, la population utilisée pour la cartographie et représentée par le plus grand nombre d'individus (459 plantes) fut dérivée du croisement entre la variété d'écologie d'eaux profondes d'*O. glaberrima* TOG5681 et l'accession *O. sativa* ssp. *indica* IR64. Etant donné les différences existantes – au moins au niveau phénotypique – entre les accessions d'écologie pluviale stricte et inondée, il apparaît

nécessaire de tester l'hypothèse que l'interaction épistatique entre les trois facteurs S_1A , S_1 et S_1B , est présente chez les deux écotypes d'*O. glaberrima*.

Afin de tenter de répondre à cette question, nous avons utilisé une banque BAC de l'accession TOG5681, construite à l'IRD il y a quelques années. Le criblage de la librairie nous a permis de construire la carte physique de la région contenant l'intervalle de probabilité maximale de localisation de S_1A , le locus S_1 , et une partie de l'intervalle de probabilité de localisation de S_1B . Le séquençage préliminaire des quatre clones BAC qui constituent le *minimum tiling path* (MTP) a permis de mettre en évidence une très forte identité entre les séquences codantes de CG14 et TOG5681.

Nous avons ensuite réalisé un examen de la structure de la région $S_1A - S_1 - S_1B$ dans une collection représentative de la diversité d'*O. glaberrima* et de son ancêtre *O. barthii*. Cette analyse, réalisée avec des marqueurs d'insertion d'éléments transposables et des marqueurs microsatellites, a mis en évidence le fait que les hybrides interspécifiques naturels rapportés par la littérature ou identifiés dans le cadre de nos recherches possèdent une configuration haplotypique correspondant soit à *O. sativa*, soit à *O. glaberrima*, sur l'intégralité de la région définie par l'intervalle $S_1A - S_1B$. En outre, la fertilité en croisement avec *O. sativa* de ces lignées hybrides s'avère plus élevée pour celles qui possèdent l'haplotype *O. sativa*. Cette dichotomie de la région et ces données de fertilité représentent une confirmation du modèle génétique de sélection gamétique que nous proposons au chapitre III.

Les résultats présentés dans ce chapitre représentent une compilation de mes propres résultats, de ceux obtenus par Julie Orjuela, dont j'ai co-encadré le stage de Master 2 « Biologie, Géosciences, Agroressources et Environnement » (BGAE) de l'Université Montpellier II (spécialité Biotraçabilité, Biodétection et Biodiversité (BBB)), et de ceux issus de la collaboration avec les différents partenaires du projet iBridges.

V.1 Matériels et méthodes

V.1.1 Construction de la carte physique du locus S_I dans l'accension *O. glaberrima* TOG5681

Afin de construire la carte physique autour du locus S_I , la banque BAC de l'accension TOG5681 fut criblée par PCR avec les marqueurs SSR de la région définie par les marqueurs RMC6_21494, RMC6_21989, C6_20017 (InDel ; Chapitres III et IV), et les marqueurs désignés auparavant pour confirmer le chevauchement des clones BAC de CG14 :

BES45G15.f (F_GCGATCCAGAAGCTCAAGAC, R_GCGAAGAAATCAGCGGATAG) et BES49I08.f (F_ATCTCCCCAAGGGATTGAAC, R_AGCCATGTTCATGACACCAA).

Les amplifications furent réalisées premièrement à partir de la banque « poolée » en plaques de 384 puits (992 pools). Les 24 clones contenus dans chaque pool positif furent ensuite utilisés comme matrice pour réaliser une amplification individuelle, ce qui a permis de sélectionner les clones positifs. Les PCR furent réalisées avec un programme de type « TouchDown » (14 min. à 94°C, 10 cycles de 30 sec. à 94°C, 1 min. à 65°C avec une diminution de 1°C par cycle, et 1 min. à 72°C ; puis 30 cycles de 30 sec. à 94°C, 1 min. à 55°C, 1 min. à 72°C ; puis une élongation finale de 10 min.), dans un volume final de 15µl (Tampon standard 1X, 200µM de dNTP, 2 pmol de chaque amorce, 1,5mM de MgCl₂, 1U de Taq polymérase). Les produits de l'amplification PCR furent visualisés sur gel d'agarose à 3% coloré au BET.

L'ADN des clones isolés fut extrait par lyse alcaline, et la taille moléculaire de tous les inserts fut estimée par la fonction « Analyse Gels » et « Molecular weight » du logiciel ImageJ (Abramoff *et al.*, 2004), à partir des fragments obtenus par digestion par l'enzyme NotI, séparés par électrophorèse en champ pulsé (22 heures, 5.5v/m, angle 120°, 14°C), et visualisés par coloration au BET. La confirmation de l'appartenance des clones positifs à la région du locus S_I fut réalisée par PCR, selon la méthodologie décrite au Chapitre III.

V.1.2 Séquençage, annotation et comparaison des séquences partielles de TOG5681 avec CG14

Les clones choisis pour constituer le MTP de la région de S_I furent séquencés par la technique de pyroséquençage 454 (Roche). Un premier lot de clones fut traité par Cogenics Genome Express (Grenoble, France) en séquençage 454 FLX, avec une librairie PET (Paired-End Tag Library). Un deuxième lot de clones fut séquencé chez MWG–Operon (Ebersberg, Germany), par 454 FLX Titanium.

Tableau V-1. Marqueurs moléculaires utilisés pour l'évaluation de la structure de la région des locus S_1A - S_1 - S_1B .

Marqueur Moléculaire	Position Chr6 Nipponbare	Type de marqueur	Insertion chez :	Amorce Forward	Amorce Reverse	Amorce LTR	T° d'hybridation (°C)
RM589	1 380 876	SSR	-	(IRGSP, 2005)		/	55
Biodiv_1	1 903 169	Transposon	O.s	CGTATTGTGGCAAAGCTTGA	GGCAGGCTAGTGAGAGGTTG	/	60
Biodiv_2	2 055 080	MITE	O.g	GTGGTGTGCTCATCCCTCTT	TGGCCATTGCAGTATTTTGA	/	60
Biodiv_3	2 074 138	Retrotransposon	O.g	GGATTATCCACGGGATTGG	TGTACGGGGAATTGGATAGG	TCTTCCGCTGATGTTCTCTG	60
Biodiv_3.1	2 108 991	MITE	CG14	GCGATAGAGTGGCAAAGGAG	TCTGCGTCAGTTCCTCAGTG	/	60
Biodiv_4	2 111 570	MITE	CG14	GCTTAACCCACCTGTCAAAT	TGGGTTTGCATTGGATTCT	/	60
Biodiv_4.1	2 127 056	InDel	TOG5681	TAGGTCCTATAGCCCTCAGA	CGTACAGGCCATACATGC	GCCCAAACCAGTTATTAACGA	60
Biodiv_5	2 124 189	MITE	O.s	TTTGGCCTTACACGACACCT	CCATCCAACCTGCACAACAAC	/	60
Biodiv_6	2 160 957	Retrotransposon	O.g	CAAATCCTCCAGTGAACGTG	AGAACAAATGCGCTCAAGGT	GTTACGGCCAAGTTCGGTAA	60
Biodiv_7	2 177 373	MITE	O.s	TCACCACCTTGGTCCCTAGC	GTGAGGTGTTTGTGGTGACG	CACTGTTTCTGGGACTGGT	60
Biodiv_8	2 180 740	Transposon	O.g	AACCCATGTGGAGATGAAGG	GGATAGGCGATCCCTAGAGC	/	60
Biodiv_9	2 187 141	Transposon	O.g	TGGTTGGTTCCTTTTCTACCC	GTATAACGTGGGGCCACTAA	/	60
Biodiv_10	2 194 355	MITE	O.g	GCACGTGTATGGGCTATTTCT	CAGTACTGGATGGACGGACA	/	60
Biodiv_11	2 214 226	Retrotransposon	O.g	TGGGGATGCTGAATCTGATG	ATCTTGTTGGCAGGGTGAAC	/	60
Biodiv_12	2 215 549	Retrotransposon	O.s	AAGGCCGATAGACAGACAGC	GGTGGTGTCTGAAACTGATGAA	GATTCCGCCGAGTGTGTTAT	60
Biodiv_13	2 310 640	MITE	O.g	TTTTGCCTTGGAACGATTTC	TCTTATGCCAAACGGGTAGG	AACCCAGAACAAGGGAGAGG	60
Biodiv_14	2 314 032	MITE	O.s	CGGCCATAGGTGACTGAACT	AAGACCCATTTACGTTTGC	/	60
Biodiv_15	2 362 914	Retrotransposon	O.g	CCAGACAAATTGAGCGACAA	CCACTGGTCTTTTGCAAGT	/	60
Biodiv_16	2 408 450	MITE	O.s	GAGGAATTGGTCCGTGGATA	GACACCATTCACTCTTTCC	GCTGGACATGGCTTTTAGG	60
Biodiv_17	2 450 385	MITE	O.g	TGGTCCCACGTAAAAGAAGG	CCATGCTCACCACCGTATC	/	60
Biodiv_17.1	2 459 719	Transposon	O.s	ATCTCACGGCATAGGACAGC	TTTTGTCCGTGAACTGAAGC	/	60
Biodiv_18	2 462 035	Retrotransposon	O.s	GGTGACTTCGACGGTTTGAC	GCCCGGTCTTAAAAGGAGAG	/	60
Biodiv_18.1	2 473 216	MITE	O.s	CCGGTCCGACACAATTAATA	CCAACCTCACCATCTTTTCC	CAAGTCAATTGCGTTCCTGC	60
Biodiv_18.2	2 542 692	InDel	O.s	AAAACCTCCGTCTTAAACC	ATGTGCAAAGGTGGCATGTA	/	60
Biodiv_19	2 591 061	MITE	O.g	CCAGGGATCGAAGGAAGATA	TGGGTTTCGCTAATCACTCA	/	60
Biodiv_20	2 688 320	MITE	O.s	TAGGGGATGCTCTCTTCTCT	GAGGTGCAAGCGCCTTAGC	/	60
Biodiv_21	2 726 066	InDel	O.s	CCCCCTTACGAACAAGTCA	ATTACCTGCGGGAAGTTGG	/	60
C6_27332	2 733 214	MITE	O.s	GACAGTGTAAAAGGGACATTGG	CAGCCTAGAATTGGGTGGA	/	61
C6_27423	2 742 310	MITE	O.g	TCCGACGATATCGCTTACTC	GGGGCAGGGTTCACAATC	/	61
RMC6_27601*	2 760 116	SSR	na O.g	TCGAAGCAAATTTCCACAGT	AGTTCTCTGGACAAGGTACGC	/	59
RMC6_28000*	2 800 078	SSR	na O.g	CCCAAGGCGATAGGATGG	CTAGCTGTCTCGCACCTCAC	/	59
RM19400	2 810 860	SSR	-	(IRGSP, 2005)		/	55
RM510	2 831 513	SSR	-	(IRGSP, 2005)		/	55
RM19406	2 841 418	SSR	na O.g	(IRGSP, 2005)		/	55
RM3805	2 853 061	SSR	-	(IRGSP, 2005)		/	55

* Marqueurs désignés à partir des annotations de TIGR V6, respectivement putative_ssr_21540 et putative_ssr_21730. **O.s**, Insertion chez *O. sativa*. **O.g**, insertion chez *O. glaberrima*. **na**, pas d'amplification détectée chez *O. glaberrima*.

L'assemblage de la séquence à partir des contigs obtenus fut réalisé en utilisant la séquence de CG14 comme matrice, par alignement graphique de type « dot-plot » avec l'aide du logiciel Dotter (Sonnhammer et Durbin, 1995). L'annotation et la comparaison de la séquence préliminaire obtenue avec celle de CG14 ont été réalisées selon la méthode décrite au chapitre III.

V.1.3 Analyse de la structure de la région $S_{IA} - S_I - S_{IB}$ chez différentes accessions

O. glaberrima et *O. barthii*

L'analyse de la structure de la région $S_{IA} - S_I - S_{IB}$ fut réalisée au moyen de marqueurs d'insertion d'éléments transposables et de marqueurs microsatellites, identifiés par comparaison entre les séquences disponibles d'*O. sativa* et d'*O. glaberrima*. Les polymorphismes générés par l'insertion d'un élément chez *O. sativa* ont été choisis lorsque l'élément était présent chez Nipponbare et 93-11 (si la qualité de l'assemblage de 93-11 le permettait), tandis que les insertions chez *O. glaberrima* pouvaient être exclusifs à CG14, à TOG5681, ou partagées par les deux accessions. Les insertions d'éléments de type LTR Retrotransposons furent détectées par une analyse de type RBIP (Retrotransposon-Based Insertion Polymorphism) (Andrew *et al.*, 1998). Un couple d'amorces fut désigné dans la région flanquant l'insertion, et une amorce additionnelle dans le LTR. Dans le cas des transposons et des éléments MITE, une simple amplification par PCR ciblée fut utilisée afin de détecter un polymorphisme de taille. Le Tableau V-1 récapitule les principales caractéristiques des différents marqueurs utilisés.

En outre, des marqueurs microsatellites déjà disponibles (IRGSP, 2005) (Chapitre III), ou désignés *de novo* à cette fin (voir Tableau V-1), et qui se situent à l'extérieur du contig de CG14, furent utilisés pour examiner la région.

Les PCR furent effectuées avec le programme suivant : 3 min. à 94°C, 35 cycles de 30 sec. à 94°C, 1 min. à 60°C, 2 min. à 72°C, élongation finale de 5 min. dans 20µl (Tampon standard 1X, 200µM de dNTP, 2 pmol de chaque amorce, 1,5mM de MgCl₂, 1U de Taq polymérase), avec environ 20 ng d'ADN. Les produits de l'amplification furent visualisés sur gel d'agarose à 1,2% coloré au BET. Les microsatellites furent évalués comme décrit précédemment, avec 1,5mM de MgCl₂ et une température d'hybridation dépendant du couple d'amorces (Tableau V-1, Chapitre III et IV), et visualisés sur agarose à 3%.

Tableau V-2. Accessions évaluées pour la région de S_1A - S_1 - S_1B .

Espèce	Accession	Code IRGC	Pays	Observations
<i>O. glaberrima</i>	CG14	96717	Sénégal	
	DC Kono		Sierra Leone	Parent iBridges
	IG02		Côte d'Ivoire	Parent iBridges
	MG12	103544	Mali	Parent iBridges
	MG19	103549	Mali	Parent iBridges
	Pa DC Kono		Sierra Leone	Parent iBridges
	RAM 111		Mali	Parent iBridges
	RAM 24		Mali	Parent iBridges
	RAM 48		Mali	Parent iBridges
	RAM 55		Mali	Parent iBridges
	RAM 63		Mali	Parent iBridges
	RAM 90		Mali	Parent iBridges
	RAM 95		Mali	Parent iBridges
	Saliforeh		Sierra Leone	Parent iBridges
	TOG5283	96720	Côte d'Ivoire	
	TOG5284	96721	Côte d'Ivoire	
	TOG5293	96725	Nigéria	
	TOG5307	96726	Nigéria	Parent iBridges
	TOG5390	96740	Nigéria	
	TOG5420	96744	Nigéria	
	TOG5429	112564	Nigéria	
	TOG5435	96750	Nigéria	
	TOG5438	96751	Nigéria	
	TOG5495	96768	Nigéria	
	TOG5602	86762	Ghana	
	TOG5603	86763	Ghana	
	TOG5620	86764	Ghana	
	TOG5641	96786	Nigéria	
	TOG5673	96789	Nigéria	
	TOG5680	86772	Nigéria	
	TOG5681	96793	Nigéria	
	TOG5776	86780	Libéria	
	TOG5785	86781	Libéria	
	TOG5795	96800	Libéria	
	TOG5810	86784	Libéria	
	TOG5815	86785	Libéria	
	TOG5817	112566	Libéria	
	TOG5828	96805	Libéria	
	TOG5885	112568	Libéria	
	TOG5887	86789	Libéria	
	TOG5919	86790	Libéria	
	TOG5923	86791	Libéria	
	TOG5980	96816	Nigéria	Parent iBridges
	TOG6017	96827	Nigéria	
	TOG6080	96832	Nigéria	
	TOG6144	96834	Nigéria	
	TOG6165	112575	Nigéria	
	TOG6195	96837	Guinée	
	TOG6197	96839	Guinée	

Tableau V-2 (Suite)

Espèce	Accession	Code IRGC	Pays	Observations
	TOG6198	96840	Guinée	
	TOG6202	86805	Guinée	
	TOG6203	86806	Guinée	
	TOG6206	96841	Zimbabwe	
	TOG6208	96842	Guinée	Parent iBridges
	TOG6210	96843	Sierra Leone	
	TOG6220	112577	Burkina Faso	
	TOG6221	96849	Nigéria	Parent iBridges
	TOG6231	96854	Mali	
	TOG6232	86810	Libéria	
	TOG6234	112569	Libéria	
	TOG6237	112611	Libéria	
	TOG6238	86812	Libéria	
	TOG7020	96888	Sierra Leone	
	TOG7173	96893	Sénégal	
	TOG7402	86826	Ghana	
	YG351/YS351		Guinée	
	YS309		Guinée	
	-	86761	Nigéria	
	-	104589	-	Parent iBridges
<i>O. barthii</i>	-	105613	Botswana	
	UB 5	86525	Cameroun	
	-	104072	Cameroun	
	-	104136	Cameroun	
	-	104138	Cameroun	
	-	104141	Cameroun	
	-	104142	Cameroun	
	-	104143	Cameroun	
	-	104296	Cameroun	
	-	105983	Cameroun	
	-	105985	Cameroun	
	-	105987	Cameroun	
	-	105988	Cameroun	
	-	105989	Cameroun	
	-	83799	Tchad	
	TB 41	86524	Tchad	
	-	103578	Tchad	
	-	103579	Tchad	
	-	103580	Tchad	
	-	103581	Tchad	
	-	103582	Tchad	
	-	104066	Tchad	
	-	104069	Tchad	
	-	104102	Tchad	
	-	104104	Tchad	
	-	104107	Tchad	
	-	104114	Tchad	
	-	104115	Tchad	
	-	104116	Tchad	

Tableau V-2. (Suite)

Espèce	Accession	Code IRGC	Pays	Observations
	-	104117	Tchad	
	-	104120	Tchad	
	-	104121	Tchad	
	-	104122	Tchad	
	-	104124	Tchad	
	-	105571	Tchad	
	-	105606	Tchad	
	-	105611	Tchad	
	-	106294	Tchad	
	-	106295	Tchad	
	-	106299	Tchad	
	-	106302	Tchad	
	-	103534	Mali	
	Ka Malo	104284	Mali	
	-	104061	Nigéria	
	-	104076	Nigéria	
	-	105609	Nigéria	
	-	103909	Tanzanie	
	-	86481	Zambie	
	-	106489	Zambie	
<i>O. sativa</i>	Nipponbare			<i>japonica</i>
	IR64			<i>indica</i>
	Caïapo			<i>japonica</i> tropical
	Curinga			<i>japonica</i> tropical
	Azucena			<i>japonica</i> tropical
	KDM105			<i>indica</i>
	Basmati			aromatique
<i>O. longistaminata</i>	LL109		Mali	
	AL302		-	
Hybrides	AL302 Obak		Madagascar	Hybride naturel entre <i>O. sativa</i> ssp. <i>indica</i> et <i>O. longistaminata</i>
	NIPLONG			Hybride entre Nipponbare et <i>O. longistaminata</i> (IRD-CIAT)
	85CA-1			Hybride BC ₄ F ₁ entre Caiapo (<i>japonica</i> tropical) et MG12 (IRD-CIAT)

Les noms des accessions en caractère gras correspondent aux parents des populations étudiées au cours de ce travail de thèse.

Nous avons utilisé l'ensemble de ces marqueurs pour évaluer les matériels suivants : (1) une collection expertisée, représentative de la diversité d'*O. glaberrima* (2) des variétés utilisées comme parents dans le projet iBridges, (3) plusieurs accessions d'*O. barthii* (4) des accessions-contrôle d'*O. sativa* et d'*O. longistaminata*, deux hybrides *O. sativa* x *O. longistaminata* (un hybride naturel, et un obtenu par croisement), et une lignée BC₄F₁ dérivée de la population CSSL Caiapo x MG12 (voir chapitre I et Tableau V-2).

V.1.4 Evaluations morphologiques des accessions

Dans le but de confirmer le caractère hybride de certaines des accessions classées parmi les *O. glaberrima*, des caractères morphologiques spécifiques à *O. glaberrima* et *O. sativa* furent évaluées à différents stades de développement des plantes et de quelques accessions-contrôle. Les caractères évalués comprenaient : la forme et la taille de la ligule, la coloration des talles, la coloration des apicules, la pubescence des glumelles, la présence de ramifications secondaires sur les panicules, le port des panicules, et la couleur du stigmate.

V.1.5 Evaluations de la fertilité des variétés du projet iBridges en croisement avec *O. sativa*

Le niveau de transmission des allèles d'*O. sativa* fut évalué de la manière suivante : les lignées parentales du projet iBridges furent croisées avec les variétés correspondantes d'*O. sativa* par les partenaires du projet (voir Chapitre I, Tableau I-10). Les plantes F₁ – utilisées comme femelle – obtenues furent rétrocroisées avec leur parent *O. sativa*. Les plantes BC₁F₁ obtenues furent évaluées génotypiquement avec le marqueur RM190 lié à *S₁*. Le pourcentage de transmission de l'allèle *O. sativa* pour ce marqueur à partir de la F₁ fut utilisé comme mesure approximative de la transmission de l'allèle *S₁*^s.

Par ailleurs, la fertilité des plantes BC₁F₁ fut estimée par comptage du nombre de grains vides et pleins sur trois panicules par plante.

**IR64 x TOG5681
Chr 6**

**Nipponbare
Chr 6**

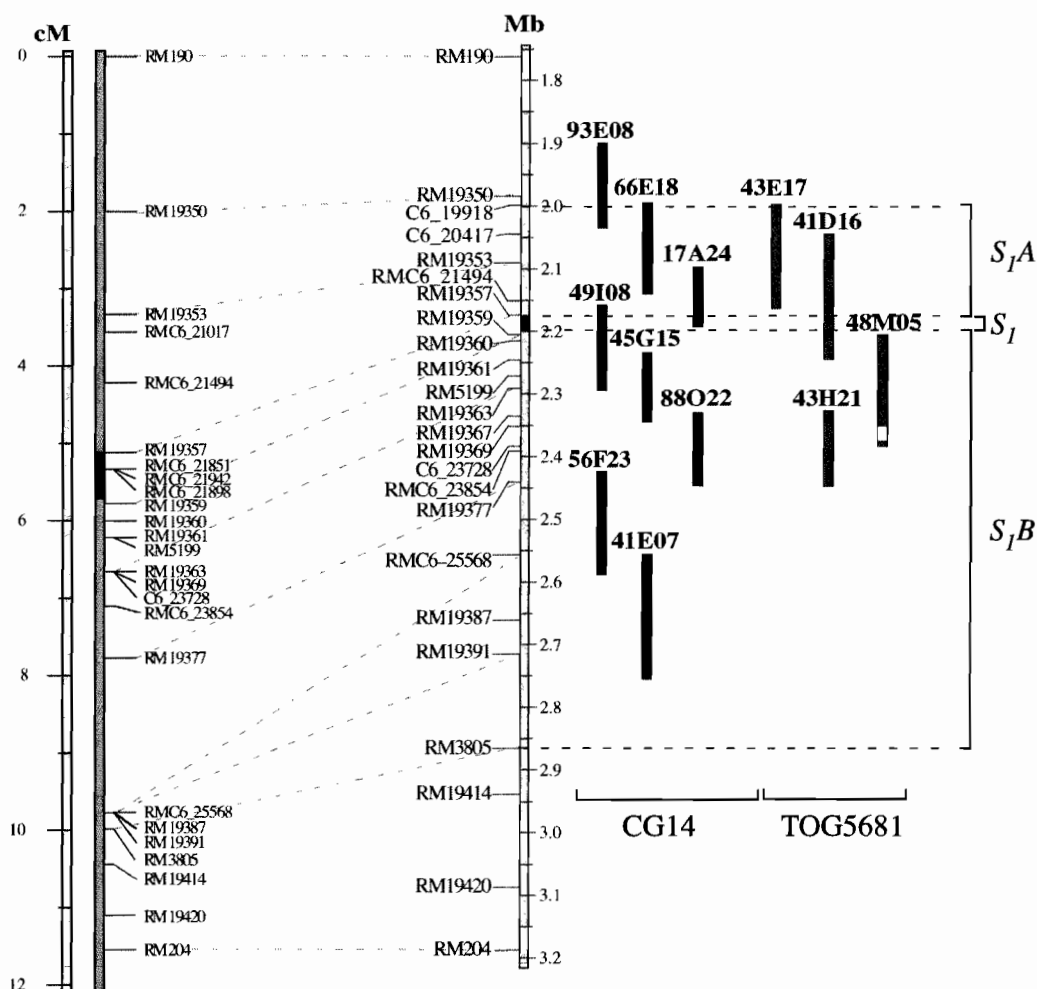


Figure V-1. Carte physique de la région du locus S_I chez *O. glaberrima* cv. TOG5681. Comparaison de la carte génétique de la région $S_I A$ - S_I - $S_I B$ pour le croisement IR64/TOG5681//IR64, avec la carte physique de la région des accessions Nipponbare (*O. sativa*), CG14 et TOG5681 (*O. glaberrima*). Certains marqueurs évalués ne sont pas montrés. La région représentée en blanc sur le clone TOG_48M05 correspond à un segment de l'inversion observée entre *O. sativa* et *O. glaberrima*.

V.2 Résultats

V.2.1 Carte physique de la région de S_I chez TOG5681

Afin d'évaluer le niveau de divergence de la région du locus S_I entre les accessions appartenant aux deux écotypes d'*O. glaberrima*, un criblage fut réalisé pour rechercher les clones couvrant la région de S_I dans la banque BAC de la variété irriguée TOG5681. 20 clones de différentes tailles furent identifiés comme appartenant à la région. Après évaluation avec les marqueurs microsatellites et Indel, utilisés pour la construction de la carte génétique du croisement IR64 x TOG5681, et détermination de leur taille moléculaire par CHEF, un MTP de quatre clones (TOG568_43E17 et TOG5681_41D16, TOG568_48M05 et TOG5681_43H21) fut déterminé sur la base de leur niveau de couverture de la région (voir Figure V-1). Par cette approche, nous avons obtenu une carte physique de la région qui contient la totalité du segment $S_I A$ et S_I , et une partie du segment $S_I - S_I B$.

V.2.2 Séquençage préliminaire de TOG5681 et comparaison avec CG14

Nous avons réalisé le séquençage des quatre clones identifiés, afin de réaliser une comparaison avec la séquence de l'accession CG14. Comme nous disposions déjà d'une séquence de qualité (celle de CG14), nous avons décidé de procéder par reséquençage en adoptant la méthode de pyroséquençage. Ainsi, deux clones sélectionnés dans un premier temps, TOG568_48M05 et TOG5681_43H21, furent traités par séquençage sur 454 FLX, avec une librairie PET. Les autres deux clones sélectionnés furent envoyés postérieurement pour un séquençage par 454 FLX Titanium.

Des séquences d'approximativement 162, 204, 135 et 150 kb furent obtenues respectivement pour les clones TOG5681_43E17, TOG5681_41D16, TOG568_48M05 et TOG5681_43H21. L'ordre et l'orientation des 15, 32, 12 et 24 contigs représentant respectivement chaque clone furent déterminés par comparaison graphique avec la séquence de CG14, et à l'aide de la librairie PET pour les clones 48M05 et 43H21. La séquence préliminaire de la région obtenue pour la variété TOG5681 consiste donc en un « superscaffold » d'approximativement 498 000 pb, contenant des interruptions dues aux incertitudes sur la séquence (voir Figure V-2 dans la page suivante).

Nous avons annoté et analysé cette séquence préliminaire afin de déterminer le niveau de divergence par rapport au segment orthologue de CG14, en prêtant une attention particulière aux polymorphismes de type InDel et SNP. En ce qui concerne les polymorphismes du type InDel, une comparaison graphique directe a permis de montrer la présence dans les régions intergéniques plusieurs Indel de séquence courte, et d'autres causés par

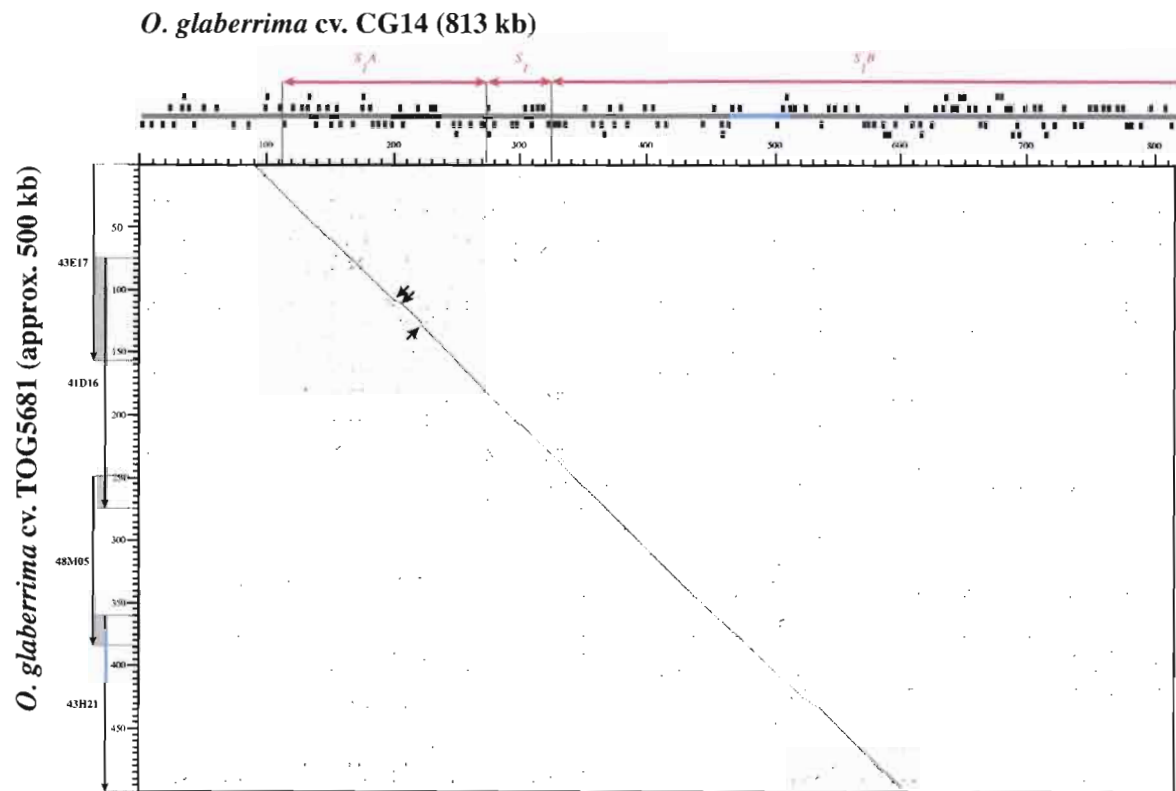


Figure V-2. Comparaison graphique entre les contigs de CG14 et TOG5681. Comparaison graphique entre les contigs des accessions *O. glaberrima* CG14 et TOG5681. Comparaison de type dot-plot de la séquence de CG14, avec la séquence préliminaire de TOG5681. Les gènes annotés chez CG14 sont représentés par des rectangles noirs sur l'axe horizontal, et les régions de probabilité d'emplacement des locus S_1A , S_1I et S_1B par les flèches rouges. Les clones de TOG5681 sont symbolisés par des traits pleins et les zones de recouvrement par des rectangles gris. La région de l'inversion entre *O. sativa* et *O. glaberrima* est représentée par une ligne bleue dans chaque accession. Les flèches noires indiquent les polymorphismes d'insertion d'éléments transposables. Celles qui pointent vers le bas indiquent une insertion chez CG14, tandis que celles qui pointent vers le haut indiquent une insertion chez TOG5681.

l'insertion différentielle de trois éléments transposables, deux chez CG14 et un chez TOG5681 (voir Figure V-2). En dehors de ces insertions, les deux séquences montrent une colinéarité presque parfaite. Au niveau des SNP, nous avons restreint la comparaison aux gènes dont la séquence chez TOG5681 ne présentent pas de rupture. Globalement, la fréquence de SNP observés dans les 47 séquences codantes comparées est très faible. La presque totalité des gènes identifiés (41) présentent une identité nucléotidique de 100%, à l'exception de deux gènes dont les SNP correspondent à des mutations synonymes (n'entraînant pas de changement dans la séquence protéique), et de quatre gènes montrant des changements non synonymes (voir Tableau V-3 dans les pages suivantes).

La divergence la plus forte est trouvée dans un gène à domaine LRR, TOG5681_43E17.20, avec 62 SNP. L'identité nucléotidique des gènes candidats pour *S₁A* (TOG5681_43E17.14) et *S₁* (TOG5681_41D16.10), ainsi que pour le reste des gènes du locus *S₁* est de 100%, à l'exception d'un des gènes (TOG5681_41D16.6) qui codent pour des protéines présentant une homologie avec les nodulines.

En résumé, la comparaison entre les séquences disponibles de deux accessions représentant les deux écotypes d'*O. glaberrima*, irrigué et pluvial stricte, a montré une très forte conservation nucléotidique entre les deux accessions, à l'exception toutefois de l'insertion d'éléments transposables et de plusieurs SNP dans les régions codantes de certains gènes.

V.2.3 Evaluation de la région génomique autour de *S₁* dans une collection

d'*O. glaberrima* et d'*O. barthii*

La forte conservation de séquence de la région *S₁A* - *S₁* - *S₁B* entre les deux accessions TOG5681 et CG14 nous a conduits à nous intéresser à la structure allélique de la région dans d'autres accessions d'*O. glaberrima* et de son ancêtre *O. barthii*. Une analyse des insertions d'éléments transposables fut entreprise à partir d'une collection expertisée de variétés *O. barthii* et *O. glaberrima*, et chez les parents des croisements du projet iBridges.

Nous observons tout d'abord que la structure génomique est globalement identique entre toutes les variétés *O. glaberrima* et *O. barthii*, avec peu de polymorphisme entre les deux espèces (voir Figure annexe, pour une représentation graphique de la structure de population déterminée à partir des données de marqueurs d'insertion d'éléments transposables).

Par ailleurs, un résultat important est que six des accessions cataloguées traditionnellement comme *O. glaberrima* présentent une structure allélique haplotypique proche de celle de certains *O. sativa* (les variétés Basmati et KDM105), et ce sur toute la longueur du contig (voir Figure V-3). Des résultats antérieurs avaient déjà montré que deux de ces accessions (DCKono et PaDCKono) pouvaient être considérées comme des « *admixture* » (Semon *et al.*, 2005) Parmi les quatre accessions restantes, une (YS309) a même parfois été cataloguée

Tableau V-3. Pourcentage d'identité des séquences protéiques et nucléotidiques entre les gènes de CG14 et TOG5681.

Gène de TOG5681 (TOG5681_)	Gène orthologue chez CG14 (OG-BBa00)	Fonction putative	Identité protéique (%)	Identité nucléotidique (%)	Nombre de SNP	Longueur des CDS alignés (pb)
43E17.1p	93E08.14	Putative Armadillo protein	/	/	/	/
43E17.2*	93E08.15	Putative protein	/	/	/	/
43E17.3	93E08.16	Putative Auxin responsive protein	100	97,7	1	358
43E17.4	93E08.17	Putative protein	100	100	0	1214
43E17.6*	93E08.18	Putative Ribosome Binding factor	/	/	/	/
43E17.7	93E08.19	Putative Epimerase	99,7	99,9	1	1126
43E17.8	93E08.20	Putative transcription factor	100	100	0	1081
43E17.9*	93E08.21	Putative methionine sulfoxide reductase	/	/	/	/
43E17.10	93E08.22	Putative protein	100	100	0	1849
43E17.11*	66E18.1	Putative CAF1 ribonuclease	/	/	/	/
43E17.12*	66E18.2	Putative protein	/	/	/	/
43E17.13	66E18.3	Putative protein	100	100	0	1051
43E17.14	66E18.4	Putative F-box protein	100	100	0	1560
43E17.15*	66E18.5	Putative protein	/	/	/	/
43E17.16	66E18.6	Putative HAD phosphatase	100	100	0	880
43E17.17	66E18.7	Putative proteasome subunit	100	100	0	739
43E17.18	66E18.8	Putative LRR protéine	100	100	0	2814
43E17.19	17A24.1	Putative methyltransferase	99,7	99,7	1	1017
43E17.20	17A24.2	Putative LRR protein	82,9	88	62	2953
43E17.21*	17A24.3	Putative LRR protein	/	/	/	/
43E17.22*	17A24.4	Putative protein	/	/	/	/
43E17.23	17A24.5	Hypothetical protein	100	99,5	1	195
43E17.24*	17A24.6	Putative protein	/	/	/	/
43E17.25	17A24.7	Putative serine/Threonine protein kinase	100	100	0	2419
43E17.26	17A24.8	Putative protein	100	100	0	205
41D16.1*	17A24.9	Putative protein	/	/	/	/
41D16.2*	17A24.10	Putative NADH pyrophosphatase protein	/	/	/	/
41D16.3	49I08.2	Putative Zinc finger protein	100	100	0	434
41D16.4	49I08.3	Putative ENOD93 protein	/	/	/	/
41D16.5	49I08.4	Hypothetical protein	100	100	0	642
41D16.6	49I08.5	Putative ENOD93 protein	99,1	99,7	1	351
41D16.7	49I08.6	Putative ENOD93 protein	100	100	0	348
41D16.8	49I08.7	Hypothetical protein	100	100	0	786
41D16.9	49I08.8	Putative protein	100	100	0	2073
41D16.91*	49I08.85*	Hypothetical protein	/	/	/	/
41D16.92*	49I08.9*	Hypothetical protein	/	/	/	/
41D16.95*	49I08.10*	Hypothetical protein	/	/	/	/
41D16.10	49I08.11	Putative F-box protein	100	100	0	1542
41D16.11	49I08.12	Putative ENOD93 protein	100	100	0	348
41D16.12	49I08.13	Putative ENOD93 protein	100	100	0	348
41D16.13	49I08.14	Putative ENOD93 protein	100	100	0	351
41D16.14	49I08.15	Putative ENOD93 protein	100	100	0	348
41D16.15	49I08.16	Putative Serine/Threonine protein kinases	100	100	0	2223
41D16.16	49I08.17	Putative Early flowering protein	/	/	/	/
48M05.1*	49I08.18	Putative Serine/Threonine protein kinases	/	/	/	/

Tableau V-3. (Suite)

Gène de TOG5681 (TOG5681_)	Gène orthologue chez CG14 (OG-BBa00)	Fonction putative	Identité protéique (%)	Identité nucléotidique (%)	Nombre de SNP	Longueur des CDS alignés (pb)
48M05.2	49I08.19	Putative cytochrome oxidase	100	100	0	465
48M05.3	49I08.20	Putative methyltransferase protein	100	100	0	1143
48M05.4	49I08.21	Putative 1-deoxy-D-xylulose-5-phosphate synthase	100	100	0	2169
48M05.5	49I08.22	Putative Superoxide dismutase	100	100	0	768
48M05.6	49I08.23	Putative protein	100	100	0	108
48M05.7	49I08.24	Putative Myristoyl-acyl carrier protein thioesterase	100	100	0	1264
48M05.8*	49I08.25	Putative PPR protein	/	/	/	/
48M05.9	49I08.26	Putative protein	100	100	0	1296
48M05.10*	49I08.27	Putative sulfate transporter	/	/	/	/
48M05.11*	45G15.1	Putative Coatomer protein	/	/	/	/
48M05.12	45G15.2	Putative BRCA1 protein	100	100	0	1042
48M05.13	88O22.1	Putative Zinc finger protein	100	100	0	573
48M05.14*	88O22.2	Putative pectate lyase protein	/	/	/	/
48M05.15	88O22.3	Putative pectate lyase protein	100	100	0	1351
48M05.16*	88O22.4	Putative GTP-binding protein	/	/	/	/
48M05.17*	88O22.5	Putative carboxypeptidase	/	/	/	/
43H21.1	88O22.6	Hypothetical protein	100	100	0	657
43H21.2*	88O22.7	Putative pectate lyase protein	/	/	/	/
43H21.3*	88O22.8	Putative Transferase	/	/	/	/
43H21.4*	88O22.9	Putative Transferase	/	/	/	/
43H21.5*	88O22.10	Putative Transferase	/	/	/	/
43H21.6*	88O22.11	Putative Transferase	/	/	/	/
43H21.65*	88O22.12	Putative AP2 proyein	/	/	/	/
43H21.7	88O22.13	Putative Whirly transcription factor protein	100	100	0	819
43H21.8	88O22.14	Putative protein	100	100	0	3079
43H21.9	88O22.15	Putative protein	100	100	0	316
43H21.10*	88O22.16	Putative WRKY protein	/	/	/	/
43H21.11*	88O22.17	Putative protein	/	/	/	/
43H21.12	88O22.18	Putative Iron-sulphur protein	100	100	0	561
43H21.13	88O22.19	Hypothetical protein	100	100	0	369
43H21.14	56F23.1	Putative protein	100	100	0	348
43H21.15	56F23.2	Putative protein	100	100	0	276
43H21.16	56F23.3	Putative protein	100	100	0	357
43H21.17*	56F23.4	Putative protein	/	/	/	/

P gène partiel, * gène à séquence incomplète chez TOG5681, non comparé.

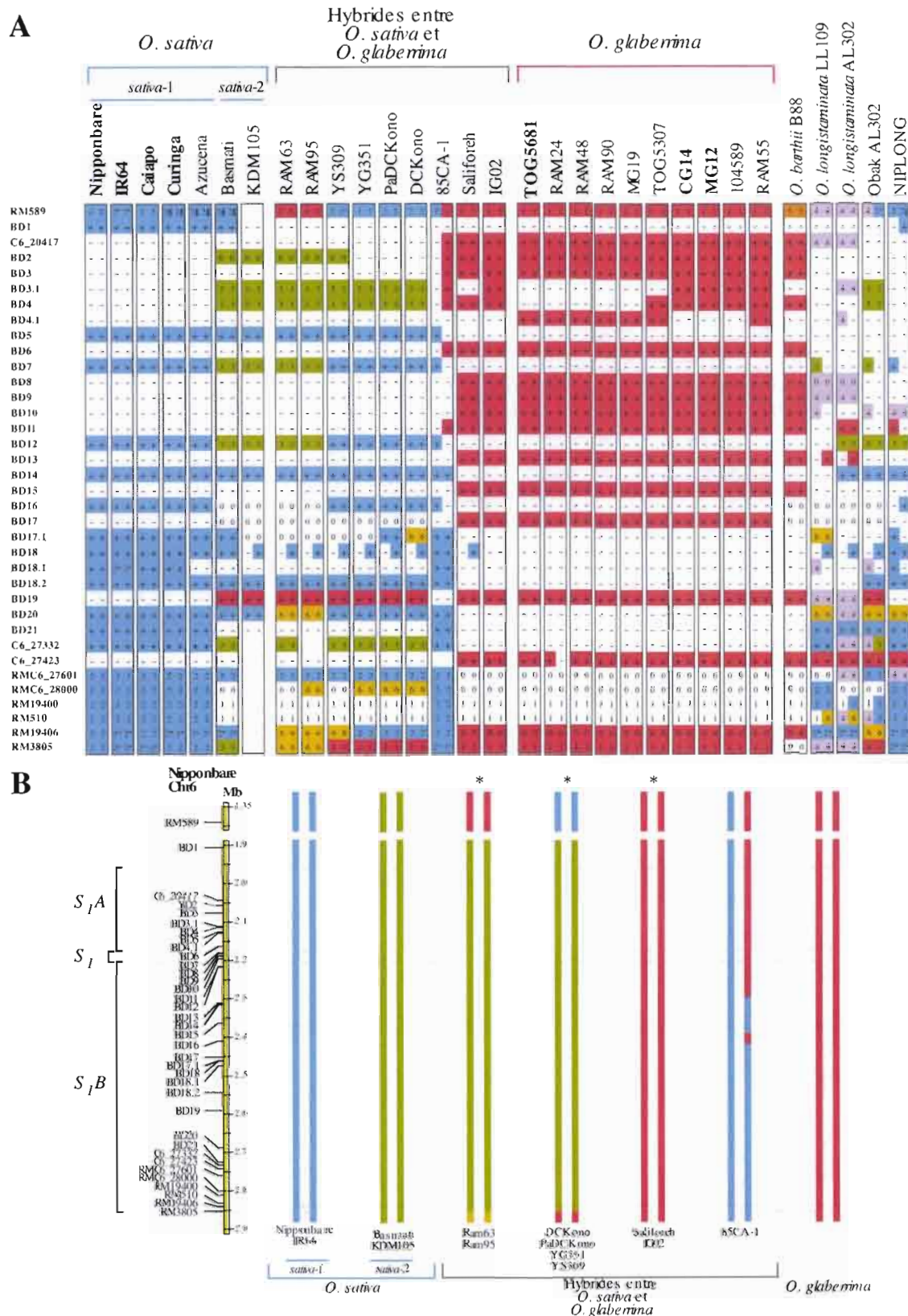


Figure V-3. Evaluation de la structure de la région S_{IA} - S_I - S_{IB} dans une collection d'accèsions *O. glaberrima*. A. Représentation graphique des génotypes pour des marqueurs d'insertion d'éléments transposables et microsatellites. Les deux allèles de chaque locus sont représentés.

comme *O. sativa* (Pham, 1992).

Comme la région étudiée, qui correspond à l'haplotype *O. sativa* pour ces six accessions, ne recouvre pas la totalité du segment qui contient les locus putatifs S_1A et S_1B , nous avons réalisé une analyse complémentaire à l'aide des marqueurs microsatellites situés aux bordures du contig. On observe que l'introggression *sativa* a pour limite supérieure le marqueur RM589 pour deux des accessions, tandis que les autres portent les allèles d'*O. sativa* au-delà de RM589. Le marqueur RM3805 définit la limite inférieure du segment introgressé pour quatre des accessions (YS309, YG351, PaDCKono, DCKono), comme dans le cas des populations BC_1F_1 que nous avons étudiées. Pour les deux autres accessions (RAM63 et RAM95), un allèle distinct des allèles portés par les contrôles *O. sativa* semble limiter la région pour le même marqueur. Il peut être intéressant de noter que cet allèle est très proche en taille moléculaire de l'allèle porté par les variétés *O. glaberrima* (voir Figure V-3).

Dans le but de confirmer la nature hybride des six accessions cataloguées comme *O. glaberrima* et qui présentent un haplotype *O. sativa* dans la région des locus S_1A , S_1 et S_1B , une évaluation pour des caractères morphologiques distinctifs entre *O. glaberrima* et *O. sativa* fut entreprise à différents stades de développement de la plante. Les résultats (voir Tableau V-4 dans la page suivante) montrent que ces accessions (RAM63, RAM95, YS309, YG351, PaDCKono et DCKono) ainsi que deux autres (Saliforeh et IG02) présentent des caractères morphologiques intermédiaires entre *O. glaberrima* et *O. sativa*, confirmant leur nature hybride.

En résumé, nos résultats indiquent que les accessions d'*O. glaberrima* et d'*O. barthii* analysées présentent un haplotype commun dans la région des trois locus de stérilité, et que certaines accessions qui s'avèrent être des hybrides naturels entre *O. sativa* et *O. glaberrima* peuvent posséder l'haplotype d'une espèce ou de l'autre sur toute la longueur de cette région.

Figure V-3. (Suite). Les allèles d'insertion (+) caractéristiques d'*O. sativa* et d'*O. glaberrima* sont colorés en bleu et rouge, respectivement. Les non-insertions sont représentés par le signe (-). Les allèles exclusifs aux accessions *O. sativa* Basmati et KDM105 (vert), à *O. longistaminata* (mauve), à *O. barthii* (orange), et ceux qui ne correspondent à aucune des catégories (jaune) sont également indiqués. Codes numériques utilisés pour les allèles des microsatellites et certains polymorphismes d'insertion d'ET : (0) non-amplification, (1) *O. glaberrima* cv. CG14, (2) Nipponbare, (2* et 2§) *O. sativa* différent de Nipponbare, (3) *O. sativa* Basmati et KDM105, (4) *O. longistaminata*, (5) *O. barthii*, (6) autres allèles ne correspondant à aucune des catégories décrites, (allèle en blanc) donnée manquante. **(b)** Représentation graphique de la correspondance de la région des locus S_1A , S_1 et S_1B chez les différentes accessions identifiées comme hybrides naturels (marquées par une *). Les couleurs bleu et vert représentent les allèles *O. sativa*, la couleur rouge les allèles *O. glaberrima*.

Tableau V-4. Analyse préliminaire des caractères morphologiques des formes intermédiaires entre *O. glaberrima* et *O. sativa*.

Accession	Forme et Taille de la ligule	Coloration du tallo	Apicules	Pilosité des glumelles	Ramification de la panicule	Forme de la panicule	Coloration du stigma	Coloration de la graine	Interpretation
CG14	gla	gla	gla	gla	gla	gla	gla	Marron	<i>O. glaberrima</i>
DcKono	gla	gla	-	gla	-	-	-	-	-
IG02	gla	gla	gla	gla	gla	gla	sat	Marron clair	intermediaire
MG19	gla	gla	gla	sat	gla	-	gla	Marron	intermediaire
PaDCKono	sat	gla	sat	gla	sat	sat	gla	Marron clair	intermediaire
Ram111	gla	gla	gla	sat	gla	-	gla	-	-
Ram24	gla	gla	gla	sat	gla	gla	gla	Marron clair	intermediaire
Ram48	gla	gla	gla	sat	gla	gla	gla	Marron	intermediaire
Ram55	gla	gla	gla	gla	gla	-	gla	Marron	-
Ram63	sat	gla	gla	sat	sat	sat	gla	Marron clair	intermediaire
Ram95	sat	gla	gla	sat	sat	sat	gla	Jaune	intermediaire
Saliforeh	gla	gla	gla	gla	gla	gla	gla	Marron	<i>O. glaberrima</i>
TOG 5681	gla	gla	gla	gla	gla	gla	gla	-	<i>O. glaberrima</i>
TOG 5307	gla	gla	gla	gla	-	-	-	-	-
YG351	sat	gla	sat	gla	sat	-	sat	Jaune	intermediaire
YS309	gla	gla	gla	gla	sat	sat	gla	Marron clair	intermediaire
IR64	sat	sat	sat	sat	sat	sat	sat	Jaune	<i>O. sativa</i>

Les caractères furent évalués comme : Forme et Taille de la ligule Courte (gla) / Longue (sat) ; Tallo coloré (gla) / non coloré (sat) ; Apicules coloré (gla) / non coloré (sat) ; Pilosité des glumelles - non poilu (gla) / poilu (sat) ; Ramification de la panicule primaire (gla) / secondaire (sat) ; Forme de la panicule droite (gla) / curbée (sat) ; Coloration du stigma Purpre (gla) / blanc (sat).

V.2.4 Sélection allélique et fertilité en croisement des variétés hybrides naturels

La constatation que certaines accessions *O. glaberrima* correspondent en fait à des hybrides naturels interspécifiques et présentent une dichotomie pour leur haplotype autour du locus S_I nous a conduit à examiner plus en détail les différences au niveau de la transmission des allèles après croisement avec *O. sativa*, en fonction de l'haplotype de l'accession. Nous avons observé le niveau de transmission de l'allèle S_I^s dans les plantes F_1 issues des croisements interspécifiques du projet iBridges. Nous avons par ailleurs examiné le niveau de fertilité des plantes BC_1F_1 résultantes. Ces données préliminaires représentent une compilation des expérimentations menées par le CIAT (M. Lorieux), l'ADRAO (M.N. Ndjondjop et l'IRD (A. Ghesquière). Le résumé de ces données est présenté dans le Tableau V-5, qui se trouve dans la page suivante.

On constate tout d'abord, en réunissant l'ensemble des plantes évaluées et qui dérivent de rétrocroisements sur les trois variétés d'*O. sativa*, que les valeurs moyennes de transmission de l'allèle S_I^s sont très semblables entre les combinaisons, l'indice k variant entre 0,945 et 0,959. Les niveaux de transmission de S_I^s pour les croisements impliquant des variétés *O. glaberrima* « vraies » oscillent entre $k = 0,887$ et $k = 1$, avec une moyenne de 0,951. En revanche, pour deux des variétés identifiées comme hybrides naturels (RAM63 et RAM95), on observe une transmission de S_I^s très élevée (respectivement 0,620 et 0,555). Ce résultat est à rapprocher du fait que ces deux accessions possèdent un haplotype *O. sativa* tout au long de la région $S_I A - S_I - S_I B$. Il est étonnant de constater qu'une autre variété portant l'haplotype *O. sativa* (DCKono) ne montre pas un niveau de transmission supérieur à celui des autres variétés *O. glaberrima*, mais le faible nombre d'individus évalués pour cette accession ne permet pas de tirer de conclusion claire. On observe enfin que les valeurs de transmission de l'allèle S_I^s des deux hybrides naturels qui possèdent l'haplotype *O. glaberrima*, IG02 et Saliforeh, ne sont pas significativement différents de celles des autres variétés d'*O. glaberrima*.

Au niveau de la fertilité des plantes BC_1F_1 , on remarque que le pourcentage de plantes fertiles parmi les homozygotes pour l'allèle *O. sativa* au marqueur RM190 est beaucoup plus élevé que celui des plantes hétérozygotes. Cette différence est encore plus claire pour les populations issues de croisements avec les hybrides naturels, avec un pourcentage de plantes fertiles – et en principe homozygotes – atteignant 98 %.

Si l'on prend en compte l'ensemble des données de transmission des allèles *O. sativa*, on peut conclure que la fertilité en croisement avec *O. sativa* dépend en grande partie de l'haplotype du parent *O. glaberrima* autour du locus S_I . Les variétés présentant une introgression d'*O. sativa* à ce locus présentent une fertilité en croisement plus élevée associée à une

Tableau V-5. Données de transmission de l'allèle S_I^s et de fertilité à partir des croisements du projet iBridges.

Parent	Nombre de combinaisons interspécifiques	Nombre de BC ₁ F ₁ évaluées	Transmission approximative de S_I^s (RM190)	k	% de BC ₁ F ₁ fertiles (S_I^s/S_I^s)	% de BC ₁ F ₁ fertiles (S_I^s/S_I^s)
<i>O. sativa</i>						
IR64	12	831	46	0,945	54	2
Curinga	11	797	33	0,959	45	5
WAB165	11	669	33	0,951	15	3
<i>O. glaberrima</i>						
104589	1	41	4	0,902	75	27
CG14	3	332	5	0,985	0	2
CG17	2	198	6	0,970	0	14
MG12	3	326	21	0,936	81	0
MG19	2	143	11	0,923	9	6
RAM3	1	51	0	1,000	0	0
RAM24	2	67	1	0,985	0	0
RAM48	1	41	2	0,951	0	23
RAM55	2	70	5	0,929	0	6
RAM90	2	125	10	0,920	1	2
RAM111	2	116	12	0,897	17	2
TOG5890	1	54	0	1,000	-	-
TOG5666	1	25	2	0,920	0	4
TOG5672	2	194	7	0,964	43	5
TOG5674	1	58	3	0,948	0	0
TOG5681	2	260	8	0,969	0	0
TOG5882	-	5	0	1,000	-	-
TOG5887	2	53	6	0,887	33	2
TOG6211	1	133	2	0,985	0	0
TOG6308	1	59	6	0,898	33	19
TOG7106	2	38	4	0,895	75	9
Total (non-introgressées)	34	2297	112	0,951	40	3
Saliforeh	1	67	4	0,940	50	6
IG02	3	121	10	0,917	40	4
DC Kono	1	30	2	0,933	0	14
RAM63	2	258	98	0,620	98	53
RAM95	2	191	85	0,555	75	42
TOTAL (introgressées)	10	878	233	0,735	72	23

transmission nettement supérieure de l'allèle S_I^s , ce qui représente un élément de confirmation important de l'hypothèse que cette région génomique est le déterminant clé de la fertilité des hybrides interspécifiques *O. sativa* x *O. glaberrima*.

V.3 Discussion

Nous avons vu que la forte barrière reproductive qui sépare les deux espèces de riz cultivé semble être indépendante des accessions d'*O. glaberrima* et *O. sativa* étudiées (Sano, 1990; Pham, 1992; Doi et al., 1998b; Aluko et al., 2004; Koide et al., 2008c; et nos propres résultats). Nous nous attendons donc à ce que la structure allélique du locus $S_1A - S_1 - S_1B$ soit fortement conservée entre les différentes variétés d'*O. glaberrima*. En outre, l'observation d'une distorsion de ségrégation dans la région de S_1 sur une carte génétique *O. sativa* x *O. barthii* (M. Lorieux, communication personnelle), et la bonne fertilité des hybrides *O. glaberrima* x *O. barthii* (Bezançon, 1994; nos propres données, non présentées ici) suggèrent que l'haplotype *O. barthii* pour $S_1A - S_1 - S_1B$ doit être très proche de celui d'*O. glaberrima*. Afin de tester ces hypothèses, deux approches furent utilisées. La première consiste à séquencer le locus S_1 dans une deuxième accession *O. glaberrima*, la variété d'écologie irriguée TOG5681 ; la deuxième est l'analyse de la configuration allélique de $S_1A - S_1 - S_1B$ dans une collection d'accessions d'*O. glaberrima* et *O. barthii*.

En accord avec la faible diversité génétique observée chez *O. glaberrima* (Bezançon, 1994), nous avons pu démontrer la forte identité nucléotidique entre les deux variétés séquencées d'*O. glaberrima* pour le locus qui nous intéresse. Des Indel de courte séquence, et un faible nombre de SNP dans les régions codantes constituent les principales variations examinées. Deux des gènes qui présentent des SNP impliquant des changements d'acide aminé, dont celui avec le nombre le plus élevé de SNP (43E17.20), sont situés à proximité du gène *spd6* responsable de la différence de taille de la panicule entre *O. sativa* et son ancêtre (Shan et al., 2009). À partir de leurs résultats, Shan et al. concluent que le plus fort candidat pour *spd6* correspond à une Methyltransferase (43E17.19), car les trois autres gènes de la région ont un domaine LRR homologue à celui du gène *HcrVf3* qui confère la résistance à la maladie cryptogamique de la pomme causée par *Venturia inaequalis* (Vinatzer et al., 2001). On peut faire l'hypothèse que 43E17.20 correspond à un gène de résistance, et que la forte divergence entre les deux variétés dans sa partie codante est due à une évolution rapide caractéristique de ce type de gène permettant la reconnaissance des éliciteurs en mutation constante.

Une différence notable entre les séquences des deux espèces cultivées est l'insertion d'éléments transposables. Ces polymorphismes représentent une opportunité pour l'évaluation de la région, puisqu'au contraire des marqueurs microsatellites, les polymorphismes dus à

l'insertion de transposons, MITE et Retrotransposons du type LTR sont peu ou pas réversibles; une fois insérés, les éléments peuvent subir différents réarrangements, mais rarement être excisés sans laisser de trace. Nous avons donc utilisé des marqueurs d'insertion d'éléments transposables, pour déterminer les configurations alléliques autour de S_I d'une collection expertisée d'accessions d'*O. glaberrima* et *O. barthii*, ainsi que d'accessions du projet iBridges. L'approche a montré une divergence réduite entre *O. glaberrima* et *O. barthii* dans la région des trois locus, mais nous a permis de mettre en évidence une dichotomie dans cette région pour les hybrides interspécifiques naturels. Cette dichotomie se traduit en un haplotype *O. sativa* ou *O. glaberrima* sans jamais observer d'haplotype recombiné entre les trois locus qui interagissent en épistasie pour causer la stérilité partielle femelle des hybrides. L'existence de ces hybrides naturels s'explique évidemment par le contact étroit entre les deux espèces de riz cultivé qui a eu lieu depuis l'introduction d'*O. sativa* sur le continent Africain. Ce contact a favorisé, probablement à plusieurs reprises, des échanges génétiques entre *O. sativa* et *O. glaberrima*, par des croisements et rétrocroisements aléatoires des hybrides vers l'une ou l'autre espèce. Un nombre indéterminé d'événements d'hybridation ont donc, grâce à la sélection naturelle ou anthropique après plus de 500 ans de sympatrie, abouti à l'émergence de ces lignées fertiles, adaptées aux conditions environnementales locales, et possédant l'haplotype autour du locus S_I de leur « parent récurrent ». On entend ici par « parent récurrent » une ou plusieurs accessions représentant le fonds génétique majoritaire de l'hybride. Il n'est probablement pas fortuit que le locus dont dépend la survie des gamètes des hybrides F_1 artificiels soit conservé entièrement (sans recombinaison) dans les *admixture* naturels sous forme d'haplotype *O. sativa* ou *O. glaberrima*; la dichotomie de cette région chromosomique représente une évidence évolutive qui vient conforter le modèle génétique développé pour expliquer la semi stérilité femelle des hybrides F_1 interspécifiques.

Un autre aspect intéressant de nos résultats est qu'ils montrent l'existence de deux haplotypes distincts à l'intérieur d'*O. sativa*. L'haplotype *sativa* 2, que partagent les variétés Basmati (groupe des *basmati*, ou « *aromatics* », apparenté aux *japonica*) et KDM105 (*indica*) est celui que l'on retrouve dans les hybrides naturels. Deux des hybrides naturels (RAM63 et RAM95) portent cet haplotype *sativa* 2. D'autres (YS309, YG351, PaDCKono et DCKono) présentent une mosaïque allélique entre les haplotypes *sativa* 2 et *sativa* 1, qui est celui des variétés d'*O. sativa* représentatives des groupes *indica*, *japonica* et *japonica* tropical. Ces résultats amènent à se demander si les différentes configurations alléliques des *admixture* sont le résultat d'un contact plus prolongé d'*O. glaberrima* avec des accessions *O. sativa* porteuses de l'haplotype *sativa* 2, introduites plus précocement en Afrique, ou sont plutôt le fruit d'une meilleure compatibilité de l'haplotype *sativa* 2 avec *O. glaberrima*. La première hypothèse

semble la plus probable, et se trouve peut-être renforcée par l'observation de l'haplotype de l'hybride naturel entre *O. sativa* et *O. longistaminata* (Obak AL302), très similaire à *sativa 2* (voir Figure V-3). Les donneurs originels de l'haplotype *sativa 2* restent à identifier par une étude comprenant un plus grand nombre d'accessions. On pourrait alors tester l'hypothèse d'une meilleure compatibilité en croisement de ces donneurs avec *O. glaberrima*. Plus directement, on pourra tester la fertilité des hybrides entre les accessions à haplotype *sativa 2* (Basmati ou KDM105) avec *O. glaberrima*.

Le caractère hybride de certaines accessions au niveau du « système S_I » (i.e., le gène S_I et les locus putatifs S_{IA} et S_{IB}) pose bien entendu la question de la fertilité en croisement de ces matériels, en relation avec l'haplotype qu'elles portent. Sous l'hypothèse que l'haplotype S_I d'une accession *O. glaberrima* déterminerait la fertilité de son hybride avec les deux espèces cultivées, on s'attend à ce que les accessions portant un haplotype *sativa 2* soient plus fertiles en croisement avec *O. sativa* qu'avec *O. glaberrima*.

Afin de tester cette hypothèse, nous avons examiné les différences de pourcentage de transmission des allèles *O. sativa* et la fertilité des hybrides naturels, selon le type d'haplotype, à partir des données obtenues par les partenaires du projet iBridges. Plusieurs conclusions peuvent être tirées de ces analyses. (1) Les valeurs des moyennes de transmission de l'allèle S_I^s sont comparables entre variétés *O. sativa*, ce qui indique que les différences ne dépendent pas exclusivement du fonds génétique d'*O. sativa*, comme suggéré préalablement (Koide *et al.*, 2008c), mais plutôt du jeu d'allèles parentaux mis en interaction par le croisement, comme nous l'avons suggéré au Chapitre III. (2) La transmission très préférentielle de S_I^s pour deux des accessions hybrides naturels qui possèdent l'haplotype *sativa 2* est corrélée à une meilleure fertilité de ces accessions en croisement avec *O. sativa*. Le système S_I à lui seul contrôle donc une large part de la fertilité des plantes F_1 . (3) Le niveau de transmission de S_I^s chez les hybrides naturels qui possèdent l'haplotype *O. glaberrima* n'est pas significativement différent de celui observé chez les variétés « vraies » d'*O. glaberrima*, ce qui représente une confirmation de la subsistance de la barrière.

En conclusion, la fertilité en croisement avec *O. sativa* d'une accession *admixture* dépendra principalement – mais pas exclusivement – de l'haplotype que la variété hybride présente dans la région autour du locus S_I .

Figure V-4. Structure populationnelle des accessions d'*O. glaberrima* et *O. barthii* trouvé par l'évaluation des marqueurs d'éléments transposables. Dendrogramme construit selon la méthode du Neighbor-Joining, calculé à partir de la matrice de dissimilarités des marqueurs d'insertion d'éléments transposables de la région du locus *S_I*. Les codes de couleurs sont les mêmes que pour la Figure V-3.

VI

DISCUSSION

VI.1 Récapitulation des principaux résultats

On peut classer les résultats obtenus au cours du présent travail de thèse, par l'implémentation de différentes approches de génétique et de génomique comparative, en deux types : ceux qui apportent une contribution à l'amélioration génétique du riz, à partir des outils développés ou améliorés, et ceux qui apportent une information ponctuelle sur la barrière de reproduction entre les deux espèces de riz cultivé, et par extension, contribuent à une meilleure connaissance de la génétique de l'isolement reproductif chez les plantes.

Dans le premier groupe de résultats se trouve le développement d'une carte génétique universelle du riz (UCGM), et d'une méthodologie pour la cartographie de gènes de stérilité. La carte universelle du riz s'est avérée être un outil fiable pour la construction et la comparaison de cartes génétiques de différents croisements interspécifiques chez le riz. En outre, son utilisation pour le génotypage d'une population interspécifique (IR64/TOG5681//IR64) a mis en évidence la co-localisation des locus de stérilité décrits entre les deux espèces de riz cultivé, et de régions en distorsion de ségrégation sur la carte génétique de ce croisement interspécifique. Cette association nous a permis de définir une méthodologie pour la cartographie de gènes de stérilité, basée sur l'effet de l'élimination gamétique sur la ségrégation des marqueurs qui leur sont génétiquement liés.

C'est avec l'application de cette méthode de cartographie génétique que les premiers résultats informatifs sur la stérilité hybride entre *O. sativa* et *O. glaberrima* ont été acquis. L'application de la méthode sur quatre populations interspécifiques (BC_1F_1) a permis la cartographie fine du locus S_I , ainsi que la mise en évidence de deux autres facteurs qui interagissent probablement en épistasie avec S_I dans le contrôle du phénotype de stérilité femelle des hybrides, et dénommés $S_I A$ et $S_I B$. L'analyse détaillée du génotype des gamètes qui survivent à l'élimination sélective, et les informations disponibles dans la littérature, nous ont permis de développer un modèle génétique basé sur des interactions épistatiques entre les allèles incompatibles, pour expliquer à la fois le taux d'élimination gamétique et la distorsion de la ségrégation des marqueurs moléculaires liés à S_I . Par ailleurs, nous avons aussi pu déduire la localisation probable de $S_I A$ et $S_I B$ à partir de nos données génotypiques.

A travers l'approche de génomique comparative, il a été possible de construire une carte physique de la région chez CG14 (accession d'écologie pluviale stricte d'*O. glaberrima*) représentée par les huit clones BAC du chemin minimal de recouvrement, d'obtenir la séquence de ces clones, de l'annoter et de la comparer avec la région orthologue disponible d'*O. sativa*. Cette comparaison nous a permis de mettre en évidence l'existence d'une bonne

colinéarité au niveau du contenu en gènes, avec des exceptions dues à des réarrangements (inversion, duplication, insertion/délétion), à la présence de différences structurales significatives dans les régions intergéniques, et à une évolution divergente au niveau des séquences codantes d'un certain nombre de gènes. Ensemble, ces analyses nous ont permis de proposer un modèle d'évolution pour une région impliquée dans l'isolement reproductif et nous ont fourni des indices sur les gènes candidats possibles pour deux des locus (S_I et $S_I A$).

En suivant une approche similaire de génomique comparative, la carte physique autour de S_I fut établie pour une deuxième accession d'*O. glaberrima*, l'accession d'écologie irriguée TOG5681. Le pyroséquençage des quatre clones BAC qui constituent cette carte a permis de mettre en évidence une très forte similarité entre les séquences codantes de CG14 et TOG5681.

Enfin, nous avons réalisé un examen de la structure de la région des trois locus $S_I A$ - S_I - $S_I B$ dans une collection d'*O. glaberrima* et de son ancêtre *O. barthii*. Cette analyse a mis en évidence une forte conservation de la région entre ces deux espèces africaines, ainsi que l'existence de deux types d'hybrides interspécifiques naturels, différenciés en fonction de la configuration haplotypique (type *sativa* ou *glaberrima*) qu'ils possèdent dans cette région. Confirmant notre modèle génétique, la fertilité en croisement avec *O. sativa* de ces lignées hybrides s'est avérée plus élevée pour celles qui possèdent l'haplotype *sativa*.

VI.2 Construction, analyse et critique du modèle génétique proposé

Sur la base des résultats générés par les approches génétiques utilisées dans le cadre de nos travaux, et de ceux disponibles dans la littérature, nous avons construit un modèle génétique explicatif du phénomène de stérilité interspécifique causée par l'élimination allélique des gamètes femelles produits par les hybrides *O. sativa* x *O. glaberrima*. Pour la construction du modèle, nous avons pris en compte les données suivantes : (1) La localisation de S_I déduite à partir du pic de TRD. (2) La transmission systématique d'un bloc génomique autour de S_I^s , qui indique la présence d'au moins deux locus additionnels situés de part et d'autre de S_I . (3) La perte de la stérilité et du TRD dans les recombinants hétérozygotes pour S_I qui ont récupéré l'état homozygote *sativa* dans la région antérieure à S_I (mise en évidence par le gène *Wx* (Sano, 1990)), ou postérieure à S_I (mise en évidence par le recombinant T65 S_I (r) (Koide *et al.*, 2008c)). (4) Les observations cytologiques de deux types de gamètes femelles avortés caractérisés par l'absence de sac embryonnaire, probablement suite à un avortement post-méiotique, et d'autres présentant un sac embryonnaire de moins de sept cellules, sans polarisation (Bouharmont *et al.*, 1985; Koide *et al.*, 2008c).

Plusieurs modèles d'élimination allélique au cours de la sporogénèse ou de la gamétogénèse furent considérés. Nous avons retenu celui pour lequel l'incompatibilité entre les trois locus a lieu pendant la méiose, car c'est le seul qui puisse expliquer une élimination élevée – mais pas totale – des allèles S_I^s . Le modèle est capable de prédire : (1) Les fréquences alléliques de S_I^s et S_I^g observées dans les populations synthétiques BC₁F₁, qui sont directement liées aux proportions de gamètes viables produits par la F₁. (2) Les taux de recombinaison entre S_I et les marqueurs qui limitent la région *sativa* héritée en bloc, et entre la région des trois locus et le centromère. (3) La fertilité paniculaire observée. Cependant, pour les trois croisements vers le fonds génétique *japonica*, et celui vers le fonds génétique *indica*, la fréquence allélique prédite pour S_I^s est respectivement plus haute et plus basse que celle observée dans les populations BC₁F₁. Nous avons donc appliqué deux correctifs afin de mieux ajuster le modèle:

- La transmission extrêmement réduite de S_I^s dans le croisement IR64/TOG5681//IR64 (fonds génétique *indica*) peut être expliquée par la présence d'un quatrième locus, $S_I C$, qui aurait un effet gamétophytique additionnel et dépendant de l'effet du complexe de S_I . L'existence de ce locus a été déduite de la même façon que celle de $S_I A$ et $S_I B$, à partir des données génotypiques qui montrent une transmission systématique d'un haplotype *sativa* de très longue taille (environ 10 Mb) chez les plantes S_I^s/S_I^s , alors que le taux de recombinaison qui sépare $S_I B$ de $S_I C$ devrait permettre d'observer de nombreux recombinants, comme c'est le cas pour les autres croisements (voir Figure III-5). Le modèle qui prend en compte la présence de $S_I C$ et le taux de recombinaison entre $S_I B$ et le marqueur qui limite l'intervalle de présence de $S_I C$ s'ajuste mieux aux fréquences alléliques observées pour ce croisement.
- Le deuxième correctif est l'application d'un coefficient de survie ν à certaines des cellules S_I^s pour les croisements à fonds génétique *japonica* Caiapo/MG12//Caiapo, Curinga/CG14//Curinga et Nipponbare/MG12//Nipponbare. Le fait de requérir artificiellement à ce coefficient afin de mieux ajuster le modèle à une transmission de S_I^s plus élevée est une manière de prendre en compte l'existence d'un facteur supplémentaire dont l'effet sur la survie des gamètes vient moduler celui du complexe S_I . Nous ignorons tout de ce facteur, qui pourrait être un (ou plusieurs) autre gène localisé ailleurs dans le génome, ou des différences entre les allèles d'*indica* et de *japonica* pour un ou plusieurs des locus S_I . Ce dernier point représente une des faiblesses du modèle proposé et laisse donc un terme inconnu au modèle. On peut tenter une explication : par exemple, si les gènes pour S_I et $S_I A$ correspondent bien aux gènes F-box, (voir section VI.7.5) les différences entre le nombre de copies chez *indica* (93-11) et *japonica* (Nipponbare) du gène F-box situé dans la région du

locus S_1A pourraient être à l'origine du taux de survie plus élevé des gamètes S_1^s dans le fonds génétique *japonica*. Un autre point faible du modèle est la prise en compte d'un possible effet du fonds génétique d'*O. sativa* sur l'élimination des gamètes. Le facteur qui assure la survie de certains gamètes S_1^s dans les croisements *japonica* et qui est pris en compte par le coefficient v , et le locus additionnel S_1C semblent expliquer les différences dans la fertilité et la transmission des allèles *sativa* entre *indica* et *japonica*, mais cependant il reste à déterminer si l'effet de S_1C est particulier au fonds génétique *indica*, ou si il est plutôt dû à son incompatibilité avec certaines accessions d'*O. glaberrima*.

Trois types d'évidences favorisent plutôt la deuxième explication : (1) La faible transmission de S_1^s dans le croisement *Caiapo* x TOG5681, comparée à celle observée pour le croisement *Caiapo* x MG12 (voir Table III-1) (2) Les auteurs de plusieurs études ont observé que des variétés *indica* présentent une fertilité plus haute en croisement avec *O. glaberrima* que des variétés *japonica* (Morishima *et al.*, 1962; Bougerol et Pham, 1989). (3) L'accession *japonica* Taichung 65 utilisée par le groupe de Y. Sano, qui montre un TRD équivalant à un k_f de 0,81 en croisement avec l'accession *O. glaberrima* W025, montre un TRD plus élevé ($k_f = 0,92$) en croisement avec l'accession IRGC104038 (Doi *et al.*, 1998b). (4) Les données de fertilité et k pour les accessions *indica* (IR64) et *japonica* (Curinga et WAB165) obtenues pour le projet iBridges (voir Tableau V-5). Il apparaît donc probable que les différences observées entre *indica* et *japonica* au niveau de la transmission de S_1^s ne sont pas dues exclusivement au fonds génétique *sativa*, mais plutôt au jeu d'allèles en présence dans un certain croisement. Toutefois, des expériences complémentaires sont nécessaires pour pouvoir confirmer cette hypothèse.

VI.3 Vers un modèle général de la barrière reproductive entre *O. sativa* et

O. glaberrima

Les résultats obtenus au cours du présent travail de thèse nous ont permis de mieux comprendre le fonctionnement de la stérilité femelle causée par le locus S_1 . A présent, et à partir des informations disponibles sur les autres locus de stérilité hybride, nous pouvons essayer de construire un modèle général de la barrière reproductive entre *O. sativa* et *O. glaberrima*.

Le groupe de Y. Sano avait décrit S_1 comme un locus complexe, qui abrite deux composants qui affectent séparément la gamétogenèse mâle et femelle, et des modificateurs qui augmentent le fTRD (Koide *et al.*, 2008c). Le composant qui cause la stérilité mâle et la

mTRD fut cartographié en utilisant des lignées dérivées d'un matériel BC₁₂, dans un intervalle équivalent à 45 kb du génome de Nipponbare, tandis que la localisation du facteur qui affecte la stérilité femelle a été inférée dans l'intervalle 2,28 - 4,36 Mb du bras court du chromosome 6. Or, nos résultats placent le composant qui affecte la stérilité femelle dans la même région où se cartographie le facteur mâle, et placent l'un des locus qui interagissent avec S_I , S_IB , dans l'intervalle génomique déduit des travaux du groupe de Y. Sano pour le facteur à effet femelle.

Nous avons vu au Chapitre III que la co-localisation des facteurs mâle et femelle pourrait indiquer qu'un seul facteur est impliqué dans l'élimination des deux types de gamètes. Dans ce cas, le gène qui code pour S_I aurait un effet pléiotropique sur la gamétogenèse mâle et femelle, capable à lui seul d'induire l'élimination des gamètes mâles S_I^s , mais requérant l'interaction avec d'autres locus pour éliminer les gamètes femelles. Il reste à déterminer le mécanisme d'élimination totale des gamètes mâle portant l'allèle S_I^s . D'après les évaluations cytologiques réalisées par Koide et al., (2008c), le développement des microspores s'arrête après la première division mitotique. A cette étape, chaque microspore a déjà hérité d'une des quatre chromatides portant un allèle de S_I . On peut se demander comment une incompatibilité entre les produits de S_I^s et S_I^g peut avoir lieu, quand à cette étape il n'y a plus de plasmodesme ni de canaux cytomictiques (i.e. de longs canaux de connexion entre méiocytes permettant l'échange d'organelles) pour connecter les microspores (Mamun *et al.*, 2005). On pourrait faire l'hypothèse que le produit codé par S_I soit synthétisé avant la disparition complète des canaux dans la tétrade, et qu'il perdure dans les microspores. Le produit de l'allèle S_I^g interférerait alors postérieurement au développement des microspores S_I^s . Une autre hypothèse, proposée pour expliquer un phénomène similaire dans l'élimination allélique des gamètes *SaM* chez *O. sativa*, serait l'existence d'un mécanisme de transport de protéines entre les cellules déjà isolées (Long *et al.*, 2008).

Malgré son rôle important dans la stérilité des hybrides *O. sativa* x *O. glaberrima*, S_I n'est pas le seul locus à conférer une barrière reproductive entre les deux espèces. Il existe au moins dix autres locus et QTL décrits à partir de différentes populations interspécifiques, répartis sur l'ensemble du génome, et qui confèrent une stérilité plus ou moins prononcée de ces hybrides (voir Tableau I-9 et Figure VI-1).

Une analyse comparative des positions de ces locus avec les différentes cartes interspécifiques disponibles devrait permettre d'identifier les locus à effet congruent (i.e., partagé par différents croisements), et les locus à effet spécifique d'un croisement. En croisant cette information avec les positions des marqueurs présentant un TRD, on obtiendra une vue

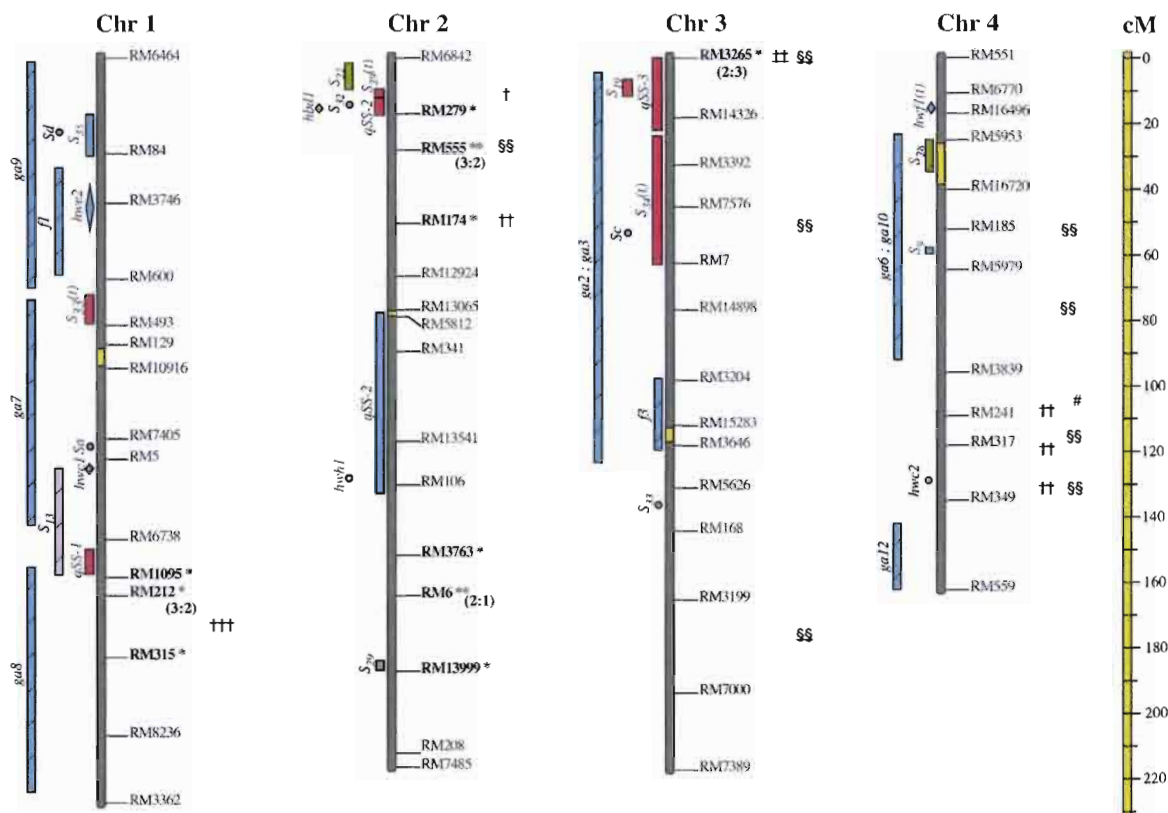


Figure VI-1. Co-localisation des différents locus de barrières reproductives chez le riz. Projection approximative des locus impliqués dans une barrière reproductrice chez le riz sur la carte génétique du croisement IR64/TOG5681//IR64 (Chapitre III). Les marqueurs qui présentent une distorsion de la ségrégation sont représentés ($P < 0,05$ (*), $P < 0,01$ (**), et $P < 10^{-5}$ (***) selon le rapport de leur ségrégation (*O. sativa*/*O. sativa* : *O. sativa*/*O. glaberrima*). Les locus sont représentés par un couleur représentant le croisement pour lequel ils ont été décrits : Intraspécifiques *O. sativa* (bleu), *O. sativa* x *O. rufipogon* (jaune foncé), *O. sativa* x *O. glaberrima* (rouge), *O. sativa* x *O. glumaepatula* (vert), et *O. sativa* x *O. longistaminata* (mauve). Les cercles pleins représentent les locus cartographiés finement, les ovales représentent les locus de létalité hybride, les rectangles représentent les locus de stérilité hybride, et les losanges représentent les locus de « hybrid breakdown ». La taille du symbole est proportionnelle à la taille de la région du génome associée avec les locus (bornée par des marqueurs moléculaires). Les centromères de chaque chromosome sont représentés par des rectangles jaunes. Les locus pour lesquels aucun, ou seulement un marqueur a pu être localisé sur la carte, ou dont la localisation est approximative, sont symbolisés en hachuré.

d'ensemble des différentes régions génomiques impliquées dans la barrière de reproduction. Quatre locus ou QTL, localisés sur les chromosomes 1 (*qSS-1*), 2 (*S_{29(t)}* ou *qSS-2*), 3 (*S₁₉* ou *qSS-3*), 6 (*S₁*), sont identifiés ou présentent un TRD dans au moins quatre des croisements comparés, voir T65/IRGC104038//T65 (Doi et al., 1998b), CG14/WAB450-16//WAB450-16 (Li et al., 2008b), BC3HD de *Caiapo* x MG12 (Aluko et al., 2004), et notre croisement IR64/TOG5681//IR64 (Figure VI-1). Parmi ces quatre locus, trois (*qSS-2*, *qSS-3* et *S₁*) ont été reportés comme ayant un effet sur la fertilité pollinique et des épillets (Koide *et al.*, 2008c; Li *et al.*, 2008b). Pour le quatrième, *qSS-1*, le fait d'observer un fTRD dans nos populations BC₁F₁ qui co-localise avec sa position génomique pourrait impliquer que son action ne se limite pas à la gamétogenèse mâle, mais qu'il peut aussi avoir un rôle dans un ou plusieurs des processus de gamétogenèse femelle, pollinisation, fécondation, ou dans la survie des zygotes ou des endospermes, en résumé dans la fertilité des épillets. Six autres locus (*S_{34(t)}*, *S₂₀* ou *qSS-7a*, *qSS-7b*, *S₂₁*, *S₁₈* et *S₃*) sont communs entre deux croisements. A l'exception de *qSS-7a*, ces locus ont été décrits comme causant une stérilité pollinique. Le fait de ne pas trouver de fTRD, mais seulement une TRD dans la population BC₃HD (issue de culture d'anthères) pourrait indiquer que leurs actions se limitent aux gamètes mâles. Enfin, un locus (*S_{33(t)}*) n'est présent dans aucun des croisements examinés, remettant en question son rôle dans la fertilité.

En ce qui concerne les sites de TRD observés pour les quatre populations disponibles, on observe que leur intensité et leur sens (contre-sélection de l'allèle *sativa* ou *glaberrima*) sont congruents entre les croisements. Dans la plupart des cas, c'est l'allèle *sativa* qui est favorisé, et c'est seulement dans trois cas (un site sur le chromosome 3 et deux sur le 6), que l'allèle hérité préférentiellement est celui d'*O. glaberrima*. La seule région du génome où le sens de la TRD varie selon le croisement est celle du centromère du chromosome 11. Ici, l'allèle *sativa* est surreprésenté dans la population *Caiapo* x MG12 BC₃HD, tandis que dans la population IR64/TOG5681//IR64 c'est l'allèle *glaberrima* qui est favorisé. Cette donnée pourrait indiquer que le locus qui cause cette TRD est polymorphe entre les accessions des deux espèces, et que selon les allèles confrontés dans un croisement donné, l'élimination favorisera l'un ou l'autre des allèles. On notera que cette région (représentée par les mêmes ancres de la carte UGCM) présente une TRD avec surreprésentation des allèles *sativa*, dans le croisement entre *O. sativa* et *O. meridionalis* (L. Moreno, communication personnelle).

Finalement, cinq régions du génome présentent une TRD sans être associées à un locus de stérilité répertorié entre les deux espèces (chromosomes 2, 3, 4 et 5). Cela pourrait signifier que ces régions portent un facteur altérant la ségrégation mendélienne des marqueurs,

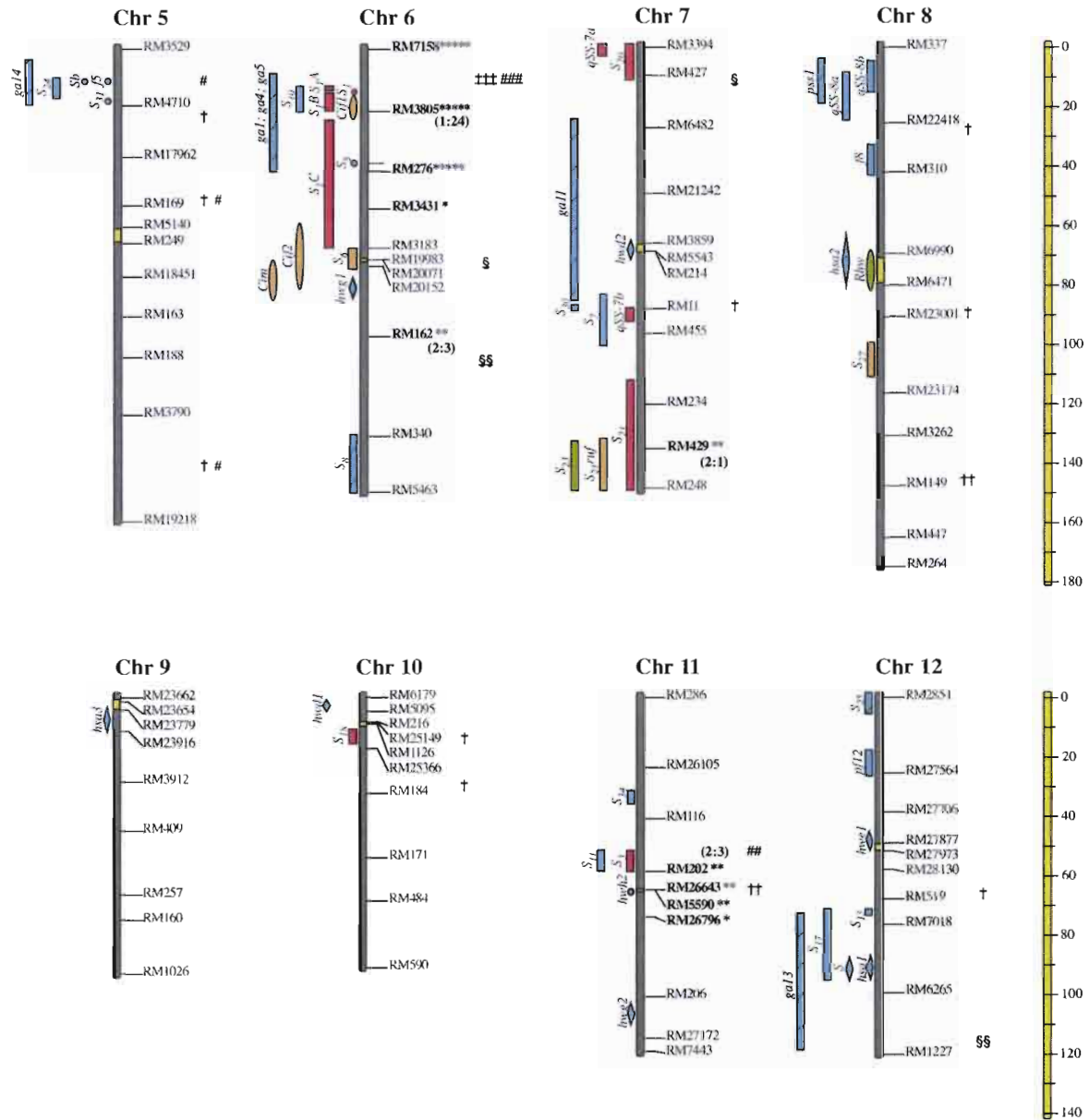


Figure VI-1. (Suite). Les localisations approximatives des marqueurs moléculaires qui présentent un TRD dans les populations interspécifiques T65/IRGC104038//T65 (#) (Doi *et al.*, 1998b), CG14/WAB450-16//WAB450-16 (§) (Li *et al.*, 2008b), et BC₃HD de Caiapo x MG12 (†) (Aluko *et al.*, 2004) sont indiquées. L'intensité de la distorsion est représentée par une (P < 0,05), deux (P < 0,01), ou trois (P < 0,001) répétitions des symboles. Les symboles surlignés indiquent une surreprésentation des allèles *glaberrima*, ceux non surlignés une surreprésentation de l'allèle *sativa*.

par divers mécanismes décrits, par exemple : (1) l'élimination allélique des gamètes correspondant à une stérilité hybride. (2) Une préférence de la provenance du pollen au moment de la fécondation ou (3) la germination du pollen, comme dans le cas d'une barrière pré-zygotique. Ou, finalement, (4) une influence dans la survie des zygotes ou de l'endosperme des hybrides, comme dans la faiblesse hybride de la F_1 ou de la F_2 (voir section I.6.4).

Le cas du TRD observé au niveau du centromère et du bras long du chromosome 6 est particulier, car le déséquilibre allélique ne semble pas avoir la même intensité selon le croisement étudié (si l'on prend en compte les populations additionnelles développées pour ce travail de thèse, voir Figure III-2). A ce stade, il est difficile de déterminer la cause de ce phénomène, mais trois observations principales nous font penser qu'il s'agit d'une région importante pour la viabilité des zygotes ou de l'endosperme des hybrides : (1) la même région a été décrite comme contenant plusieurs gènes de létalité des hybrides entre *O. sativa* et *O. rufipogon*. Ces gènes, dont deux sont localisés entre le centromère et le bras long du chromosome 6 (*Cif₂* et *CiM*), et l'autre (*Cif₁*) dans la région de *S₁B* (Matsubara *et al.*, 2003; Koide *et al.*, 2008b), interagissent pour causer une létalité unidirectionnelle de l'endosperme. *O. rufipogon* possède les allèles *Cif₁*, *Cif₂* et *Cim*, tandis que *O. sativa* possède les allèles *cif₁*, *cif₂* et *cim*. L'avortement de l'endosperme survient seulement lorsqu'une plante *Cif* est utilisée comme parent femelle et une plante *cim* comme mâle. (2) Le TRD n'est pas observé dans cette région dans la population BC₃HD du croisement MG12/*Caiapo*, mais il est bien observé dans la population BC₁F₁ *Caiapo*/MG12//*Caiapo*. On notera le rapprochement entre cette distorsion unidirectionnelle et celle mise en évidence pour le système *Cif*/*Cim*. Il se pourrait qu'un mécanisme similaire ou des allèles différents pour les mêmes locus soient portés par *O. sativa* et *O. glaberrima*. (3) Il est intéressant de noter – même s'il s'agit peut-être d'une coïncidence – que, parmi nos croisements, le TRD le plus fort est observé dans la population (Nipponbare/MG12//Nipponbare) pour laquelle l'obtention des individus BC₁F₁ a été la plus problématique (voir Figure 2, Chapitre III). De nouveau, les mêmes ancres de la carte UGCM présentent un TRD avec surreprésentation des allèles *sativa*, dans le croisement entre *O. sativa* et *O. meridionalis* (L. Moreno, communication personnelle). Cette région héberge peut-être un ou plusieurs locus impliqués dans un mécanisme-clé pour la survie des embryons ou des endospermes dans les espèces du groupe génomique AA.

Une analyse plus approfondie de ces différentes régions sera nécessaire pour pouvoir établir leur fonction dans la barrière reproductive entre *O. sativa* et *O. glaberrima*.

Il y a de fortes chances que des interactions épistatiques aient lieu entre les différents locus de stérilité ou les régions qui présentent une TRD. L'existence d'une épistasie négative entre locus de stérilité de régions chromosomiques différentes (S_{24} et S_{35}) a été préalablement démontrée dans des croisements intraspécifiques d'*O. sativa*. Dans cet exemple, les gamètes portant les allèles *japonica* à S_{35} sont avortés seulement en présence de l'allèle *indica* de S_{24} (Kubo *et al.*, 2008). Il n'est pas exclu que des interactions de ce type soient présentes dans les croisements interspécifiques entre *O. sativa* et *O. glaberrima*. Li et ses collaborateurs ont suggéré l'existence d'une interaction épistatique entre les QTL $qSS-1$ et $qSS-7a$, car l'effet sur la fertilité pollinique de chaque locus n'apparaît que lorsque l'autre se trouve à l'état homozygote (Li *et al.*, 2008b). Des données préliminaires obtenues pendant notre thèse avec des lignes CSSL (résultats non présentés), nous font penser à une possible additivité des effets des locus ou à une interaction épistatique entre eux. Cette type d'interaction pourrait être responsable des différences observées dans le niveau d'élimination de l'allèle S_I^s selon le fonds génétique d'*O. sativa*, dont l'équation du modèle tient compte par l'inclusion du coefficient v .

En conclusion, la stérilité des hybrides F_1 obtenus par croisement entre *O. sativa* et *O. glaberrima* est le résultat de l'action cumulée de plusieurs locus. La mesure de la stérilité femelle est fortement, voire totalement due à l'action du locus S_I , tandis que le phénotype de stérilité mâle dépend tant de S_I comme des effets des autres locus de stérilité pollinique distribués sur tout le génome. C'est donc la somme des effets de l'ensemble de ces locus qui rendent les hybrides F_1 totalement mâle stériles et partiellement femelle stériles.

VI.4 Vues sur la spéciation : conclusions sur la divergence des complexes

O. sativa-O. rufipogon* et *O. glaberrima-O. barthii

Les informations disponibles sur la littérature placent la divergence des espèces *O. sativa* et *O. glaberrima* il y a environ 700.000 ans (Ma et Bennetzen, 2004; Zhu et Ge, 2005; Ammiraju *et al.*, 2008). Les processus de domestication pour chacune des deux espèces ayant eu lieu longtemps après leur divergence à partir d'un ancêtre commun (il y a au moins 7 000 ans pour *O. sativa* (Fuller *et al.*, 2009), et 3 000 ans pour *O. glaberrima* (Portères, 1962; Murray, 2004)), il est indiscutable que cette séparation entre les espèces est survenue entre leurs ancêtres sauvages respectifs, et qu'elle se répercute sur les espèces cultivées. En termes de spéciation, on parlera donc de la divergence des complexes *O. sativa-O. rufipogon* et *O. glaberrima-O. barthii*, plutôt que de la divergence entre *O. sativa* et *O. glaberrima*.

La région génomique qui contient le locus S_I a été étudiée en détail, impliquant des comparaisons minutieuses entre les séquences orthologues d'*O. sativa* et *O. glaberrima*, ce qui nous a permis d'émettre des hypothèses sur de possibles mécanismes de divergence de ces complexes. Pour l'ensemble de la région de S_I , le patron d'évolution génomique ne semble pas différent à celui des autres régions du génome d'*O. glaberrima* séquencées par d'autres groupes de recherche (Ammiraju *et al.*, 2008; Lu *et al.*, 2009), ce qui pourrait être en accord avec un modèle de spéciation en allopatrie, dans lequel une barrière post-zygotique est un produit secondaire de la divergence évolutive sous l'action de la sélection naturelle. Au niveau des gènes, nous observons que la plupart d'entre eux ont subi une évolution purifiante, tandis que les autres ont subi une évolution neutre, ou neutre-accélérée, et sous une évolution diversifiante (sélection positive). Le patron d'évolution des régions intergéniques est différent, avec de fortes restructurations localisées principalement aux alentours des meilleurs gènes candidats. Il est cependant difficile de déterminer si ces restructurations ont directement contribué au processus de spéciation, ou si au contraire elles sont survenues postérieurement et ne sont qu'un résultat de la divergence des complexes *O. sativa*-*O. rufipogon* et *O. glaberrima*-*O. barthii*.

Si les candidats les plus probables pour ces locus se confirmaient (voir section VI.7.5), cela représenterait une évidence supplémentaire de l'implication directe de la duplication suivie de divergence de gènes dans les incompatibilités de type BDM, et par conséquent, dans les barrières de reproduction post-zygotiques. De même, la présence d'une inversion dans la région pourrait représenter une preuve directe de l'implication des réarrangements chromosomiques dans la spéciation. On sait que pour plusieurs systèmes, les changements structuraux dans les chromosomes peuvent constituer par eux mêmes des barrières reproductives entre deux espèces (Coyne et Orr, 2004), mais peu d'information biologique permet d'établir un rôle direct de ces réarrangements avec l'accumulation d'incompatibilités de type BDM, par exemple en générant une liaison génétique plus étroite des gènes incompatibles impliqués (Navarro et Barton, 2003). La présence d'une inversion chromosomique dans la région de S_I pourrait avoir ce type d'effet même si la réduction de la recombinaison entre les facteurs est faible.

Par ailleurs, il est remarquable que le locus de stérilité S_{I0} entre *indica* et *japonica* soit précisément localisé dans la même région que S_I (Sano *et al.*, 1994; Zhu *et al.*, 2005a). Il se peut que les différences d'intensité du TRD observées dans nos populations interspécifiques en fonction du fonds génétique *indica* ou *japonica* reflètent des différences alléliques entre les facteurs impliqués, comme nous le suggérons dans la section précédente, et que, de plus, ces

différences alléliques soient suffisantes pour causer une incompatibilité BDM et induire une stérilité entre les deux sous-espèces d'*O. sativa*.

Pour l'heure, il est difficile d'établir avec certitude si la barrière exercée par S_I , S_{IA} et S_{IB} entre les complexes *O. sativa*-*O. rufipogon* et *O. glaberrima*-*O. barthii* fut antérieure à la divergence des populations (effet fondateur sur la spéciation), ou si au contraire, elle est survenue après la divergence des deux complexes (comme conséquence de la spéciation). A ce propos, la caractérisation de certaines barrières reproductives chez les animaux, dont les gènes ont été identifiés, a permis aux auteurs de proposer l'hypothèse d'une implication directe de ces barrières dans le processus de spéciation. Par exemple, chez la souris le fait que l'isolation reproductrice entre les deux sous-espèces étudiées (*Mus musculus* ssp. *musculus* et *Mus musculus* ssp. *domesticus*) ne soit pas totale, a amené les auteurs à postuler que le gène *Prdm9* fait partie d'une barrière naissante (Mihola *et al.*, 2009). D'une façon similaire, les auteurs, qui ont identifié le gène *Ovd*, considèrent que celui-ci est une composante importante de la spéciation en cours, et non une barrière ayant évolué après l'isolement reproductif (Phadnis et Orr, 2009). Ils s'appuient sur le fait que ce gène, et les autres locus qui interagissent en épistasie avec lui, sont essentiels pour la manifestation de la stérilité hybride entre les deux sous-espèces de *D. pseudoobscura*, *pseudoobscura* et *bogotana*. Par analogie, nous pouvons formuler l'hypothèse que le complexe composé par S_{IA} , S_I , et S_{IB} , responsable de l'élimination d'un nombre élevé des gamètes mâles et femelles (environ 50 %), pourrait être cataloguée comme une composante majeure de la spéciation des complexes *O. sativa*-*O. rufipogon* et *O. glaberrima*-*O. barthii*, mais que le processus de spéciation est à un état plus avancé que les deux exemples reportés, du fait de la présence d'autres locus ayant un effet additionnel sur la stérilité hybride.

Si on se reporte aux premières observations réalisées sur la fertilité des hybrides dérivés des croisements avec *O. glaberrima* réalisés par Oka (Morishima *et al.*, 1962), qui montrent une fertilité hybride plus élevée d'*O. glaberrima* avec *O. rufipogon* (5 à 25 % pour leurs hybrides F_1 en comparaison avec la stérilité totale des hybrides *O. sativa*/*O. glaberrima*), on peut imaginer que la barrière reproductrice post-zygotique initiale se serait renforcé après la domestication d'*O. sativa*. Cela pourrait traduire un effet des locus additionnels dont l'action ajoutée produirait la forte barrière reproductrice post-zygotique entre *O. sativa* et *O. glaberrima*, par une accumulation d'incompatibilités du type BDM, en accord avec un effet « snowball » (Orr et Turelli, 2001).

La co-localisation des locus de stérilité entre *O. sativa* et *O. glaberrima* avec ceux identifiés pour les croisements entre *O. sativa* et les autres espèces du génome AA est très intéressante.

En plus de la région de S_1 et S_{10} , cinq autres régions du génome (chromosomes 1, 2, 7, 8, et 11) hébergent des locus de barrières post-zygotiques pour au moins deux croisements interspécifiques ou intraspécifiques (Figure VI-1). Les cas les plus visibles de co-localisation de locus de stérilité sont ceux qui impliquent S_{21} , et $S_{23} - S_{21}ruf$ sur le chromosome 7, et celui entre S_{22} et $S_{29}(t) - S_{32} - qSS-2$ sur le chromosome 2. La question est de savoir si cette co-localisation pourrait constituer une évidence que les mêmes locus sont impliqués dans les barrières de stérilité entre plusieurs espèces. Si tel est le cas, cela voudrait dire que les facteurs impliqués sont représentés dans chaque espèce par un allèle différent, et que des incompatibilités entre ces allèles sont responsables de la stérilité. Les informations obtenues pour d'autres systèmes d'isolement reproductif chez les plantes (Christie et Macnair, 1987; Sweigart *et al.*, 2007), montrent que les locus d'incompatibilités du type BDM peuvent être hautement polymorphes entre les populations, et qu'ils peuvent constituer postérieurement une barrière post-zygotique établie, si les populations ont fixé des allèles incompatibles. Il est vraisemblable que chaque espèce du genre *Oryza* ait fixé un allèle différent pour ces locus, et que l'incompatibilité soit aussi présente dans des croisements qui n'impliquent pas *O. sativa*.

VI.5 Conséquences sur l'étude des variétés hybrides naturels

Il est intéressant de constater l'existence d'hybrides naturels ou variétés *admixture* parmi les accessions prospectés en Afrique, malgré la puissante barrière de stérilité exercée par S_1 et les autres locus de stérilité. Si on compare avec les hybrides inter-subspécifiques *sativa* x *japonica*, on constate que leur facilité d'obtention est fortement liée aux allèles neutres possédés par les WCV (Wang *et al.*, 1998). Les allèles neutres des locus de stérilité entre *indica* et *japonica*, en particulier celui du locus S_5 (S_5^n), ont permis un flux de gènes qui a agi en sens inverse du processus de spéciation, et a peut-être empêché l'isolement total (Chen *et al.*, 2008).

Pour le cas d'*O. sativa* et *O. glaberrima* il n'a pas été détecté jusqu'à présent de système de ce type qui s'opposerait au processus de spéciation. A ce sujet, une étude plus approfondie de l'haplotype S_1 des variétés *admixture* est nécessaire pour pouvoir déterminer si la présence de ces allèles (qui correspondent à l'haplotype *sativa-2*) est simplement due à une durée de sympatrie plus longue avec *O. glaberrima*, que dans le cas des variétés d'*O. sativa* portant l'haplotype *sativa-1* ; ou, si la présence systématique de cet haplotype dans les hybrides naturels est due à une compatibilité plus élevée avec *O. glaberrima* que celle présenté par l'haplotype *sativa-1*. Si la deuxième hypothèse s'avérait être la bonne, ses implications dans l'utilisation d'*O. glaberrima* pour l'amélioration d'*O. sativa* seraient très importantes, car les

variétés à haplotype *sativa*-2 seraient de meilleurs candidats pour la création de ponts interspécifiques, apportant une compatibilité similaire à celle des accessions WCV.

Les variétés révélées comme hybrides naturels ont été soumises durant des siècles à des processus de sélection qui leur ont conféré un haut niveau d'autofertilité, malgré leur caractère hybride. En étudiant d'une part la fertilité de ces *admixture* en croisement avec *O. sativa*, *O. glaberrima* et entre eux, et d'autre part leur contenu allélique sur l'ensemble du génome, on peut espérer repérer les régions associées à la fertilité vers l'une ou l'autre espèce. Une telle approche élargirait la vision que l'on a sur les barrières à contourner pour améliorer les techniques d'hybridation entre les deux espèces.

VI.6 Conséquences sur les stratégies d'amélioration

Les résultats rapportés dans le présent travail de thèse contribuent de deux façons principales à l'établissement de stratégies d'amélioration du riz. La première contribution se base sur l'application des outils développés, pour une large gamme de situations, et la deuxième est en relation directe avec l'exploitation d'*O. glaberrima* comme source de gènes d'intérêt.

En ce qui concerne les deux outils développés au cours de la thèse, ils représentent un avancement important, et pourraient avoir des répercussions directes dans la façon de concevoir les stratégies d'amélioration, non seulement sur le riz, mais aussi pour toute espèces où la disponibilité de séquences génomiques permet d'envisager le développement d'outils équivalents. L'UCGM est un outil fiable et puissant qui facilitera la construction et la comparaison des cartes génétiques et des localisations des gènes et QTL cartographiés dans différentes populations, ainsi que la sélection assistée par marqueurs. L'autre outil développé, une méthodologie pour la cartographie fine des gènes de stérilité, réduit de manière considérable le temps requis pour identifier les régions génomiques qui portent les facteurs causant des distorsions de ségrégation. La méthode est applicable à toute espèce pour laquelle la disponibilité en séquences génomiques permet la définition de marqueurs très saturants.

La conséquence la plus importante de nos travaux sur les stratégies d'amélioration du riz est à mettre en relation avec les objectifs originaux du projet iBridges : l'application des résultats obtenus permet d'envisager une utilisation facilitée et courante du germplasma d'*O. glaberrima* dans les programmes d'amélioration variétale. L'identification et la localisation des facteurs-clés de la barrière de reproduction permet d'envisager le développement de lignées hybrides à fertilité optimale en croisement avec *O. sativa* et qui conservent une proportion importante du génome d'*O. glaberrima*, par la méthode de rétrocroisement assisté par marqueurs.

Une évaluation des individus BC₁F₁ obtenus pour chacun des croisements du projet avec des marqueurs moléculaires localisés aux bornes du fragment génomique *S₁A* - *S₁B*, et un marqueur localisé à proximité de *S₁*, permettra de sélectionner précocement les plantes qui auront hérité de l'allèle *sativa* dans la région, et pour lesquelles la barrière exercée par le locus sera donc désactivée dans les générations suivantes. La possibilité d'évaluer rapidement et avec une technologie simple des matériels, dépend des caractéristiques des marqueurs, par exemple un polymorphisme suffisant pour être repéré en gel d'agarose. Par exemple, C6_20017 et RM3805 sont de bons marqueurs autour de *S₁A* et *S₁B*, ainsi que RMC6_21942 pour déterminer l'état allélique à *S₁*.

Certains caractères resteront très difficiles à introgresser d'*O. glaberrima* vers *O. sativa*. Il s'agit des gènes localisés dans la région de *S₁A*, *S₁* et *S₁B*, et peut-être des autres locus de stérilité. Par exemple, un locus qui influence la taille de la panicule des hybrides *O. rufipogon* x *O. sativa* a été récemment cartographié dans la région de *S₁A* (Shan *et al.*, 2009). S'il s'avérait que ce locus contrôle le même caractère chez les hybrides *O. sativa* x *O. glaberrima*, il sera difficile à introgresser vers *O. sativa* tout en conservant l'allèle *s* à *S₁*, du fait de la liaison génétique étroite entre les deux facteurs. La même difficulté sera rencontrée pour les gènes de résistance putatifs de la région (e.g. les gènes à domaine LRR très polymorphe).

Enfin, la forte conservation de structure génomique autour de *S₁* entre *O. glaberrima* et *O. barthii* permet d'espérer que les résultats et la stratégie pour le développement de lignées hybrides fertiles pourra s'appliquer assez directement aux croisements entre *O. sativa* et *O. barthii*. Ainsi, ces deux espèces africaines pourront être utilisées comme sources de variation allélique importantes pour l'amélioration du riz. L'application de ce concept de ponts interspécifiques pourra être étendue aux autres espèces de riz du groupe génomique AA, en se basant sur la localisation des principaux facteurs de stérilité pour chaque combinaison interspécifique.

VI.7 Les gènes candidats

Nous avons délimité la région qui porte le locus *S₁* à un segment de 50,3 kb (chez *O. glaberrima*). Cette région contient dix gènes, dont cinq seulement sont présents dans la région orthologue de 27,8 kb chez *O. sativa*. Ces dix gènes représentent les candidats pour le facteur responsable de la stérilité hybride entre les deux espèces. Ils appartiennent à des familles géniques très différentes, dont les domaines et fonctions identifiés dans la littérature peuvent donner des indices sur leur implication dans notre phénotype.

VI.7.1 Les gènes hypothétiques

La différence dans le nombre de gènes présents dans la région de S_I entre les deux espèces est due à l'acquisition de différents gènes putatifs. Parmi ceux-ci, trois des cinq gènes additionnels chez *O. glaberrima* et le gène additionnel chez *O. sativa* sont des gènes hypothétiques trouvés par les algorithmes de prédiction de gènes, sans aucune donnée d'expression, aucun domaine reconnu, ni protéine homologue répertoriée, et dont l'origine se trouve dans l'insertion d'éléments transposables. Par exemple, le gène hypothétique Os06g04960 a subi une rupture due à l'insertion de plusieurs transposons chez *O. glaberrima*, ce qui a produit un autre gène hypothétique, le 49I08.7. Les deux gènes hypothétiques additionnels (49I08.85 et 49I08.9) proviennent de l'insertion de deux Pack-MULE. Tous les gènes additionnels de la région de S_I chez CG14 sont aussi présents chez TOG5681, indiquant une origine probable postérieure à la différenciation d'avec *O. sativa*. En ce qui concerne leur fonction, l'absence de données d'expression et d'identification de domaines fonctionnels ne permet pas de déterminer leur éventuelle implication dans le phénomène de stérilité hybride.

VI.7.2 Les gènes homologue aux Nodulines

Trois gènes de la région de S_I , et quatre autres localisés à proximité, présentent une homologie avec le gène Noduline-93 (N93) de *G. max*. Généralement, les gènes des protéines « Nodulines » sont exprimés exclusivement dans les nodules symbiotiques des légumineuses induits par la colonisation des bactéries fixatrices de l'azote, mais certains s'expriment aussi dans d'autres tissus. C'est le cas du gène N93 qui s'exprime dans les racines, les jeunes pousses et les plantules du soja (Lee *et al.*, 2004). Les données d'expression disponibles pour les homologues de ces gènes chez le riz, obtenus par profil transcriptionnel MPSS (Massively Parallel Signature Sequencing) (Nobuta *et al.*, 2007), montrent une expression généralisée de ces gènes. Selon les profils de signatures (uniques ou partagés par la famille), ils sont fortement transcrits dans les graines en développement et en germination, mais aussi – bien que dans une bien plus faible mesure – dans une grande variété de tissus comme les cals, les racines, les ovaires matures, les feuilles ainsi qu'après inoculation par *M. oryzae* et *X. oryzae*. Le mécanisme par lequel ces gènes pourraient être impliqués dans la stérilité des hybrides n'est donc pas clair. Cependant, ils ne peuvent pas non plus être écartés définitivement comme candidats.

VI.7.3 Le gène régulateur de la biogenèse des ribosomes

Le gène qui possède un domaine homologue à une protéine régulatrice de la biogenèse des ribosomes appartient à la famille des gènes *Rrs1*, hautement conservée parmi les eucaryotes.

Il est impliqué dans la maturation et l'assemblage des sous-unités ribosomales. Chez la levure *S. cerevisiae*, un homologue qui porte le même domaine est essentiel pour la croissance végétative (Tsuno *et al.*, 2000). Les données disponibles sur l'expression de ce gène sont très limitées, et seulement trois signatures MPSS chez le riz hybrident sur ce gène avec des patrons d'expression dans des tissus différents pour chacune (tissus méristématiques et floraux, ainsi qu'après inoculation par *M. oryzae* et *X. oryzae*, et stress hydrique). On imagine bien qu'une incompatibilité qui impliquerait ce type de protéine pourrait causer une létalité hybride, on voit moins bien comment elle pourrait causer une stérilité.

VI.7.4 Le Pack-MULE portant un fragment de AP2

L'autre gène spécifique d'*O. glaberrima* est un Pack-MULE qui porte un fragment du facteur de transcription APETALA (AP2). Des double mutants pour deux gènes de la famille AP2 montrent des problèmes de fertilité du fait de l'arrêt de la mégagamétoogenèse après la première division méiotique (Byzova *et al.*, 1999). Comme il a été mentionné (voir Chapitre III), les Pack-MULE sont capables de capturer et de transposer des fragments de gènes (Hanada *et al.*, 2009). Ils peuvent aussi interférer avec l'expression des gènes dont ils se dérivent, par un mécanisme de régulation post-transcriptionnel par silençage. Dans ce contexte, le gène 49I08.10 pourrait silencer l'expression de facteurs – putatifs – de transcription nécessaires à la mégagamétoogenèse possédant une homologie avec le fragment AP2. Ce silençage pourrait ainsi déclencher une dysgenèse hybride. Cependant, deux éléments remettent en question l'implication de ce mécanisme dans la barrière reproductive que nous étudions. Premièrement, la dysgenèse hybride est un phénomène unidirectionnel, c'est à dire qu'il se manifeste quand un parent femelle non porteur d'un certain transposon est croisé avec un mâle qui le possède, mais pas dans le cas réciproque. Dans le cas du système de stérilité hybride entre *O. sativa* et *O. glaberrima*, la barrière est active quand *O. sativa* (qui ne possède pas ce transposon dans cette région génomique) est utilisé comme parent male et femelle (Sano *et al.*, 1979; Bougerol et Pham, 1989). Deuxièmement, sauf à faire l'hypothèse (déjà écartée), de la présence de copies additionnelles du même transposon dans les intervalles délimités pour *S₁A* et *S₁B*, qui par un effet de dosage pourraient induire le silençage des homologues, il est difficile d'expliquer comment des incompatibilités du type BDM impliquant les Pack-MULE peuvent apparaître avec ces régions.

VI.7.5 Le gène F-box

Le dernier gène présent dans l'intervalle où nous avons cartographié S_I , code pour une protéine contenant un domaine F-box et un domaine LRR. Plusieurs éléments suggèrent son implication dans le mécanisme de stérilité hybride, que nous discuterons ici.

Les gènes F-box constituent une famille très diverse, et sont principalement impliqués dans la dégradation par ubiquitination de plusieurs protéines régulatrices et de signalisation. Elles font partie du complexe Ubiquitin ligase SCF (Skp1 – Cullin – F-box) dans lequel leur rôle principal est la reconnaissance et l'interaction avec la protéine-substrat afin de l'amener dans le complexe de dégradation (Zheng *et al.*, 2002). Les protéines F-box et les complexes SCF sont impliquées dans une large gamme de processus chez les plantes, comme les réponses hormonales (Dharmasiri *et al.*, 2005), la formation et le développement des organes (Durfée *et al.*, 2003), la progression du cycle cellulaire (Pesin et Orr-Weaver, 2008), et l'autoincompatibilité (McClure, 2009). Les gènes codant pour les protéines F-box constituent une famille génique particulièrement grande chez le riz, avec plus de 700 gènes et pseudogènes (Thibaud-Nissen *et al.*, 2009; Xu *et al.*, 2009). Leur vitesse d'évolution et de divergence est très élevée du fait d'événements de duplication (souvent en tandem), de réarrangements, et de changement de domaines, ce qui leur confère cette plasticité au niveau fonctionnel (Jain *et al.*, 2007).

L'annotation et la comparaison détaillée de la séquence de CG14 avec la pseudomolécule de Nipponbare autour du locus S_I , nous a permis d'identifier la présence de deux copies du même gène codant pour une protéine F-box, la première dans l'intervalle où se trouve le locus S_I , et la deuxième dans la région de S_{IA} .

Plusieurs données liées à leur biologie, leur expression et leur fonction, indiquent une probable implication de ces gènes F-box dans la barrière reproductive causée par le locus S_I :

- L'impossibilité d'hériter simultanément d'allèles incompatibles à deux locus (dans notre cas, les combinaisons alléliques S_I^s/S_{IA}^g et S_I^s/S_{IB}^g) peut être la manifestation d'une incompatibilité du type BDM entre un ou plusieurs facteurs. Ce type d'incompatibilité a été démontré entre les accessions Col et Cvi de *A. thaliana* (Bikard *et al.*, 2009). Dans ce système, la duplication d'un gène essentiel pour la synthèse de l'histidine, et sa divergence dans chaque population suivie de l'inactivation des copies alternées du gène, cause la létalité des hybrides qui héritent simultanément des deux copies non actives du gène.

La duplication à l'origine des deux copies du gène F-box a eu lieu avant la divergence des deux espèces. Postérieurement, chaque copie du gène a subi une évolution divergente chez chaque espèce, et nos analyses ont mis en évidence qu'il s'agissait d'une accumulation de

changements non synonymes, de l'insertion d'un élément transposable dans un intron pour le gène 48I09.11 localisé dans la région de S_I , et d'une duplication chez Nipponbare qui a engendré les gènes Os06g04690 et Os06g04710 (voir Chapitre IV), localisés dans la région de S_{IA} . Comme nous le mentionnons plus haut, les gènes des F-box appartiennent à une famille génique complexe, présentant un taux d'évolution très élevé lié à leur fonction d'interaction avec une grande variété de protéines. Cette évolution rapide qui leur permet de se spécialiser dans des processus différents (néofonctionnalisation), de développer des fonctions originales (subfonctionnalisation), ou d'avoir des fonctions redondantes sur la même voie métabolique, pourrait être à l'origine des incompatibilités alléliques entre des espèces différentes. En conséquence, leur rôle dans les interactions protéine–protéine pourrait bien être à l'origine d'un cas d'incompatibilité de type BDM.

Cette incompatibilité pourrait être causée par différents mécanismes : (1) des différences dans le niveau d'expression, la régulation transcriptionnelle ou après la transcription, (2) des différences fonctionnelles liées à l'affinité avec la protéine-cible ou son complexe, ou (3) des différences de localisation cellulaire de la protéine produite par chaque allèle de chaque locus (comme c'est le cas pour l'allèle S_5'' (Chen *et al.*, 2008)). Dans le cas d'une fonction redondante ou complémentaire entre les deux copies, ces phénomènes peuvent se traduire en une concurrence pour les substrats engendrée par un excès en protéine, soit au contraire, par une concentration insuffisante en protéines (comme c'est le cas chez *Arabidopsis* (Bikard *et al.*, 2009)). L'hypothèse d'une concurrence semble plus plausible, le gène additionnel chez Nipponbare, Os06g04710, semblant avoir perdu l'extrémité 5' où se trouve le domaine F-box. Si c'est bien le cas, cet allèle pourrait interagir avec la cible mais pas avec le complexe SCF. Le phénomène de concurrence et la présence de ce gène en copie additionnelle par rapport à *O. glaberrima* mais aussi par rapport à *indica* (voir Chapitre IV), pourrait aussi contribuer aux différences entre l'élimination des allèles S_I^s dans les fonds génétiques *indica* et *japonica*, et expliquer l'existence d'un gène de stérilité entre les deux sous-espèces dans cette région précise.

- Une étude approfondie de la classification et de l'expression des gènes des protéines F-box chez le riz a montré que ces trois gènes en particulier sont exprimés dans les tissus floraux (Jain *et al.*, 2007). Des analyses d'expression par puce à ADN (Affymetrix), réalisées spécifiquement avec la sonde Os.3577.1.S1_x_at qui hybride avec les trois gènes Os06g04690, Os06g04710 (région de S_{IA}) et Os06g04980 (région de S_I), a montré une surexpression de ces gènes dans les épillets pendant la méiose, à l'étape de jeune microspore, et dans le pollen vacuolisé, avec un pic d'expression à l'étape de jeune microspore. En outre,

le signal de cette sonde a montré une sous-expression significative de ces gènes dans des mutants du facteur de transcription *Udt1* (*Undeveloped tapetum 1*) (Jung *et al.*, 2005), un gène qui joue un rôle important dans la différenciation des méiocytes. Ces données indiquent que les trois gènes s'expriment dans le tissu approprié et aux étapes de développement où l'incompatibilité se manifeste selon notre modèle d'élimination gamétique. Toutefois, la technique ne permet pas de différencier les niveaux d'expression entre les copies du gène.

- Un couple de gènes adjacents composé d'une protéine F-box (*SaF*) et d'une protéine similaire à une SUMO E3 ligase (*SaM*) sont responsables de la stérilité mâle entre les deux sous-espèces d'*O. sativa* causée par le locus *Sa* (Long *et al.*, 2008). *Indica* et *japonica* possèdent respectivement les allèles SaF^+SaM^+ et SaF^-SaM^- . Chez *indica*, le motif auto-inhibiteur de SaM^+ empêche l'élimination des gamètes après l'interaction entre SaF^+ et SaM^+ , alors que ce motif est absent chez la protéine SaM^- chez *japonica*. L'interaction entre les allèles SaF^+ et SaM^- cause donc l'avortement des microspores qui possèdent l'allèle SaM^- produit par les plantes hétérozygotes SaF^+SaM^+ / SaF^-SaM^- . L'implication des protéines F-box déjà démontrée dans un mécanisme de stérilité chez le riz représente un indice supplémentaire en faveur de l'hypothèse de l'implication de ces gènes dans la barrière de reproduction *O. sativa* / *O. glaberrima*.

- Plusieurs gènes F-box et leurs complexes SCF sont impliqués dans la régulation du cycle cellulaire. Par exemple dans le déroulement de la deuxième mitose du pollen (Gusti *et al.*, 2009), et le passage du stade leptotène au pachytène (Wang et Yang, 2006), chez *Arabidopsis* ; ou le système APC/C (Anaphase-promoting complex / Cyclosome) qui joue différents rôles et agit comme régulateur durant la méiose et la mitose, chez les levures et les animaux (Pesin et Orr-Weaver, 2008). Dans notre modèle génétique de stérilité femelle causée par *S_I*, seules les cellules qui, après chaque division cellulaire méiotique et mitotique, possèdent une combinaison compatible des allèles *S_IA*, *S_I*, et *S_IB* peuvent continuer leur développement. Le rôle, bien documenté, des protéines avec un domaine F-box dans le déroulement de la méiose et de la mitose fait d'elles de sérieux candidats, notamment parce que l'existence d'incompatibilités qu'elles présentent avec d'autres peut empêcher le passage à la suivante phase du cycle, causant l'élimination des gamètes qui portent les allèles incompatibles.

- Des gènes homologues (mais pas d'orthologues) aux trois copies des F-box localisées sur la région de *S_IA* et *S_I* ont été trouvés chez le sorgho (SB08G000945) et chez l'espèce d'orge sauvage *Hordeum bulbosum* (HSLF1/HSLF2) (voir Chapitre IV). Parmi ceux-ci, les deux protéines de l'orge sont impliquées dans le phénotype d'auto-incompatibilité (Kakeda,

2009). Ce phénomène de SI (*self-incompatibility*) est contrôlé par l'interaction entre des protéines F-box produites par le pollen, et des RNases S produites par le pistil. Même si le contexte et le mécanisme sont différents, il est clair que des protéines avec une structure similaire à celles codées par les gènes F-box Os06g04690, Os06g04710 et Os06g04980 sont capables de générer une barrière reproductive chez les céréales.

VI.7.6 Candidats pour S_1A et S_1B

Les éléments décrits plus haut nous amènent à considérer les gènes Os06g04690, Os06g04710 et Os06g04980 comme les meilleurs candidats pour S_1A et S_1 . En revanche, nous n'avons pas de données permettant d'inférer un candidat pour le locus S_1B .

La présence d'une inversion dans la région de S_1B pourrait avoir un rôle dans la barrière de reproduction entre *O. sativa* et *O. glaberrima*. Dans le scénario le plus simple, la présence de cette inversion, qui se révèle sur la carte génétique par une restriction totale de la recombinaison entre les marqueurs qui la délimitent, contribue à augmenter la liaison génétique entre les gènes impliqués dans la barrière (Navarro et Barton, 2003), d'une façon semblable – avec un effet moins sévère toutefois – que celui observé pour l'haplotype-t décrit chez la souris (Schimenti, 2000).

Dans le modèle à deux gènes F-box agissant comme facteurs de la stérilité, le troisième facteur S_1B pourrait correspondre à l'un des composants du complexe de dégradation SCF, comme par exemple le gène qui code pour la cible à dégrader, ou encore un activateur-régulateur du complexe de dégradation (par exemple un facteur de transcription spécifique d'une étape précise du cycle cellulaire). Dans ce cas, le gène LOC_Os06g05760 qui code pour une protéine Ubiquitine putative, tout comme les autres protéines F-box localisées dans la région, seraient candidats pour S_1B . Il faudra de toute façon cartographier finement le facteur S_1B afin de réduire l'espace de recherche de candidats.

VI.8 Conclusion générale

Par des approches de génétique et de génomique comparée, nous avons pu déchiffrer et mieux comprendre le fonctionnement de la barrière reproductive qui sépare les deux espèces de riz cultivé, *O. sativa* et *O. glaberrima*, exercée par le locus *S_I*. Nos données ont permis la détection d'autres locus génétiquement liés à *S_I*, qui agissent probablement en épistasie par le biais d'incompatibilités alléliques entre eux. Sur cette base, nous avons construit un modèle génétique qui explique la semi stérilité des hybrides interspécifiques, et exploré la région génomique qui porte ces locus, chez les espèces du complexe *O. glaberrima/O. barthii*, et finalement identifié des gènes candidats pour deux des locus.

Le développement et l'implémentation de nouveaux outils et méthodologies, comme la carte génétique universelle du riz (UCGM), ou la stratégie de cartographie fine des gènes de stérilité basée sur l'étude du TRD, représentent un progrès important pour l'amélioration du riz, et peuvent être pris comme base pour établir des stratégies similaires pour d'autres plantes d'importance agronomique.

Les croisements naturels entre *O. sativa* et *O. glaberrima* sont à l'origine de l'existence d'hybrides naturels, qui selon leur génotype autour du locus *S_I*, s'avèrent plus ou moins fertiles en croisement avec *O. sativa*. Ces hybrides, exposés à la sélection naturelle pendant des siècles, constituent des ponts interspécifiques naturels. Leur existence, et leurs niveaux de fertilité avec *O. sativa* confirment le concept duquel est dérivé le projet iBridges. En conséquence, le projet iBridges, et les lignées générés par ce projet, devraient constituer un excellent outil pour exploiter le potentiel génétique d'*O. glaberrima* pour l'amélioration du riz cultivé.

L'ensemble de résultats présentés dans ce travail de thèse apportent des éléments pour une meilleure compréhension de la stérilité des hybrides causée par le locus *S_I*, et apporte une meilleure connaissance des barrières reproductives chez le riz en général. Ces connaissances, qui viennent s'ajouter à celles générées par d'autres équipes de recherche, permettront à terme au sélectionneurs de contourner efficacement ces barrières de reproduction et d'utiliser pleinement le pool génétique disponible dans les espèces apparentées au riz asiatique cultivé.

Il sera nécessaire de relier toutes les informations acquises avec des données biologiques, afin de mieux comprendre le mécanisme sous-jacent à la barrière de reproduction, notamment par l'identification des protéines impliquées et des processus biologiques perturbés par la barrière.

VI.9 Perspectives

VI.9.1 Cartographie fine de S_1A et S_1B

Grace à la méthodologie pour la cartographie des gènes de stérilité basée sur l'évaluation de l'intensité de la distorsion de la ségrégation aux locus marqueurs, et à l'analyse du génotype des gamètes S_1^s qui survivent dans les hybrides, nous avons mis en évidence l'existence d'au moins deux locus interagissant avec S_1 pour causer l'avortement des gamètes S_1^s , et nous avons déduit leur emplacement dans des intervalles équivalant à 183 et 654 kb du génome de Nipponbare. D'après le modèle génétique développé pour expliquer la stérilité femelle, il n'est pas possible de dériver des génotypes issus de gamètes – non viables – $S_1A^g-S_1^s-S_1B^g$, $S_1A^g-S_1^s-S_1B^s$, ou $S_1A^s-S_1^s-S_1B^g$. Pour cette raison, il semblait irréalisable d'isoler les facteurs S_1A^g ou S_1B^g du facteur S_1^g , par exemple dans une lignée NIL à fonds génétique *O. sativa*, dans le but d'en réaliser le clonage positionnel et d'évaluer leur effet individuel dans les recombinants.

Il est donc nécessaire d'envisager une stratégie différente pour cartographier finement les deux locus puis identifier les gènes candidats. La méthodologie devrait se diviser en deux approches visant à déterminer les bornes internes et externes des intervalles de probabilité de présence des facteurs (voir la Figure VI-2). L'analyse des points de recombinaison, dont la limite tend asymptotiquement vers les locus S_1A et S_1B , permettra de définir les bornes de la région minimale héritée par les gamètes S_1^s .

La stratégie pour déterminer les bornes internes des intervalles de présence de S_1A et S_1B se basera quant à elle sur l'analyse de la décroissance de la semi stérilité femelle des recombinants et de la distorsion de ségrégation des marqueurs, les deux phénomènes ayant la même cause, à savoir l'absence d'interaction entre S_1 et l'un des deux locus S_1A ou S_1B . L'hypothèse de travail est qu'un individu hétérozygote pour les trois facteurs devrait à la fois être semi-stérile et présenter une distorsion de ségrégation pour la région dans sa descendance en autofécondation, alors qu'un individu homozygote pour l'un des gènes doit récupérer l'équilibre allélique Mendélien dans sa descendance (Chapitre III; Koide et al., 2008c). A partir de rétrocroisement entre une lignée qui possède l'haplotype *glaberrima* autour de S_1 et son parent *O. sativa*, on obtiendra des individus recombinants dans la région entre S_1A et S_1B , dont la majorité seront S_1^s/S_1^g du fait de la distorsion de la ségrégation. Les individus recombinants et hétérozygotes pour S_1 seront autofécondés et évalués pour leur fertilité pollinique et paniculaire. En outre, un certain nombre de leurs descendants seront évalués avec un marqueur de S_1 afin d'évaluer sa ségrégation. Les individus qui auront perdu un des

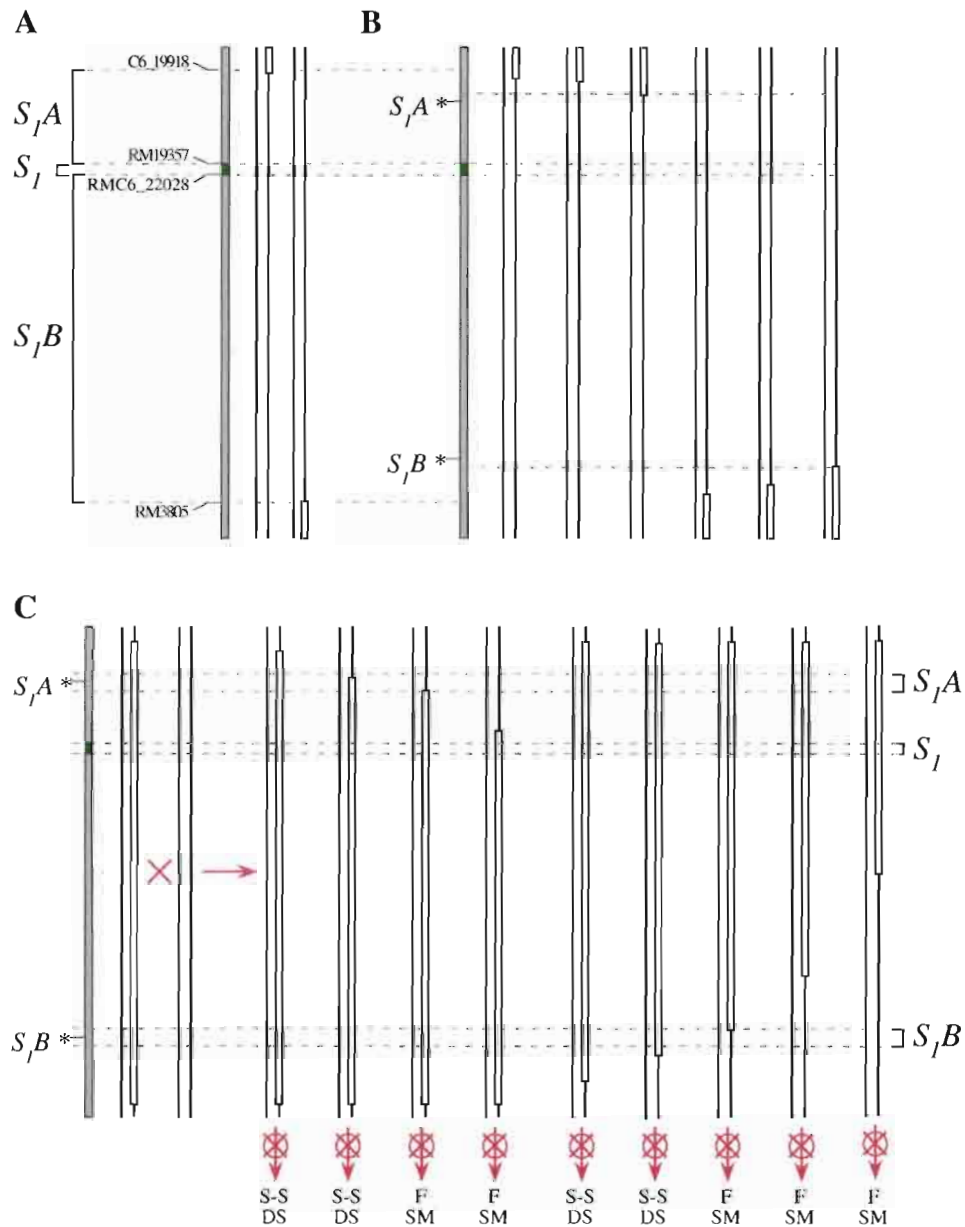


Figure VI-2. Stratégie de cartographie fine des locus S_1A et S_1B . A. Délimitation des bornes externes de la position de S_1A et S_1B à partir du fragment minimal *O. sativa* hérité par les plantes S_1^s/S_1^s . B. Délimitation des bornes internes de la position de S_1A et S_1B par l'étude de la décroissance de la distorsion de la ségrégation et de la semi stérilité femelle des recombinants. L'analyse est faite à partir des plantes S_1^s/S_1^g qui présentent une recombinaison dans la région des trois allèles identifiés. Les plantes issues d'une recombinaison qui élimine un des allèles présentent une récupération de la fertilité et n'engendre pas de distorsion de ségrégation dans leur descendance issue d'autofécondation. Les deux chromatides du bras court du chromosome 6 sont représentées pour la région des trois locus. Les barres vides et les lignes pleines représentent respectivement les chromatides provenant d'*O. glaberrima* et d'*O. sativa*. (S-S) Semi-stérile, (F) Fertile, (DS) Distorsion de la ségrégation, (SM) Ségrégation mendélienne.

facteurs $S_I A$ ou $S_I B$ seront identifiées par la récupération de la fertilité paniculaire tout en conservant la semi stérilité pollinique (du fait de l'interaction S_I^s/S_I^g), et par la transmission Mendélienne des allèles au niveau des individus F2 en proportion S_I^s/S_I^g 1 : 1 (et non 1 : 2 : 1 puisque le pollen S_I^s est non viable).

Pour cette analyse, nous disposons déjà de matériels végétaux adéquats, spécifiquement deux lignées BC₄F₁ dérivées des BC₃-DH du croisement *Caiapo* x MG12. L'une (103-CA1) possède l'haplotype *glaberrima* pour la région des trois gènes, et l'autre (85-CA1) possède un segment introgressé *O. glaberrima* plus court (85-CA1, voir Figure V-3), et peut-être recombinaison entre S_I et $S_I B$. Le fonds génétique de ces lignées étant pratiquement exempt d'autres fragments génomiques d'*O. glaberrima*, l'effet observé sur la fertilité sera en principe dû exclusivement à S_I et aux locus qui lui sont liés. De plus, nous disposons d'une lignée CSSL (M128-1-I1) dérivée du croisement IR64 x TOG5681 BC₃F₁, hétérozygote pour toute la région autour de S_I , et présentant un faible nombre d'introgessions de TOG5681, ainsi que les graines obtenues par autofécondation des plantes BC₁F₁ recombinantes autour de S_I et que nous avons étudiées au cours du présent travail (Voir Chapitre III).

Le rôle de l'inversion dans la barrière de reproduction causée par S_I devra être étudié de manière plus approfondie. A cette fin, nous proposons une analyse de sa présence chez *O. sativa*, *O. glaberrima*, leurs ancêtres, et les autres espèces du groupe du génome AA. Pour cela, un jeu de trois amorces permettant une amplification différentielle selon la présence ou l'absence de l'inversion est déjà disponible et pourra être testé sur les différentes accessions. Etant donné les nombreuses répétitions de séquence dans la région de l'inversion, il sera nécessaire de recourir à des PCR de type *long-range* permettant des amplifications de 9,5 kb.

VI.9.2 Identification et clonage du facteur S_I

Nous avons vu plus haut dans ce document que neuf gènes définissent la région où S_I a été cartographié. D'après nos analyses, nous pensons que le meilleur candidat pour S_I est le gène qui code pour une protéine avec un domaine F-box. Cependant, il est encore trop tôt pour écarter définitivement l'hypothèse d'une possible implication d'un des huit autres gènes. Des analyses plus approfondies au niveau de l'expression dans les tissus floraux des hybrides, des recombinants et des lignées parentales sont nécessaires pour déterminer l'implication de chacun de ces gènes dans le phénotype de stérilité.

Des études montrent que les gènes des F-box s'expriment dans le tissu et au stade correspondant, et causent l'élimination allélique des gamètes (Jain *et al.*, 2007). Néanmoins, il reste à déterminer la ou les copies du gène dont l'expression a été mise en évidence. Afin de

réaliser une étude d'expression par RT-PCR et Q-PCR qui permette de différencier les copies de chacun des locus chez *O. sativa* et *O. glaberrima*, nous avons désigné plusieurs amorces spécifiques de chaque allèle de chaque copie du gène sur la base de leur alignement et des SNP, que nous pourrions tester dès que le matériel végétal adéquat sera disponible.

Par ailleurs, une analyse d'hybridation *in situ* sur des ovules et anthères en méiose pourrait être envisagée. Cependant, une sonde traditionnelle obtenue par transcription *in vitro* ou par PCR pourrait ne pas être assez spécifique pour différencier les copies du gène. On pourrait envisager l'utilisation comme sonde d'oligonucléotides de 20 à 25 nt, modifiés pour contenir des acides nucléiques bloqués ou LNA (Kutter *et al.*, 2007), spécifiques pour les allèles des gènes candidats.

Si nous parvenons à établir un lien solide entre un des gènes candidats et le phénotype de stérilité, nous envisageons d'entamer un séquençage allélique de ces gènes chez plusieurs accessions d'*O. glaberrima*, *O. sativa*, les accessions hybrides naturels entre les deux espèces, et leurs ancêtres, afin de confirmer les différences entre les allèles des deux complexes (*O. sativa/O. rufipogon* et *O. glaberrima/O. barthii*), et de mieux étudier l'évolution de la barrière de reproduction.

Une autre approche qui peut être envisagée pour confirmer les gènes qui pourraient interagir avec les allèles S_I^s ou S_I^g et qui ont un rôle dans la barrière de stérilité est l'utilisation de la méthode double hybride (*two yeast hybrid*). Cette méthodologie a déjà été utilisée pour déterminer l'interaction entre les allèles des deux gènes impliqués dans la barrière de stérilité intersubspécifique causé par le locus *Sa* (Long *et al.*, 2008), et pourrait donc faciliter l'identification des protéines qui interagissent avec S_I . L'information générée par cette approche aidera à la détermination d'un candidat pour S_{IB} , ou d'un autre facteur impliqué dans la barrière et localisé dans une autre région génomique.

On peut aussi envisager de conduire des études fonctionnelles complémentaires qui viseraient à confirmer le gène candidat pour S_I . Pour cela, une approche de clonage de l'allèle S_I^g d'*O. glaberrima* et de transformation d'*O. sativa* (de préférence *japonica*, plus facile à transformer) avec cet allèle peut être utilisée. Puisque le phénotype de semi stérilité femelle fait intervenir d'autres facteurs que S_I , l'évaluation de l'effet de cet allèle sur la fertilité devra probablement être faite sur la base d'évaluations de fertilité pollinique. Par conséquent, si le produit codé par le gène S_I agit en *trans*, un phénotype d'élimination des gamètes mâles S_I^s devrait se produire, générant une haute stérilité pollinique des transformants. Une expérience complémentaire serait le clonage de l'allèle S_I de *O. sativa* et la transformation d'*O. glaberrima* avec cet allèle, mais ne sera réalisable que lorsqu'un protocole de

transformation d'*O. glaberrima* aura été mis au point. Pour l'instant, ce protocole n'est pas disponible, alors que la transformation de lignées hybrides interspécifiques de type NERICA a été obtenue avec succès (Ishizaki et Kumashiro, 2008).

Finalement, on pourrait essayer de mettre à profit les collections de mutants d'insertion de riz disponibles. Cependant, l'approche est assez risquée si l'on considère que les mutants pour S_I ou pour les autres locus en interaction avec S_I ne présenteront pas nécessairement de phénotype stérile et que des croisements mutant x *O. glaberrima* pourraient s'avérer nécessaires pour révéler de possibles effets sur la stérilité. Par exemple, une mutation pourrait modifier l'expression de l'un des facteurs épistatiques, ce qui aurait pour effet d'augmenter la fréquence de transmission des allèles *O. sativa* pour la région de S_I , et par conséquent la fertilité en croisement interspécifique. Une vingtaine de lignées T-DNA avec des FST (Flanking Sequence Tag) dans la région de S_I ont été identifiés dans la collection de mutants Oryza Tag Line (Flanking Sequence Tag) dans la région de S_I existent dans la collection de mutants Oryza Tag Line (Larmande *et al.*, 2008), dont huit présentent une insertion dans le gène qui code pour la protéine F-box. Ces lignées pourront être croisées avec *O. glaberrima* afin d'étudier la fertilité pollinique et paniculaire des hybrides. Comme le phénotype de la semi stérilité femelle dépend de l'interaction entre différents locus, il pourrait s'avérer nécessaire de recourir à la création par croisement de lignées à double ou triple mutation, afin de modifier en parallèle l'expression des différents facteurs impliqués.

VI.9.3 Cartographie fine et identification des autres locus de stérilité entre *O. sativa* et *O. glaberrima*

Même lorsque la barrière reproductive exercée par S_I est contournée, c'est à dire que l'individu interspécifique porte l'haplotype *sativa*, sa fertilité restera déterminée par sa configuration allélique à des gènes de stérilité secondaires. Il est donc important de préciser la localisation exacte de ces locus, de déterminer d'éventuelles interactions entre eux, et de quantifier leurs effets, afin de pouvoir construire des stratégies visant à rétablir totalement la fertilité des lignées introgressées. A cette fin, on pourra réaliser une caractérisation de lignées CSSL à fonds génétique *O. sativa* et portant des fragments du génome d'*O. glaberrima* incluant individuellement chacun des locus de stérilité répertoriés ainsi que les régions de distorsion de ségrégation de nos populations interspécifiques. Une corrélation entre la fertilité pollinique ou paniculaire et le génotype de ces lignées apporterait des renseignements sur l'effet de chaque région sur la fertilité hybride. De la même façon, l'analyse des lignées qui portent des combinaisons d'au moins deux de ces locus permettrait d'analyser les interactions

possibles à l'origine d'une baisse de fertilité et/ou d'une distorsion de ségrégation, par évaluation du déséquilibre de liaison entre des régions du génome non liées physiquement (Torjek *et al.*, 2006), et corrélation avec la fertilité des plantes. Nous avons réalisé une évaluation préliminaire de l'effet de certains segments chromosomiques sur la fertilité pollinique et paniculaire, et mis en évidence une association possible entre des régions des chromosome 7 et 10 (données non présentées dans ce document). La poursuite de ces évaluations, avec un génotypage plus fin que celui basé sur la carte UCGM, pourrait permettre d'apporter une vision plus globale du contrôle génétique de la barrière reproductive entre *O. sativa* et *O. glaberrima*, et cette connaissance pourra être utilisée en amélioration, dans des stratégies du type « iBridges » afin de maximiser la fertilité hybride des lignées interspécifiques avec *O. sativa*.

VI.9.4 Etude des hybrides naturels

Une autre voie à explorer pour caractériser la barrière de reproduction entre les deux espèces de riz cultivé est l'étude des hybrides naturels ou *admixture*. Ces accessions, qui résultent très probablement d'hybridations spontanées, ont été soumises à une sélection naturelle et anthropique pendant des périodes plus ou moins longues – pouvant atteindre plusieurs centaines d'années – qui leur a permis de récupérer une bonne autofertilité. Leur étude pourrait donc générer de précieuses informations sur les régions du génome contrôlant à la fois l'autofertilité et la fertilité en croisement avec *O. sativa* et *O. glaberrima*.

L'approche pour leur étude pourrait commencer par la recherche d'un plus grand nombre d'accessions de type *admixture*, à partir des matériels décrits précédemment (Semon *et al.*, 2005; Barry *et al.*, 2007; Nuijten *et al.*, 2009), et dans les différentes collections d'accessions d'*O. glaberrima* et *O. sativa* disponibles. Une évaluation de la fertilité pollinique et paniculaire des plantes F₁ obtenues par croisement de ces lignées avec *O. sativa*, *O. glaberrima* et entre elles, pourrait être réalisée, ainsi que le génotypage fin au moyen d'une plateforme à haut débit, ou même un reséquençage de leur génome. Sur la base de ces données, une analyse de cartographie d'association pourrait permettre de mieux comprendre la relation entre la fertilité en croisement et les allèles aux gènes de stérilité ou aux sites de distorsion de ségrégation.

Nous avons vu plus haut (voir Chapitre V) que les accessions *admixture* qui possèdent un haplotype *sativa*-2 dans la région de *S₁* présentent des allèles différents à ceux des variétés modèles d'*O. sativa* (haplotype *sativa*-1). Il serait donc intéressant de chercher l'origine de cet haplotype *sativa*-2, afin de tester l'hypothèse d'une meilleure compatibilité allélique de cet

haplotype avec l'haplotype *glaberrima*. Pour ce faire, un criblage dans les collections de variétés d'*O. sativa* traditionnelles africaines, puis des croisements entre ces lignées avec *O. glaberrima*, suivi de l'évaluation du pourcentage de transmission de S_I^s et de la fertilité paniculaire et pollinique des descendants, pourraient permettre de mieux comprendre les événements d'introggression qui ont conduit à favoriser cet haplotype dans les *admixture*.

VII

RÉFÉRENCES BIBLIOGRAPHIQUES

- Abe T., Edanami T., Adachi E., Sasahara T. 1999.** Phylogenetic relationships in the genus *Oryza* based on mitochondrial RFLPs. *Genes & Genetic Systems* **74**(1): 23-27.
- Abramoff M. D., Magelhaes P. J., Ram S. J. 2004.** Image processing with ImageJ. *Biophotonics International* **11**(7): 36-42.
- Abril J. F., Guigo R. 2000.** gff2ps: visualizing genomic annotations. *Bioinformatics* **16**(8): 743-744.
- Ahn S. N., Kang K. H., Suh J. P., Kwon S. J., Moon H. P., Choi H. C., McCouch S. R. 2003.** Advanced backcross analysis for transferring QTLs from *O. rufipogon*. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 126-127.
- Akimoto M., Morishima H. 2003.** Genetic population structures of *Oryza glumaepatula* and *O. grandiglumis* distributed in the Amazon flood area. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 126-127.
- Akimoto M., Shimamoto Y., Morishima H. 1998.** Population genetic structure of wild rice *Oryza glumaepatula* distributed in the Amazon flood area influenced by its life-history traits. *Molecular Ecology* **7**(10): 1371-1381.
- Albar L., Ndjioudjop M. N., Esshak Z., Berger A., Pinel A., Jones M., Fargette D., Ghesquière A. 2003.** Fine genetic mapping of a gene required for *Rice Yellow Mottle Virus* cell-to-cell movement. *Theoretical and Applied Genetics* **107**(2): 371-378.
- Almeida A., Giraldo O., Martínez C. P., Borrero J., Duque M. C., López J., Tohme J. 2004.** Poster - Identification of QTLs for yield and yield components in a backcross population between *Oryza barthii* and Lemont. En. *RedBio*. Santo Domingo, Republique Dominicaine.
- Aluko G., Martinez C., Tohme J., Castano C., Bergman C., Oard J. H. 2004.** QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* x *O. glaberrima*. *Theoretical and Applied Genetics* **109**(3): 630-639.
- Amante-Bordeos A., Sitch L. A., Nelson R., Dalmacio R. D., Oliva N. P., Aswidinnoor H., Leung H. 1992.** Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*. *Theoretical and Applied Genetics* **84**(3): 345-354.
- Ammiraju J. S. S., Lu F., Sanyal A., Yu Y., Song X., Jiang N., Pontaroli A. C., Rambo T., Currie J., Collura K., Talag J., et al. 2008.** Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* **20**(12): 3191-3209.
- Ampong-Nyarko K., Datta S. K. D. 1991.** *A handbook for weed control in rice*. Manila, Philippines: International Rice Research Institute
- Andrew J. F., Maggie R. K., Stephen R. P., Ellis T. H. N. 1998.** Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant Journal* **16**(5): 643-650.
- Asch F., Dingkuhn M., Sow A., Audebert A. 2005.** Drought-induced changes in rooting patterns and assimilate partitioning between root and shoot in upland rice. *Field Crops Research* **93**(2-3): 223-236.
- Audebert A., Fofana M. 2009.** Rice yield gap due to iron toxicity in West Africa. *Journal of Agronomy and Crop Science* **195**(1): 66-76.
- Barbash D. A., Awadalla P., Tarone A. M. 2004.** Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biology* **2**(6): e142.

- Barry M. B., Pham J. L., Noyer J. L., Billot C., Courtois B., Ahmadi N. 2007. Genetic diversity of the two cultivated rice species (*O. sativa* & *O. glaberrima*) in Maritime Guinea. Evidence for interspecific recombination. *Euphytica* 154(1): 127-137.
- Bateson W. 1909. Heredity and variation in modern lights. En: Seward A. C. ed. *Darwin and modern science*. Cambridge: Cambridge University Press, 81-101.
- Bautista N. S., Shimadzutsu K., Teranishi T., Takamatsu S., Kobayashi N., Uchida N., Kamijima O., Ishii T. 2003. Trait-improving wild QTL alleles identified using advanced backcross QTL analysis from a cross between cultivated rice, *Oryza sativa*, and wild rice, *O. rufipogon*. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 135-138.
- Bautista N. S., Solis R., Kamijima O., Ishii T. 2001. RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. *Genes & Genetic Systems* 76(2): 71-79.
- Bezançon G. 1994. *Le riz cultivé d'origine africaine Oryza glaberrima Steud. et les formes sauvages et adventices apparentées : diversité, relations génétiques et domestication*. Paris.
- Bezançon G., Bozza J., Koffi G., Second G. 1977. Diversité génétique d'*O. glaberrima* et *O. breviligulata* en observation directe et par électrophorèses d'isozymes. En ORSTOM I. *Les Espèces Africaines de Riz* Paris. 15-46.
- Bezançon G., Causse M., Ghesquière A., Kochko A. d., Pham J.-L., Gérard. S. 1989. Les riz en Afrique: diversité génétique, relations interspécifiques et évolution. *Bulletin de la Société Botanique de France. Actualités Botaniques* 136(3-4): 251-262.
- Bikard D., Patel D., Le Mette C., Giorgi V., Camilleri C., Bennett M. J., Loudet O. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* 323(5914): 623-626.
- Blumenstiel J. P., Hartl D. L. 2005. Evidence for maternally transmitted small interfering RNA in the repression of transposition in *Drosophila virilis*. *Proceedings of the National Academy of Sciences of the United States of America* 102(44): 15965-15970.
- Bomblies K., Lempe J., Eppe P., Warthmann N., Lanz C., Dangl J. L., Weigel D. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-Type incompatibility syndrome in plants. *PLoS Biology* 5(9): e236.
- Bomblies K., Weigel D. 2007. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews: Genetics* 8(5): 382-393.
- Bougerol B., Pham J. L. 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(5): 810-815.
- Bouharmont J., Olivier M., Chassart M. D. 1985. Cytological observations in some hybrids between the rice species *Oryza sativa* L. and *O. glaberrima* Steud. *Euphytica* 34(1): 75-81.
- Bradshaw H. D., Jr., Otto K. G., Frewen B. E., McKay J. K., Schemske D. W. 1998. Quantitative Trait Loci affecting differences in floral morphology between two species of Monkeyflower (*Mimulus*). *Genetics* 149(1): 367-382.
- Bradshaw H. D., Wilbert S. M., Otto K. G., Schemske D. W. 1995. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* 376(6543): 762-765.
- Brar D. S. 2005. Broadening gene pool of rice through introgression from wild species. En: *Rice is life. Proceedings of the World Rice Research Conference 2004*. Tsukuba. 157-160.

- Brar D. S., Dalmacio R., Elloran R., Aggarwal R., Angeles R., Khush G. S. 1996.** Gene transfer and molecular characterization of introgression from wild *Oryza* species into rice. En: Khush G. S. ed. *Rice Genetics III*. Manila, Philippines: IRRI, 477-486.
- Brar D. S., Elloran R., Khush G. S. 1991.** Interspecific hybrids produced through embryo rescue between cultivated and eight wild species of rice. *Rice Genetics Newsletter* **8**: 91-92.
- Brar D. S., Elloran R. M., Talag J. D., Abbasi F., Khush G. S. 1997.** Cytogenetic and molecular characterization of an intergeneric hybrid between *Oryza sativa* L. and *Porteresia coarctata* (Roxb.) Tateoka. *Rice Genetics Newsletter* **14**: 43-44.
- Brar D. S., Khush G. S. 1997.** Alien introgression in rice. *Plant Molecular Biology* **35**(1-2): 35-47.
- Brar D. S., Khush G. S. 2002.** Transferring genes from wild species into rice. En: Kang M. S. ed. *Quantitative genetics, genomics and plant breeding*. Wallingford (UK): CAB International, 197-217.
- Brar D. S., Khush G. S. 2006.** Cytogenetic manipulation and germplasm enhancement of rice (*Oryza sativa* L.). En: Singh R. J., Jauhar P. P. eds. *Genetic Resources, Chromosome Engineering, and Crop Improvement: Cereals, Volume 2*: CRC Press, 115-158.
- Brennecke J., Malone C. D., Aravin A. A., Sachidanandam R., Stark A., Hannon G. J. 2008.** An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**(5906): 1387-1392.
- Brideau N. J., Flores H. A., Wang J., Maheshwari S., Wang X., Barbash D. A. 2006.** Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* **314**(5803): 1292-1295.
- Brondani C., Brondani R. P. V., Rangel P. H. N., Ferreira M. E. 2001.** Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. *Hereditas* **134**(1): 59-71.
- Brondani C., Rangel P., Brondani R., Ferreira M. 2002.** QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers. *Theoretical and Applied Genetics* **104**(6): 1192-1203.
- Byzova M. V., Franken J., Aarts M. G. M., de Almeida-Engler J., Engler G., Mariani C., Van Lookeren Campagne M. M., Angenent G. C. 1999.** *Arabidopsis* STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes & Development* **13**(8): 1002-1014.
- Cai H.-W., Akimoto M., Morishima H. 2008.** Genetic diversity in wild relatives of rice and domestication events. *Rice Biology in the Genomics Era*, 261-275.
- Calvert L., Sanint L. R., Châtel M., J. I. 2006.** Rice production in Latin America at critical crossroads. *International Rice Commission Newsletter (FAO)* **55**: 65-73.
- Carbonell R., Yanis A., Rubial A., Pereiro G., Sousa J., Correa F., Calvert L. 2001.** *Guía para el trabajo de campo en el manejo integrado de plagas del arroz*. Cali, Colombia: CIAT.
- Carver T. J., Rutherford K. M., Berriman M., Rajandream M. A., Barrell B. G., Parkhill J. 2005.** ACT: the Artemis comparison tool. *Bioinformatics* **21**(16): 3422-3423.
- Cattani M. V., Presgraves D. C. 2009.** Genetics and lineage-specific evolution of a lethal hybrid incompatibility between *Drosophila mauritiana* and its sibling species. *Genetics* **181**(4): 1545-1555.
- Causse M., Ghesquiere A. 1991.** Prospective use of *Oryza longistaminata* for rice breeding. En: IRRI ed. *Rice Genetics II*. Manila, Philippines IRRI, 81-89.

- Causse M. A., Fulton T. M., Cho Y. G., Ahn S. N., Chunwongse J., Wu K. S., Xiao J. H., Yu Z. H., Ronald P. C., Harrington S. E., Second G., et al. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138(4): 1251-1274.
- Chaparro C., Guyot R., Zuccolo A., Piegu B., Panaud O. 2007. RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Research* 35(suppl_1): D66-70.
- Cheema K. K., Grewal N. K., Vikal Y., Sharma R., Lore J. S., Das A., Bhatia D., Mahajan R., Gupta V., Bharaj T. S., Singh K. 2008. A novel bacterial blight resistance gene from *Oryza nivara* mapped to 38-kb region on chromosome 4L and transferred to *Oryza sativa* L. *Genetics Research* 90(05): 397-407.
- Chen J., Ding J., Ouyang Y., Du H., Yang J., Cheng K., Zhao J., Qiu S., Zhang X., Yao J., Liu K., et al. 2008. A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of *indica-japonica* hybrids in rice. *Proceedings of the National Academy of Sciences* 105(32): 11436-11441.
- Chen Z., Zhuang C., Mei M., Liu Y. 2004. Genetic and physical fine-mapping of the *Sc* locus conferring *indica-japonica* hybrid sterility in rice (*Oryza sativa* L.). *Chinese Science Bulletin* 49(16): 1718-1721.
- Cheng C. Y., Tsuchimoto S., Ohtsubo H., Ohtsubo E. 2002. Evolutionary relationships among rice species with AA genome based on SINE insertion analysis. *Genes & Genetic Systems* 77(5): 323-334.
- Childs N., Kiawu J. 2009. Factors behind the rise in global rice prices in 2008. En: USDA, Economic Research Service, Outlook Report No. RCS-09D-01. 25 pp.
- Christie P., Macnair M. R. 1987. The distribution of postmating reproductive isolating genes in populations of the Yellow Monkey Flower, *Mimulus guttatus*. *Evolution* 41(3): 571-578.
- Chu Y.-E., Morishima H., Oka H.-I. 1969. Reproductive barriers distributed in cultivated rice species and their wild relatives. *Japanese Journal of Genetics* 44(4): 207-223.
- Chu Y.-E., Oka H.-I. 1970. The genetic basis of crossing barriers between *Oryza perennis* Subsp. *barthii* and its related taxa. *Evolution* 24(1): 135-144.
- Chu Y. E., Oka H. I. 1972. The distribution and effects of genes causing F1 weakness in *Oryza breviligulata* and *O. glaberrima*. *Genetics* 70(1): 163-173.
- Coyne J. A., Orr H. A. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.
- Dharmasiri N., Dharmasiri S., Estelle M. 2005. The F-box protein TIR1 is an auxin receptor. *Nature* 435(7041): 441-445.
- Dingkuhn M., Audebert A. Y., Jones M. P., Etienne K., Sow A. 1999. Control of stomatal conductance and leaf rolling in *O. sativa* and *O. glaberrima* upland rice. *Field Crops Research* 61(3): 223-236.
- Dobzhansky T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* 21(2): 113-135.
- Doi K., Izawa T., Fuse T., Yamanouchi U., Kubo T., Shimatani Z., Yano M., Yoshimura A. 2004. *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes & Development* 18(8): 926-936.
- Doi K., Sobrizal, Ikeda K., Sanchez P. L., Kurakazu T., Nagai Y., Yoshimura A. 2003a. Developing and evaluating rice chromosome segment substitution lines. En: Mew T. W., Joshi R. C., Brar D. S. eds. *Rice Science: Innovations and Impact for Livelihood*, 289-296.
- Doi K., Taguchi K., Yoshimura A. 1998a. A new locus affecting high F₁ pollen sterility found in backcross progenies of *japonica* rice and African rice. *Rice Genetics Newsletter* 15: 146-148.

- Doi K., Taguchi K., Yoshimura A. 1999.** RFLP mapping of *S20* and *S21* for F_1 pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. *Rice Genetics Newsletter* **16**: 65-68.
- Doi K., Taguchi K., Yoshimura A. 2003b.** Genetic analysis of pollen sterility loci found in hybrid progeny between *Oryza sativa* and *O. glaberrima*. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 149-151.
- Doi K., Yoshimura A., Iwata N. 1998b.** RFLP mapping and QTL analysis of heading date and pollen sterility using backcross populations between *Oryza sativa* L. and *Oryza glaberrima* Steud. *Breeding Science* **48**(4): 395-399.
- Droc G., Perin C., Fromentin S., Larmande P. 2009.** OryGenesDB 2008 update: database interoperability for functional genomics of rice. *Nucleic Acids Research* **37**(suppl_1): D992-995.
- Duan S. H., Lu B. R., Li Z., Tong J. P., Kong J., Yao W., Li S. Q., Zhu Y. G. 2007.** Phylogenetic analysis of AA-genome *Oryza* species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochemical Genetics* **45**(1-2): 113-129.
- Durfee T., Roe J. L., Sessions R. A., Inouye C., Serikawa K., Feldmann K. A., Weigel D., Zambryski P. C. 2003.** The F-box-containing protein UFO and AGAMOUS participate in antagonistic pathways governing early petal development in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **100**(14): 8571-8576.
- Eizenga G., Lee F., Jia Y. 2004.** Identification of Blast resistance genes in rice wild relatives (*Oryza* spp.) and newly introduced rice (*O. sativa*) lines En: Norman R. J., Meullenet J.-F., Moldenhauer K. A. K. eds. *BR Wells Rice Research Studies 2003*,. Arkansas: University of Arkansas Agricultural Experiment Station Research Series, 29-36.
- Eizenga G. C., Lee F. N., Rutger J. N. 2002.** Screening *Oryza* species plants for rice sheath blight resistance. *Plant Disease* **86**(7): 808-812.
- Escobar-Restrepo J.-M., Huck N., Kessler S., Gagliardini V., Gheyselinck J., Yang W.-C., Grossniklaus U. 2007.** The FERONIA Receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* **317**(5838): 656-660.
- FAO 2009.** More people than ever are victims of hunger.En. *FAO News Release*, 19 Juin. Rome.
- Fishman L., Aagaard J., Tuthill J. C. 2008.** Toward the evolutionary genomics of gametophytic divergence: Patterns of transmission ratio distortion in Monkeyflower *Mimulus* hybrids reveal a complex genetic basis for conspecific pollen precedence. *Evolution* **62**(12): 2958-2970.
- Fishman L., Willis J. H. 2001.** Evidence for Dobzhansky-Muller incompatibilities contributing to the sterility of hybrids between *Mimulus guttatus* and *M. nasutus*. *Evolution* **55**(10): 1932-1942.
- Fishman L., Willis J. H. 2005.** A novel meiotic drive locus almost completely distorts segregation in *Mimulus* (Monkeyflower) hybrids. *Genetics* **169**(1): 347-353.
- Fu X., Lu Y., Liu X., Li J., Feng J. 2007.** Cytological mechanisms of interspecific incrossability and hybrid sterility between *Oryza sativa* L. and *O. alta* Swallen. *Chinese Science Bulletin* **52**(6): 755-765.
- Fukuoka S., Namai H., Okuno K. 1998.** RFLP mapping of the genes controlling hybrid breakdown in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **97**(3): 446-449.
- Fukuoka S., Newingham M. C. V., Ishtiaq M., Nanamine T., Kawase M., Okuno K. 2005.** Identification and mapping of two new loci for hybrid breakdown in cultivated rice. *Rice Genetics Newsletter* **22**: 29-31.

- Fuller D. Q., Qin L., Zheng Y., Zhao Z., Chen X., Hosoya L. A., Sun G.-P. 2009. The domestication process and domestication rate in rice: Spikelet bases from the lower Yangtze. *Science* **323**(5921): 1607-1610.
- Futakuchi K., Watanabe H., Jones M. P. 2008. Relationship of grain protein content to other grain quality traits in interspecific *Oryza sativa* L. X *Oryza glaberrima* Steud. Progenies. *Agricultural Journal* **3**(1): 50-57.
- Garavito A. 2003. *Asociación del silenciamiento post-transcripcional de genes y la reacción de hipersensibilidad, con la resistencia transgénica al Virus de la Hoja Blanca del Arroz*. Thèse de Master, Universidad de los Andes. Bogota, Colombie. Soutenue en Janvier 2003.
- Garavito A., Guyot R., Lozano J., Gavory F., Samain S., Panaud O., Tohme J., Ghesquière A., Lorieux M. 2009. A genetic model for the female sterility barrier between Asian and African cultivated rice species. *Genetics* **Submitted**.
- Garris A. J., Tai T. H., Coburn J., Kresovich S., McCouch S. 2005. Genetic structure and diversity in *Oryza sativa* L. *Genetics* **169**(3): 1631-1638.
- Ge S., Sang T., Lu B. R., Hong D. Y. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proceedings of the National Academy of Sciences of the United States of America* **96**(25): 14400-14405.
- Ge S., Sang T., Lu B. R., Hong D. Y. 2001. Phylogeny of the genus *Oryza* as revealed by molecular approaches. En: Khush G. S., Brar D. S., Hardy B. eds. *Rice genetics IV*. Los Banos, Philippines: IRRI, 89-105.
- Ghesquière A. 1988. *Diversité génétique de l'espèce sauvage de riz, Oryza longistaminata A.Chev. et Roehr et dynamique des flux géniques au sein du groupe Sativa en Afrique*. Université Paris XI Orsay.
- Ghesquière A. 1991. Reexamination of genetic control of the reproductive barrier between *Oryza longistaminata* and *O. sativa* and relationship to rhizome expression. En: IRRI ed. *Rice Genetics II*. Manila, Philippines: IRRI, 729-730.
- Ghesquière A. 1986. Evolution of *Oryza longistaminata* En: IRRI ed. *Rice Genetics*. Manila, 15-25.
- Ghesquière A., Causs M. 1992. Linkage study between molecular markers and genes controlling the reproductive barrier in interspecific backcross between *O. sativa* and *O. longistaminata*. *Rice Genetics Newsletter* **9**: 28-31.
- Ghesquière A., Sequier J., Second G., Lorieux M. 1997. First steps towards a rational use of African rice, *Oryza glaberrima*, in rice breeding through a 'contig line' concept. *Euphytica* **96**(1): 31-39.
- Glaszmann J. C. 1987. Isozymes and classification of Asian rice varieties. *Theoretical and Applied Genetics* **74**(1): 21-30.
- Gnanamanickam S. S. 2009. Major diseases of rice. *Biological Control of Rice Diseases*, 13-42.
- Gusti A., Baumberger N., Nowack M., Pusch S., Eisler H., Potuschak T., De Veylder L., Schnittger A., Genschik P. 2009. The *Arabidopsis thaliana* F-Box protein FBL17 is essential for progression through the second mitosis during pollen development. *PLoS ONE* **4**(3): e4780.
- Haiyuan Y., Xiang R. E. N., Qingmei W., Lili Z. H. U., Guangcun H. E. 2002. Molecular mapping and genetic analysis of a rice brown planthopper (*Nilaparvata lugens* Stal) resistance gene. *Hereditas* **136**(1): 39-43.
- Hanada K., Vallejo V., Nobuta K., Slotkin R. K., Lisch D., Meyers B. C., Shiu S.-H., Jiang N. 2009. The Functional role of Pack-MULEs in rice inferred from purifying selection and expression profile. *Plant Cell* **21**(1): 25-38.

- Harushima Y., Kurata N., Yano M., Nagamura Y., Sasaki T., Minobe Y., Nakagahra M. 1996.** Detection of segregation distortions in an *indica-japonica* rice cross using a high-resolution molecular map. *Theoretical and Applied Genetics* **92**(2): 145-150.
- Harushima Y., Nakagahra M., Yano M., Sasaki T., Kurata N. 2002.** Diverse variation of reproductive barriers in three intraspecific rice crosses. *Genetics* **160**(1): 313-322.
- Hayashi Y., Kyojuka J., Shimamoto K. 1988.** Hybrids of rice (*Oryza sativa* L.) and wild *Oryza* species obtained by cell fusion. *Molecular and General Genetics MGG* **214**(1): 6-10.
- Heuer S., Miezian K. M. 2003.** Assessing hybrid sterility in *Oryza glaberrima* x *O. sativa* hybrid progenies by PCR marker analysis and crossing with wide compatibility varieties. *Theoretical and Applied Genetics* **107**(5): 902-909.
- Hirabayashi H., Kaji R., Okamoto M., Ogawa T., Brar D. S., Angeles E. R., Khush G. S. 2003.** Mapping QTLs for brown planthopper (BPH) resistance introgressed from *Oryza officinalis* in rice. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 268-270.
- Hoballah M. E., Gubitzi T., Stuurman J., Broger L., Barone M., Mandel T., Dell'Olivo A., Arnold M., Kuhlemeier C. 2007.** Single gene-mediated shift in pollinator attraction in *Petunia*. *Plant Cell* **19**(3): 779-790.
- Hodnett G. L., Burson B. L., Rooney W. L., Dillon S. L., Price H. J. 2005.** Pollen-pistil interactions result in reproductive isolation between *Sorghum bicolor* and divergent *Sorghum* species. *Crop Science* **45**(4): 1403-1409.
- Hu F. Y., Xu P., Deng X. N., Zhou J. W., Li J., Tao D. Y. 2006.** Molecular mapping of a pollen killer gene *S29(t)* in *Oryza glaberrima* and co-linear analysis with *S22* in *O. glumaepatula*. *Euphytica* **151**(3): 273-278.
- Huggan R. D. 1995.** Co-evolution of rice and humans. *GeoJournal* **35**(3): 262-265.
- Ichitani K., Namigoshi K., Sato M., Taura S., Aoki M., Matsumoto Y., Saitou T., Marubashi W., Kuboyama T. 2007.** Fine mapping and allelic dosage effect of *Hwcl*, a complementary hybrid weakness gene in rice. *Theoretical and Applied Genetics* **114**(8): 1407-1415.
- Ikeda R., Sokei Y., Akintayo I. 2009.** Seed fertility of F1 hybrids between upland rice NERICA cultivars and *Oryza sativa* L. or *O. glaberrima* Steud. *Breeding Science* **59**(1): 27-35.
- Inai S., Ishikawa K., Nunomura O., Ikehashi H. 1993.** Genetic analysis of stunted growth by nuclear-cytoplasmic interaction in interspecific hybrids of *Capsicum* by using RAPD markers. *Theoretical and Applied Genetics* **87**(4): 416-422.
- IRGSP International Rice Genome Sequencing P. 2005.** The map-based sequence of the rice genome. *Nature* **436**(7052): 793-800.
- Ishii T., Brar D. S., Multani D. S., Khush G. S. 1994.** Molecular tagging of genes for brown planthopper resistance and earliness introgressed from *Oryza australiensis* into cultivated rice, *O. sativa*. *Genome* **37**(2): 217-221.
- Ishii T., Nakano T., Maeda H., Kamijima O. 1996.** Phylogenetic relationships in A-genome species of rice as revealed by RAPD analysis. *Genes & Genetic Systems* **71**(4): 195-201.
- Ishizaki T., Kumashiro T. 2008.** Genetic transformation of NERICA, interspecific hybrid rice between *Oryza glaberrima* and *O. sativa*, mediated by *Agrobacterium tumefaciens*. *Plant Cell Reports* **27**(2): 319-327.
- Iwata N., Nagamatsu T., Omura T. 1964.** Abnormal segregation of waxy and apiculus coloration by a gametophyte gene belonging to the first linkage group in rice. *Japanese Journal of Breeding* **14**(1): 33-39.

- Jain M., Nijhawan A., Arora R., Agarwal P., Ray S., Sharma P., Kapoor S., Tyagi A. K., Khurana J. P. 2007.** F-Box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiology* **143**(4): 1467-1483.
- Jelodar N. B., Blackhall N. W., Hartman T. P. V., Brar D. S., Khush G., Davey M. R., Cocking E. C., Power J. B. 1999.** Intergeneric somatic hybrids of rice [*Oryza sativa* L. (+) *Porteresia coarctata* (Roxb.) Tateoka]. *Theoretical and Applied Genetics* **99**(3): 570-577.
- Jena K. K., Khush G. S. 1990.** Introgression of genes from *Oryza officinalis* Well ex Watt to cultivated rice, *O. sativa* L. *Theoretical and Applied Genetics* **80**(6): 737-745.
- Ji Q., Lu J., Chao Q., Gu M., Xu M. 2005.** Delimiting a rice wide-compatibility gene S5 n to a 50 kb region. *Theoretical and Applied Genetics* **111**(8): 1495-1503.
- Jiang N., Bao Z., Zhang X., Eddy S. R., Wessler S. R. 2004.** Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **431**(7008): 569-573.
- Jiang W., Chu S.-H., Piao R., Chin J.-H., Jin Y.-M., Lee J., Qiao Y., Han L., Piao Z., Koh H.-J. 2008.** Fine mapping and candidate gene analysis of hwh1 and hwh2, a set of complementary genes controlling hybrid breakdown in rice. *Theoretical and Applied Genetics* **116**(8): 1117-1127.
- Jing W., Zhang W. W., Jiang L., Chen L. M., Zhai H. Q., Wan J. M. 2007.** Two novel loci for pollen sterility in hybrids between the weedy strain Ludao and the Japonica variety Akihikari of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **114**(5): 915-925.
- Johnson D. E., Dingkuhn M., Jones M. P., Mahamane M. C. 1998.** The influence of rice plant type on the effect of weed competition on *Oryza sativa* and *Oryza glaberrima*. *Weed Research* **38**(3): 207-216.
- Jones M., Dingkuhn M., Aluko G., Semon M. 1997.** Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. *Euphytica* **94**(2): 237-246.
- Jones M., Yacouba S., Nwilene F., Beye A., Dobo M., Jalloh A. 2001.** *Oryza glaberrima* x *Oryza sativa* interspecifics for Africa. En: Rockwood W. G. ed. *Rice research and production in the 21st century*. Manila, Philippines: International Rice Research Institute, 137-149.
- Josefsson C., Dilkes B., Comai L. 2006.** Parent-dependent loss of gene silencing during interspecies hybridization. *Current Biology* **16**(13): 1322-1328.
- Joshi S. P., Gupta V. S., Aggarwal R. K., Ranjekar P. K., Brar D. S. 2000.** Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theoretical and Applied Genetics* **100**(8): 1311-1320.
- Jung K. H., Han M. J., Lee Y. S., Kim Y. W., Hwang I., Kim M. J., Kim Y. K., Nahm B. H., An G. 2005.** Rice Undeveloped Tapetum1 is a major regulator of early tapetum development. *Plant Cell* **17**(10): 2705-2722.
- Jurka J., Kapitonov V. V., Pavlicek A., Klonowski P., Kohany O., Walichiewicz J. 2005.** Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* **110**(1-4): 462-467.
- Kakeda K. 2009.** S locus-linked F-box genes expressed in anthers of *Hordeum bulbosum*. *Plant Cell Reports* **28**(9): 1453-1460.
- Kawano N., Ito O., Sakagami J. I. 2008.** Flash flooding resistance of rice genotypes of *Oryza sativa* L., *O. glaberrima* Steud., and Interspecific hybridization. *Environmental and Experimental Botany* **63**(1): 9-18.

- Khush G. S. 1979.** Genetics of and breeding for resistance to the brown planthopper. *Brown planthopper: Threat to rice production in Asia*. Manila, Philippines: IRRI, 321-332.
- Khush G. S., Bacalango E., Ogawa T. 1990.** A new gene for resistance to bacterial Blight from *O. longistaminata*. *Rice Genetics Newsletter* 7: 121-122.
- Khush G. S., Ling K. C. 1974.** Inheritance of resistance to *Grassy Stunt Virus* and its vector in rice. *Journal of Heredity* 65(3): 135-136.
- Kikuchi S., Satoh K., Nagata T., Kawagashira N., Doi K., Kishimoto N., Yazaki J., Ishikawa M., Yamada H., Ooka H., Hotta I., et al. 2003.** Collection, mapping, and annotation of over 28,000 cDNA clones from japonica Rice. *Science* 301(5631): 376-379.
- Kim H., Hurwitz B., Yu Y., Collura K., Gill N., SanMiguel P., Mullikin J., Maher C., Nelson W., Wissotski M., Braidotti M., et al. 2008.** Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*. *Genome Biology* 9(2): R45.
- Kinoshita T. 1993.** Registration of new gene symbols. *Rice Genetics Newsletter* 10: 7-10.
- Kinoshita T., Takamure I. 1984.** Inheritance and linkage relationship on zebra chlorosis and zebra necrosis in rice: Genetical studies on rice plant. *Journal of the Faculty of Agriculture Hokkaido University* 61(4): 445-455.
- Kishine M., Suzuki K., Nakamura S., Ohtsubo K. Ä. 2008.** Grain qualities and their genetic derivation of 7 New Rice for Africa (NERICA) Varieties. *Journal of Agricultural and Food Chemistry* 56(12): 4605-4610.
- Kobayashi N., Ikeda R., Domingo I. T., Vaughan D. A. 1993.** Resistance to infection of *Rice Tungro Viruses* and vector resistance in wild species of Rice (*Oryza* spp.). *Japanese Journal of Breeding* 43(3): 377-387.
- Koide Y., Ikenaga M., Sawamura N., Nishimoto D., Matsubara K., Onishi K., Kanazawa A., Sano Y. 2008a.** The evolution of sex-independent transmission ratio distortion involving multiple allelic interactions at a single locus in rice. *Genetics* 180(1): 409-420.
- Koide Y., Ikenaga M., Shinya Y., Matsubara K., Sano Y. 2008b.** Two loosely linked genes controlling the female specificity for cross-incompatibility in rice. *Euphytica* 164(3): 753-760.
- Koide Y., Onishi K., Nishimoto D., Baruah A. R., Kanazawa A., Sano Y. 2008c.** Sex-independent transmission ratio distortion system responsible for reproductive barriers between Asian and African rice species. *New Phytologist* 179(3): 888-900.
- Kouassi N. K., N'Guessan P., Albar L., Fauquet C. M., Brugidou C. 2005.** Distribution and Characterization of *Rice Yellow Mottle Virus*: A threat to African farmers. *Plant Disease* 89(2): 124-133.
- Kubo T., Eguchi M., Yoshimura A. 2000.** A new gene for F1 pollen sterility in *japonica/indica* cross of rice. *Rice Genetics Newsletter* 17: 63-64.
- Kubo T., Eguchi M., Yoshimura A. 2001.** A new gene for F1 pollen sterility located on chromosome 12 in *japonica/indica* cross of rice. *Rice Genetics Newsletter* 18: 54-55.
- Kubo T., Yamagata Y., Eguchi M., Yoshimura A. 2008.** A novel epistatic interaction at two loci causing hybrid male sterility in an inter-subspecific cross of rice (*Oryza sativa* L.). *Genes & Genetic Systems* 83(6): 443-453.
- Kubo T., Yoshimura A. 1998.** Linkage analysis of a gene controlling F2 sterility in Japonica/Indica backcrossed progenies of rice. *Rice Genetics Newsletter* 15: 149-151.
- Kubo T., Yoshimura A. 2001.** Linkage analysis of an F1 sterility gene in *japonica/indica* cross of rice. *Rice Genetics Newsletter* 18: 52-54.
- Kubo T., Yoshimura A. 2002.** Genetic basis of hybrid breakdown in a *japonica/indica* cross of rice, *Oryza sativa* L. *Theoretical and Applied Genetics* 105(6-7): 906-911.

- Kubo T., Yoshimura A. 2005.** Epistasis underlying female sterility detected in hybrid breakdown in a *japonica-indica* cross of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **110**(2): 346-355.
- Kuboyama T., Saito T., Matsumoto T., Wu J., Kanamori H., Taura S., Sato M., Marubashi W., Ichitani K. 2009.** Fine mapping of *HWC2*, a complementary hybrid weakness gene, and haplotype analysis around the locus in rice. *Rice*.
- Kurakazu T., Sobrizal, Ikeda K., Sanchez P. L., Doi K., Angeles E. R., Khush G. S., Yoshimura A. 2001.** *Oryza meridionalis* chromosomal segment introgression lines in cultivated rice, *O. sativa* L. *Rice Genetics Newsletter* **18**: 81-82.
- Kuroda Y., Sato Y.-I., Bounphanousay C., Kono Y., Tanaka K. 2005.** Gene flow from cultivated rice (*Oryza sativa* L.) to wild *Oryza* species (*O. rufipogon* Griff. and *O. nivara* Sharma and Shastry) on the Vientiane plain of Laos. *Euphytica* **142**(1): 75-83.
- Kutter C., Schob H., Stadler M., Meins F., Jr., Si-Ammour A. 2007.** MicroRNA-mediated regulation of stomatal development in *Arabidopsis*. *Plant Cell* **19**(8): 2417-2429.
- Kwon S. J., Lee J. K., Hong S. W., Park Y. J., McNally K. L., Kim N. S. 2006.** Genetic diversity and phylogenetic relationship in AA *Oryza* species as revealed by Rim2/Hipa CACTA transposon display. *Genes & Genetic Systems* **81**(2): 93-101.
- Larmande P., Gay C., Lorieux M., Perin C., Bouniol M., Droc G., Sallaud C., Perez P., Barnola I., Biderre-Petit C., Martin J., et al. 2008.** *Oryza* Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. *Nucleic Acids Research* **36**(suppl_1): D1022-1027.
- Lee H., Hur C. G., Oh C. J., Kim H. B., Pakr S. Y., An C. S. 2004.** Analysis of the root nodule-enhanced transcriptome in soybean. *Molecules and Cells* **18**(1): 53-62.
- Lee H.-Y., Chou J.-Y., Cheong L., Chang N.-H., Yang S.-Y., Leu J.-Y. 2008.** Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**(6): 1065-1073.
- Lee S. J., Oh C. S., Suh J. P., McCouch S. R., Ahn S. N. 2005.** Identification of QTLs for domestication-related and agronomic traits in an *Oryza sativa* x *O. rufipogon* BC1F7 population. *Plant Breeding* **124**(3): 209-219.
- Li D., Chen L., Jiang L., Zhu S., Zhao Z., Liu S., Su N., Zhai H., Ikehashi H., Wan J. 2007a.** Fine mapping of *S32(t)*, a new gene causing hybrid embryo sac sterility in a Chinese landrace rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **114**(3): 515-524.
- Li G., Hu W., Qin R., Jin H., Tan G., Zhu L., He G. 2008a.** Simple sequence repeat analyses of interspecific hybrids and MAALs of *Oryza officinalis* and *Oryza sativa*. *Genetica* **134**(2): 169-180.
- Li J., Xiao J., Grandillo S., Jiang L., Wan Y., Deng Q., Yuan L., McCouch S. R. 2004.** QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* **47**(4): 697-704.
- Li J., Xu P., Deng X., Zhou J., Hu F., Wan J., Tao D. 2008b.** Identification of four genes for stable hybrid sterility and an epistatic QTL from a cross between *Oryza sativa* and *Oryza glaberrima*. *Euphytica* **164**(3): 699-708.
- Li W., Jiang L., Zhou S., Wang C., Liu L., Chen L., Ikehashi H., Wan J. 2007b.** Fine mapping of *pss1*, a pollen semi-sterile gene in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* **114**(5): 939-946.
- Li W., Zeng R., Zhang Z., Ding X., Zhang G. 2006.** Fine mapping of locus *S-b* for F₁ pollen sterility in rice (*Oryza sativa* L.). *Chinese Science Bulletin* **51**(6): 675-680.

- Li W., Zeng R., Zhang Z., Ding X., Zhang G. 2008c. Identification and fine mapping of *S-d*, a new locus conferring the partial pollen sterility of intersubspecific F₁ hybrids in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 116(7): 915-922.
- Lin S. Y., Ikehashi H. 1993. A gamete abortion locus detected by segregation distortion of isozyme locus *Est-9* in wide crosses of rice (*Oryza sativa* L.). *Euphytica* 67(1): 35-40.
- Liu H. Y., Xu C. G., Zhang Q. 2004a. Male and female gamete abortions, and reduced affinity between the uniting gametes as the causes for sterility in an *indica/japonica* hybrid in rice. *Sexual Plant Reproduction* 17(2): 55-62.
- Liu K. D., Wang J., Li H. B., Xu C. G., Liu A. M., Li X. H., Zhang Q. 1997. A genome-wide analysis of wide compatibility in rice and the precise location of the *S5* locus in the molecular map. *Theoretical and Applied Genetics* 95(5): 809-814.
- Liu L., Lafitte R., Guan D. 2004b. Wild *Oryza* species as potential sources of drought-adaptive traits. *Euphytica* 138(2): 149-161.
- Londo J. P., Chiang Y.-C., Hung K.-H., Chiang T.-Y., Schaal B. A. 2006. From the Cover: Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proceedings of the National Academy of Sciences* 103(25): 9578-9583.
- Long M., Qing Z., Xian-Ping W., Han H., Li-Huang Z. 1994. RFLP analysis of progeny from hybrid of *Oryza alta* × *O. sativa*. *Acta Botanica Sinica* 36(1): 11-18.
- Long Y., Zhao L., Niu B., Su J., Wu H., Chen Y., Zhang Q., Guo J., Zhuang C., Mei M., Xia J., et al. 2008. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proceedings of the National Academy of Sciences* 105(48): 18871-18876.
- Lorieux M., Ndjioudjop M. N., Ghesquiere A. 2000. A first interspecific *Oryza sativa* × *Oryza glaberrima* microsatellite-based genetic linkage map. *Theoretical and Applied Genetics* 100(3-4): 593-601.
- Lorieux M., Reversat G., Garcia Diaz S. X., Denance C., Jouvenet N., Orioux Y., Bourger N., Pando-Bahuon A., Ghesquiere A. 2003. Linkage mapping of *Hsa-1(Og)*, a resistance gene of African rice to the cyst nematode, *Heterodera sacchari*. *Theoretical and Applied Genetics* 107(4): 691-696.
- Lu B. R., Naredo M. E. B., Juliano A. B., Jackson M. T. 2003. Genomic relationships of the AA genome *Oryza* species. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 118-121.
- Lu C. G., Takabatake K., Ikehashi H. 2000. Identification of segregation-distortion-neutral alleles to improve pollen fertility of *indica-japonica* hybrids in rice (*Oryza sativa* L.). *Euphytica* 113(2): 101-107.
- Lu F., Ammiraju J. S. S., Sanyal A., Zhang S., Song R., Chen J., Li G., Sui Y., Song X., Cheng Z., de Oliveira A. C., et al. 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proceedings of the National Academy of Sciences* 106(6): 2071-2076.
- Ma J., Bennetzen J. L. 2004. Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences* 101(34): 12404-12410.
- Maclean J. L., Dawe D. C., Hardy B., Hettel G. P. 2002. *Rice almanac: source book for the most important economic activity on earth*. Wallingford, UK: CABI Publishing, International Center for Tropical Agriculture (CIAT), West Africa Development Association (WARDA), Food and Agriculture Organization (FAO).
- Maekawa M. 1982. Studies on genetical differences between distantly related rice varieties. *Memoirs of the Faculty of Agriculture, Hokkaido University* 13(2): 146-177.

- Maekawa M., Kita F. 1985.** New gametophyte genes located in the third linkage group (Chromosome 3) of rice. *Japanese Journal of Breeding* **35**(1): 25-31.
- Mahmoud A. A., Sukumar S., Krishnan H. B. 2007.** Interspecific rice hybrid of *Oryza sativa* and *Oryza nivara* reveals a significant increase in seed protein content. *Journal of Agricultural and Food Chemistry* **56**(2): 476-482.
- Mamun E. A., Cantrill L. C., Overall R. L., Sutton B. G. 2005.** Cellular organisation in meiotic and early post-meiotic rice anthers. *Cell Biology International* **29**(11): 903-913.
- Martin N. H., Bouck A. C., Arnold M. L. 2007.** The genetic architecture of reproductive isolation in Louisiana Irises: Flowering phenology. *Genetics* **175**(4): 1803-1812.
- Martinez C. P., Arumuganathan K., Kikuchi H., Earle E. D. 1994.** Nuclear DNA content of ten rice species as determined by flow cytometry. *The Japanese Journal of Genetics* **69**(5): 513-523.
- Martinez C. P., Moncada P., López J., Almeida A., Gallego G., Borrero J., Duque M. C., Roca W., McCouch s. R., Bruzzone C., Tohme J. 2003.** Using new alleles from wild rice *Oryza rufipogon* to improve cultivated rice (*O. sativa*) in Latin America. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 138-143.
- Masly J. P., Jones C. D., Noor M. A. F., Locke J., Orr H. A. 2006.** Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* **313**(5792): 1448-1450.
- Matsubara K., Ito S., Nonoue Y., Ando T., Yano M. 2007.** A novel gene responsible for hybrid breakdown found in a cross between *japonica* and *indica* cultivars in rice. *Rice Genetics Newsletter* **23**: 11-13.
- Matsubara K., Khin T., Sano Y. 2003.** A gene block causing cross-incompatibility hidden in wild and cultivated rice. *Genetics* **165**(1): 343-352.
- Matsushita S., Iseki T., Fukuta Y., Araki E., Kobayashi S., Osaki M., Yamagishi M. 2003.** Characterization of segregation distortion on chromosome 3 induced in wide hybridization between *indica* and *japonica* type rice varieties. *Euphytica* **134**(1): 27-32.
- McClure B. 2009.** Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for S-RNase-based SI. *Journal of Experimental Botany* **60**(4): 1069-1081.
- Mendoza R. D., Molinawe J. A., Gregorio G. B., Guerta C. Q., Brar D. S. 2003.** Genetic variability of tolerance for iron toxicity in different species of *Oryza* and their derivatives. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 154-157.
- Merrill C., Bayraktaroglu L., Kusano A., Ganetzky B. 1999.** Truncated *RanGAP* encoded by the segregation distorter locus of *Drosophila*. *Science* **283**(5408): 1742-1745.
- Mihola O., Trachtulec Z., Vlcek C., Schimenti J. C., Forejt J. 2009.** A mouse speciation gene encodes a meiotic Histone H3 Methyltransferase. *Science* **323**(5912): 373 - 375.
- Milne R. I., Terzioglu S., Abbott R. J. 2003.** A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. *Molecular Ecology* **12**(10): 2719-2729.
- Mino M., Maekawa K., Ogawa K. i., Yamagishi H., Inoue M. 2002.** Cell Death Processes during Expression of Hybrid Lethality in Interspecific F1 Hybrid between *Nicotiana glauca* Domin and *Nicotiana tabacum*. *Plant Physiology* **130**(4): 1776-1787.
- Miura K., Yamamoto E., Morinaka Y., Takashi T., Kitano H., Matsuoka M., Ashikari M. 2008.** The hybrid breakdown 1(t) locus induces interspecific hybrid breakdown between rice *Oryza sativa* cv. Koshihikari and its wild relative *O. nivara*. *Breeding Science* **58**(2): 99-105.

- Miyazaki Y., Doi K., Yasui H., Yoshimura A. 2007. Identification of a new allele of F1 pollen sterility gene, S21, detected from the hybrid between *Oryza sativa* and *O. rufipogon*. *Rice Genetics Newsletter* 23: 36-38.
- Moncada P., Martinez C. P., Borrero J., Chatel M., Gauch H., Guimaraes E., Tohme J., McCouch S. R. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa* x *Oryza rufipogon* BC2F2 population evaluated in an upland environment. *Theoretical and Applied Genetics* 102(1): 41-52.
- Mori K., Kinoshita T., Takahashi M. 1973. Segregation distortion and its causation of an endosperm character in crosses of distantly related rice varieties.-Genetical studies on rice plant, LVIII-. *Memoirs of the Faculty of Agriculture, Hokkaido University* 9(1): 74-86.
- Morishima K., Hinata K., Oka H. 1962. Comparison between two Cultivated Rice Species, *Oryza sativa* L. and *O. glaberrima* STEUD. *Japanese Journal of Breeding* 12(3): 153-165
- Moyle L. C., Graham E. B. 2006. Genome-wide associations between hybrid sterility QTL and marker transmission ratio distortion. *Molecular Biology and Evolution* 23(5): 973-980.
- Moyle L. C., Nakazato T. 2008. Comparative genetics of hybrid incompatibility: Sterility in two *Solanum* species crosses. *Genetics* 179(3): 1437-1453.
- Moyle L. C., Nakazato T. 2009. Complex epistasis for Dobzhansky-Muller hybrid incompatibility in *Solanum*. *Genetics* 181(1): 347-351.
- Muller H. J. 1942. Isolating mechanisms, evolution and temperature. *Biological Symposium* 6: 71-125.
- Multani D. S., Jena K. K., Brar D. S., Reyes B. G., Angeles E. R., Khush G. S. 1994. Development of monosomic alien addition lines and introgression of genes from *Oryza australiensis* Domin. to cultivated rice *O. sativa* L. *Theoretical and Applied Genetics* 88(1): 102-109.
- Multani D. S., Khush G. S., delos Reyes B. G., Brar D. S. 2003. Alien genes introgression and development of monosomic alien addition lines from *Oryza latifolia* Desv. to rice, *Oryza sativa* L. *Theoretical and Applied Genetics* 107(3): 395-405.
- Murray S. S. 2004. Searching for the origins of African rice domestication. *Antiquity* 78 (300): <http://antiquity.ac.uk/projgall/murray/index.html>.
- Nagai Y. S., Doi K., Yoshimura A. 2002. A seed shattering gene, *Sh3*, and its inhibitor gene found in an African cultivated rice, *Oryza glaberrima* Steud. *Rice Genetics Newsletter* 19: 76-77.
- Nakagahara M. 1972. Genetic mechanism on the distorted segregation of marker genes belonging to the eleventh linkage group in cultivated rice. *Japanese Journal of Breeding* 22(4): 232-238.
- Naredo M. E. B., Juliano A. B., Lu B.-R., Jackson M. T. 1997. Hybridization of AA genome rice species from Asia and Australia I. Crosses and development of hybrids. *Genetic Resources and Crop Evolution* 44(1): 17-23.
- Naredo M. E. B., Juliano A. B., Lu B.-R., Jackson M. T. 1998. Taxonomic status of *Oryza glumaepatula* Steud. II. Hybridization between New World diploids and AA genome species from Asia and Australia. *Genetic Resources and Crop Evolution* 45(3): 205-214.
- Navarro A., Barton N. H. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57(3): 447-459.
- Ndjiondjop M. N., Albar L., Fargette D., Fauquet C., Ghesquiere A. 1999. The genetic basis of high resistance to Rice Yellow Mottle Virus (RYMV) in cultivars of two cultivated rice species. *Plant Disease* 83(10): 931-935.

- Ndjiondjop M. N., Semagn K., Sie M., Cissoko M., Fatondji B., Jones M. 2008.** Molecular profiling of interspecific lowland rice populations derived from IR64 (*Oryza sativa*) and Tog5681 (*Oryza glaberrima*). *African Journal of Biotechnology* 7(23): 4219–4229.
- Nguyen B., Brar D., Bui B., Nguyen T., Pham L., Nguyen H. 2003.** Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into *indica* rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 106(4): 583-593.
- Nguyen N. V., Ferrero A. 2006.** Meeting the challenges of global rice production. *Paddy and Water Environment* 4(1): 1-9.
- Nobuta K., Venu R. C., Lu C., Belo A., Vemaraju K., Kulkarni K., Wang W., Pillay M., Green P. J., Wang G.-l., Meyers B. C. 2007.** An expression atlas of rice mRNAs and small RNAs. *Nature Biotechnology* 25(4): 473-477.
- Noland H. M., John H. W. 2007.** Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution* 61(1): 68-82.
- Nuijten E., van Treuren R., Struik P. C., Mokuwa A., Okry F., Teeken B. I., Richards P. 2009.** Evidence for the emergence of new rice types of interspecific hybrid origin in west African farmers' fields. *PLoS ONE* 4(10): e7335.
- Nwanze K. F., Mohapatra S., Kormawa P., Keya S., Bruce-Oliver S. 2006.** Rice development in sub-Saharan Africa. *Journal of the Science of Food and Agriculture* 86(5): 675-677.
- Oka H. I. 1953.** The mechanism of sterility in the intervarietal hybrid: Phylogenetic differentiation of the cultivated rice plant. VI. *Japanese Journal of Breeding* 2(4): 217-224.
- Oka H. I. 1957a.** Genic analysis for the sterility of hybrids between distantly related varieties of cultivated rice. *Journal of Genetics* 55(3): 397-409.
- Oka H. I. 1957b.** Phylogenetic differentiation of cultivated rice. XV Complementary lethal genes in rice. *Japanese Journal of Genetics* 32(3): 83-87.
- Orjuela J., Garavito A., Bouniol M., Arbelaez J., Moreno L., Kimball J., Wilson G., Rami J.-F., Tohme J., McCouch S., Lorieux M. 2009.** A universal core genetic map for rice. *Theoretical and Applied Genetics* Doi: 10.1007/s00122-009-1176-1.
- Orr H. A., Irving S. 2005.** Segregation distortion in hybrids between the Bogota and USA subspecies of *Drosophila pseudoobscura*. *Genetics* 169(2): 671-682.
- Orr H. A., Turelli M. 2001.** The evolution of postzygotic isolation: Accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55(6): 1085-1094.
- Ouyang S., Buell C. R. 2004.** The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* 32(suppl_1): D360-363.
- Ouyang S., Zhu W., Hamilton J., Lin H., Campbell M., Childs K., Thibaud-Nissen F., Malek R. L., Lee Y., Zheng L., Orvis J., et al. 2007.** The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* 35(suppl_1): D883-887.
- Park K. C., Lee J. K., Kim N. H., Shin Y. B., Lee J. H., Kim N. S. 2003.** Genetic variation in *Oryza* species detected by MITE-AFLP. *Genes & Genetic Systems* 78(3): 235-243.
- Pathak M. D., Khan Z. R. 1994.** *Insect pests of rice*. Manila, Philippines: International Rice Research Institute, International Centre of Insect Physiology and Ecology.
- Pélissier P. 1966.** *Les paysans du Sénégal. Les civilisations agraires du Cayor à la Casamance*. Saint-Yrieix (Haute-Vienne): Imprimerie Fabrègue.

- Pesin J. A., Orr-Weaver T. L. 2008.** Regulation of APC/C activators in mitosis and meiosis. *Annual Review of Cell and Developmental Biology* **24**(1): 475-499.
- Phadnis N., Orr H. A. 2009.** A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science* **323**(5912): 376-379.
- Pham J. L. 1992.** *Evaluation des ressources génétiques des riz cultivés en Afrique par hybridation intra- et interspécifique*. Paris XI, Orsay Paris.
- Phillips G. C., Collins G. B., Taylor N. L. 1982.** Interspecific hybridization of red clover (*Trifolium pratense* L.) with *T. sarosiense* Hazsl. Using in vitro embryo rescue. *Theoretical and Applied Genetics* **62**(1): 17-24.
- Portères R. 1962.** Berceaux agricoles primaires sur le continent Africain. *The Journal of African History* **3**(2): 195-210.
- Presgraves D. C., Balagopalan L., Abmayr S. M., Orr H. A. 2003.** Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423**(6941): 715-719.
- Prot J.-C., Rahman M. L. 1994.** Nematode ecology, economic importance, and management in rice ecosystems in South and Southeast Asia. En: Teng P. S., Heong K. L., Moody K. eds. *Rice pest science and management. Selected papers from the International Rice Research Conference*. Manila, Philippines: International Rice Research Institute.
- Qiu S., Liu K., Jiang J., Song X., Xu C., Li X., Zhang Q. 2005.** Delimitation of the rice wide compatibility gene *S5n* to a 40-kb DNA fragment. *Theoretical and Applied Genetics* **111**(6): 1080-1086.
- Ramsey J., Bradshaw H. D., Schemske D. W., Morgan M. 2003.** Components of Reproductive Isolation between the Monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution* **57**(7): 1520-1534.
- Rangel P. N., Brondani R. P., Rangel P. H., Brondani C. 2008.** Agronomic and molecular characterization of introgression lines from the interspecific cross *Oryza sativa* (BG90-2) x *Oryza glumaepatula* (RS-16). *Genetics and Molecular Research* **7**(1): 184-195.
- Razdan V. K., Sabitha M. 2009.** Integrated disease management: Concepts and practices. *Integrated Pest Management: Innovation-Development Process*, 369-389.
- Ren F. G., Lu B. R., Li S. Q., Huang J. Y., Zhu Y. G. 2003.** A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. *Theoretical and Applied Genetics* **108**(1): 113-120.
- Ren G., Xu P., Deng X., Zhou J., Hu F., LI J., LI F., Zhang Z., Tao D. 2005.** A new gamete eliminator from *Oryza glaberrima*. *Rice Genetics Newsletter* **22**: 45-48.
- Reversat G., Destombes D. 1998.** Screening for resistance to *Heterodera sacchari* in the two cultivated rice species, *Oryza sativa* and *O. glaberrima*. *Fundamental and Applied Nematology* **21**(4): 307-317.
- Rha E. S., Thseng F. S., Wu S. T. 1994.** Distorted segregation of *Acp-1* for acid phosphatase in a wide cross of rice. . *Rice Genetics Newsletter* **11**: 106-108.
- Rice P., Longden I., Bleasby A. 2000.** EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**(6): 276-277.
- Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M.-A., Barrell B. 2000.** Artemis: sequence visualization and annotation. *Bioinformatics* **16**(10): 944-945.
- Sahrawat K. L., Sika M. 2002.** Comparative tolerance of *Oryza sativa* and *O. glaberrima* rice cultivars for iron toxicity in West Africa. *International Rice Research Notes* **27**(2): 30-31.

- Sakagami J.-I., Joho Y., Ito O. 2009.** Contrasting physiological responses by cultivars of *Oryza sativa* and *O. glaberrima* to prolonged submergence. *Annals of Botany* **103**(2): 171-180.
- Salamov A. A., Solovyev V. V. 2000.** *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Research* **10**(4): 516-522.
- Sanchez P. L., Sobrizal, Ikeda K., Yasui H., Yoshimura A. 2003.** Identifying late heading genes in rice using *Oryza glumaepatula* introgression lines. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 153-154.
- SanMiguel P., Gaut B. S., Tikhonov A., Nakajima Y., Bennetzen J. L. 1998.** The paleontology of intergene retrotransposons of maize. *Nature Genetics* **20**(1): 43-45.
- Sano Y. 1983.** Genetic-studies of speciation in cultivated rice .3. A new gene controlling sterility in F₁ hybrids of 2 cultivated rice species - Its association with photoperiod sensitivity. *Journal of Heredity* **74**(6): 435-439.
- Sano Y. 1986.** Sterility barriers between *Oryza sativa* and *O. glaberrima*. En: Institute I. R. R. ed. *Rice Genetics*. Manila, 109-118.
- Sano Y. 1990.** The genic nature of gamete eliminator in rice. *Genetics* **125**(1): 183-191.
- Sano Y. 1994.** Pollen-killer in rice. *Japanese Journal of Breeding* **44**((suppl 1)): 298.
- Sano Y., Chu Y.-E., Oka H.-I. 1979.** Genetic studies of speciation in cultivated rice. 1. Genetic analysis for the F₁ sterility between *O. sativa* L. and *O. glaberrima* STEUD. *Japanese Journal of Genetics* **54**(2): 121-132.
- Sano Y., Sano R., Eiguchi M., Hirano H. Y. 1994.** Gamete eliminator adjacent to the Wx locus as revealed by pollen analysis in Rice. *Journal of Heredity* **85**(4): 310-312.
- Sawamura N., Sano Y. 1996.** Chromosomal location of gamete eliminator, *S11(t)*, found in an Indica-Japonica hybrid. *Rice Genetics Newsletter* **13**: 70-72.
- Schimenti J. 2000.** Segregation distortion of mouse t haplotypes: the molecular basis emerges. *Trends in Genetics* **16**(6): 240-243.
- Second G. 1986.** Isozymes and phylogenetic relationship in *Oryza*. En: Institute I. R. R. ed. *Rice Genetics*. Manila, 27-39.
- Semagn K., Ndjioudjop M. N., Cissoko M. 2006.** Microsatellites and agronomic traits for assessing genetic relationships among 18 New Rice for Africa (NERICA) varieties. *African Journal of Biotechnology* **5**(10): 800-810.
- Semagn K., Ndjioudjop M. N., Lorieux M., Cissoko M., Jones M., McCouch S. 2007.** Molecular profiling of an interspecific rice population derived from a cross between WAB 56-104 (*Oryza sativa*) and CG 14 (*Oryza glaberrima*). *African Journal of Biotechnology* **6**(17): 2014-2022.
- Semon M., Nielsen R., Jones M. P., McCouch S. R. 2005.** The population structure of African cultivated rice *Oryza glaberrima* (Steud.): Evidence for elevated levels of linkage disequilibrium caused by admixture with *O. sativa* and ecological adaptation. *Genetics* **169**(3): 1639-1647.
- Septiningsih E. M., Prasetyono J., Lubis E., Tai T. H., Tjubaryat T., Moeljopawiro S., McCouch S. R. 2003.** Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theoretical and Applied Genetics* **107**(8): 1419-1432.
- Shan J.-X., Zhu M.-Z., Shi M., Gao J.-P., Lin H.-X. 2009.** Fine mapping and candidate gene analysis of *spd6*, responsible for small panicle and dwarfness in wild rice (*Oryza rufipogon* Griff.). *Theoretical and Applied Genetics* **119**(5): 827-836.
- Sie M., Sere Y., Sanyang S., Narteh L. T., Dogbe S., Coulibaly M. M., Sido A., Cisse F., Drammeh E., Ogunbayo S. A., Zadaj L., et al. 2008.** Regional Yield Evaluation of

- the Interspecific Hybrids (*O. glaberrima* x *O. sativa*) and Intraspecific (*O. sativa* x *O. sativa*) Lowland Rice. *Asian Journal of Plant Sciences* 7(2): 130-139.
- Sobrizal, Ikeda K., Sanchez P. L., Doi K., Angeles E. R., Khush G. S., Yoshimura A. 1999.** Development of *Oryza glumaepatula* introgression lines in rice *O. sativa* L. *Rice Genetics Newsletter* 16: 107-108.
- Sobrizal, Matsuzaki Y., Ikeda K., Sanchez P. L., Doi K., Yasui H., Angeles E. R., Khush G. S., Yoshimura A. 2003.** *Oryza glumaepatula* Steud. introgression lines in rice: identification of genes for reproductive barriers. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 128-130.
- Sobrizal, Matsuzaki Y., Sanchez L., Ikeda L., Yoshimura A. 2000a.** Identification of a gene for male gamete abortion in backcross progeny of *Oryza sativa* and *Oryza glumaepatula*. *Rice Genetics Newsletter* 17: 59-60.
- Sobrizal, Matsuzaki Y., Sanchez L., Ikeda L., Yoshimura A. 2000b.** Mapping of F1 pollen semi-sterility gene found in backcross progeny of *Oryza sativa* L. and *Oryza glumaepatula* Steud. *Rice Genetics Newsletter* 17: 61-63.
- Sobrizal, Matsuzaki Y., Yoshimura A. 2001a.** Mapping of a gene for pollen semi-sterility on chromosome 8 of rice. *Rice Genetics Newsletter* 18: 59-61.
- Sobrizal, Matsuzaki Y., Yoshimura A. 2002.** Mapping of pollen semi-sterility gene, *S28(t)*, on rice chromosome 4. *Rice Genetics Newsletter* 19: 80-82.
- Sobrizal, Tanaka C., Yasui H., Yoshimura A. 2001b.** Mapping of a gene for F2 weakness on chromosome 4 of rice. *Rice Genetics Newsletter* 18: 61-63.
- Song X., Qiu S. Q., Xu C. G., Li X. H., Zhang Q. 2005.** Genetic dissection of embryo sac fertility, pollen fertility, and their contributions to spikelet fertility of intersubspecific hybrids in rice. *Theoretical and Applied Genetics* 110(2): 205-211.
- Song Z. P., Lu B.-R., Zhu Y. G., Chen J. K. 2003.** Gene Flow from Cultivated Rice to the Wild Species *Oryza rufipogon* under Experimental Field Conditions. *New Phytologist* 157(3): 657-665.
- Sonnhammer E. L. L., Durbin R. 1995.** A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167(1-2): GC1-GC10.
- Soriano I. R., Schmit V., Brar D. S., Prot J.-C., Reversat G. 1999.** Resistance to rice root-knot nematode *Meloidogyne graminicola* identified in *Oryza longistaminata* and *O. glaberrima*. *Nematology* 1(4): 395-398.
- Sun S., Ting C.-T., Wu C.-I. 2004.** The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* 305(5680): 81-83.
- Sweigart A. L., Fishman L., Willis J. H. 2006.** A simple genetic incompatibility causes hybrid male sterility in *mimulus*. *Genetics* 172(4): 2465-2479.
- Sweigart A. L., Mason A. R., Willis J. H. 2007.** Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution* 61(1): 141-151.
- Tagawa T., Hirao K., Kubota F. 2000.** A specific feature of nitrogen utilization efficiency in leaf photosynthesis in *Oryza glaberrima* Steud. *Japanese Journal of Crop Science* 69(1): 74-79.
- Taguchi K., Doi K., Yoshimura A. 1999.** RFLP mapping of *S19*, a gene for F₁ pollen semi-sterility found in backcross progeny of *Oryza sativa* and *O. glaberrima*. *Rice Genetics Newsletter* 16: 70-71.
- Tan L. B., Liu F. X., Xue W., Wang G. J., Ye S., Zhu Z. F., Fu Y. C., Wang X. K., Sun C. Q. 2007.** Development of *Oryza rufipogon* and *O. sativa* introgression lines and assessment for yield-related quantitative trait loci. *Journal of Integrative Plant Biology* 49(6): 871-884.

- Taneichi T., KOIDE Y., Nishimoto D., Sano Y. 2005.** Hybrid sterility gene *S13* found in a distantly related rice species, *O. longistaminata*. *Genes & Genetic Systems* **80**(6): 477.
- Tang S., Presgraves D. C. 2009.** Evolution of the *Drosophila* Nuclear Pore Complex Results in Multiple Hybrid Incompatibilities. *Science* **323**(5915): 779-782.
- Tang T., Lu J., Huang J., He J., McCouch S. R., Shen Y., Kai Z., Purugganan M. D., Shi S., Wu C.-I. 2006.** Genomic variation in rice: Genesis of highly polymorphic linkage blocks during domestication. *PLoS Genetics* **2**(11): e199.
- Tao D., Hu F., Yang Y., Xu P., Li J., Sacks E., McNally K. L., Sripichitt P. 2003.** A rhizomatous individual obtained from interspecific BC1F1 progenies between *Oryza sativa* and *O. longistaminata*. En: Khush G. S., Brar D. S., Hardy B. eds. *Advances in rice genetics*. Los Baños (Philippines): International Rice Research Institute, 151-152.
- Tao Y., Araripe L., Kingan S. B., Ke Y., Xiao H., Hartl D. L. 2007a.** A sex-ratio meiotic drive system in *Drosophila simulans*. II: An X-linked distorter. *PLoS Biology* **5**(11): e293.
- Tao Y., Masly J. P., Araripe L., Ke Y., Hartl D. L. 2007b.** A sex-ratio meiotic drive system in *Drosophila simulans*. I: An autosomal suppressor. *PLoS Biology* **5**(11): e292.
- Thibaud-Nissen F., Ouyang S., Buell C. 2009.** Identification and characterization of pseudogenes in the rice gene complement. *BMC Genomics* **10**(1): 317.
- Thiel T., Michalek W., Varshney R., Graner A. 2003.** Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* **106**(3): 411-422.
- Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G. 1997.** The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**(24): 4876-4882.
- Thomson M. J., Tai T. H., McClung A. M., Lai X. H., Hinga M. E., Lobos K. B., Xu Y., Martinez C. P., McCouch S. R. 2003.** Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theoretical and Applied Genetics* **107**(3): 479-493.
- Tian F., Li D. J., Fu Q., Zhu Z. F., Fu Y. C., Wang X. K., Sun C. Q. 2006a.** Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. *Theoretical and Applied Genetics* **112**(3): 570-580.
- Tian F., Zhu Z. F., Zhang B. S., Tan L. B., Fu Y. C., Wang X. K., Sun C. Q. 2006b.** Fine mapping of a quantitative trait locus for grain number per panicle from wild rice (*Oryza rufipogon* Griff.). *Theoretical and Applied Genetics* **113**(4): 619-629.
- Ting C.-T., Tsaui S.-C., Wu M.-L., Wu C.-I. 1998.** A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**(5393): 1501-1504.
- Torjek O., Witucka-Wall H., Meyer R. C., von Korff M., Kusterer B., Rautengarten C., Altmann T. 2006.** Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theoretical and Applied Genetics* **113**(8): 1551-1561.
- Tran D. V., Nguyen N. V. 2006.** The concept and implementation of precision farming and rice integrated crop management systems for sustainable production in the twenty-first century. *International Rice Commission Newsletter (FAO)* **55**: 91-102.
- Tsuno A., Miyoshi K., Tsujii R., Miyakawa T., Mizuta K. 2000.** *RRS1*, a conserved essential gene, encodes a novel regulatory protein required for ribosome biogenesis in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **20**(6): 2066-2074.

- Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R., Leunissen J. A. M. 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* 35(suppl_2): W71-74.
- Vaughan D. A., Ge S., Kaga A., Tomooka N. 2008. Phylogeny and biogeography of the genus *Oryza*. *Rice Biology in the Genomics Era*, 219-234.
- Vietmeyer N. 1996. African Rice. *Lost Crops of Africa. Vol 1: Grains*. Washington, DC: Washington National Academy Press, 17-38.
- Vinatzer B. A., Patocchi A., Gianfranceschi L., Tartarini S., Zhang H.-B., Gessler C., Sansavini S. 2001. Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of Tomato with a cluster of genes cosegregating with Vf Apple scab resistance. *Molecular Plant-Microbe Interactions* 14(4): 508-515.
- Wan J., Ikehashi H. 1996. List of hybrid sterility gene loci (HSGLi) in cultivated rice (*Oryza sativa* L.). *Rice Genetics Newsletter* 13: 110-114.
- Wan J., Ikehasih H., Sakai M., Horisue H., Imbe T. 1998. Mapping of hybrid sterility gene *S17* of rice (*Oryza sativa* L.) by isozyme and RFLP markers. *Rice Genetics Newsletter* 15: 151-154.
- Wan J., Yamaguchi Y., Kato H., Ikehashi H. 1996. Two new loci for hybrid sterility in cultivated rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 92(2): 183-190.
- Wang C. M., Zhu C. S., Zhai H. Q., Wan J. M. 2005. Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genetical Research* 86(2): 97-106.
- Wang G. W., He Y. Q., Xu C. G., Zhang Q. F. 2006. Fine mapping of *f5-Du*, a gene conferring wide-compatibility for pollen fertility in inter-subspecific hybrids of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 112(2): 382-387.
- Wang J., Liu K. D., Xu C. G., Li X. H., Zhang Q. 1998. The high level of wide-compatibility of variety 'Dular' has a complex genetic basis. *Theoretical and Applied Genetics* 97(3): 407-412.
- Wang Y., Yang M. 2006. The ARABIDOPSIS SKP1-LIKE1 (ASK1) protein acts predominately from leptotene to pachytene and represses homologous recombination in male meiosis. *Planta* 223(3): 613-617.
- Watanabe H., Futakuchi K., Jones M. P., Sobambo B. A. 2006. Grain Protein Content of Interspecific Progenies Derived from the Cross of African Rice (*Oryza glaberrima* Steud.) and Asian Rice (*Oryza sativa* L.). *Plant Production Science* 9(3): 287-293.
- Wittbrodt J., Adam D., Malitschek B., Maueler W., Raulf F., Telling A., Robertson S. M., Schartl M. 1989. Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. *Nature* 341(6241): 415-421.
- Xiao J., Li J., Grandillo S., Ahn S. N., Yuan L., Tanksley S. D., McCouch S. R. 1998. Identification of Trait-improving Quantitative Trait Loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* 150(2): 899-909.
- Xie X., Jin F., Song M.-H., Suh J.-P., Hwang H.-G., Kim Y.-G., McCouch S., Ahn S.-N. 2008. Fine mapping of a yield-enhancing QTL cluster associated with transgressive variation in an *Oryza sativa* × *O. rufipogon* cross. *Theoretical and Applied Genetics* 116(5): 613-622.
- Xu G., Ma H., Nei M., Kong H. 2009. Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. *Proceedings of the National Academy of Sciences* 106(3): 835-840.
- Xu J.-H., Cheng C., Tsuchimoto S., Ohtsubo H., Ohtsubo E. 2007. Phylogenetic analysis of *Oryza rufipogon* strains and their relations to *Oryza sativa* strains by insertion polymorphism of rice SINEs. *Genes & Genetic Systems* 82(3): 217-229.

- Xu Y., Zhu L., Xiao J., Huang N., McCouch S. R. 1997.** Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). *Molecular and General Genetics MGG* **253**(5): 535-545.
- Xue-Lin F., Yong-Gen L., Xiang-Dong L., Jin-Quan L. 2009.** Crossability barriers in the interspecific hybridization between *Oryza sativa* and *O. meyeriana*. *Journal of Integrative Plant Biology* **51**(1): 21-28.
- Yamamoto E., Takashi T., Morinaka Y., Lin S., Kitano H., Matsuoka M., Ashikari M. 2007.** Interaction of two recessive genes, *hbd2* and *hbd3*, induces hybrid breakdown in rice. *Theoretical and Applied Genetics* **115**(2): 187-194.
- Yan H., Liu G., Cheng Z., Min S., Zhu L. 2001.** Characterization of euploid backcross progenies derived from interspecific hybrids between *Oryza sativa* and *O. eichingeri* by restriction fragment length polymorphism (RFLP) analysis and genomic in situ hybridization (GISH). *Genome* **44**(1): 86-95.
- Yan H. H., Liu G. Q., Cheng Z. K., Li X. B., Liu G. Z., Min S. K., Zhu L. H. 2002.** A genome-specific repetitive DNA sequence from *Oryza eichingeri* : characterization, localization, and introgression to *O. sativa*. *Theoretical and Applied Genetics* **104**(2): 177-183.
- Yanagihara S., Kato H., Ikehashi H. 1992.** A new locus for multiple alleles causing hybrid sterility between an Aus variety and Javanica varieties in Rice (*Oryza sativa* L.). *Japanese Journal of Breeding* **42**(4): 793-801.
- Yasui H., Iwata N. 1991.** Production of monosomic alien addition lines of *O. sativa* having a single *O. punctata* chromosome. En: IRRI ed. *Rice Genetics II*. Manila, Philippines IRRI, 147-155.
- Yi C., Tang S., Zhou Y., Liang G., Gong Z., Gu M. 2008.** Development and characterization of interspecific hybrids between *Oryza sativa* and *O. latifolia* by in situ hybridization. *Chinese Science Bulletin* **53**(19): 2973-2980.
- Yoon D. B., Kang K. H., Kim H. J., Ju H. G., Kwon S. J., Suh J. P., Jeong O. Y., Ahn S. N. 2006.** Mapping quantitative trait loci for yield components and morphological traits in an advanced backcross population between *Oryza grandiglumis* and the *O. sativa* japonica cultivar Hwaseongbyeon. *Theoretical and Applied Genetics* **112**(6): 1052-1062.
- Zeigler R., Barclay A. 2008.** The Relevance of Rice. *Rice* **1**(1): 3-10.
- Zfang Q., Lin S. C., Zhao B. Y., Wang C. L., Yang W. C., Zhou Y. L., LI D. Y., Chen C. B., Zhu L. H. 1998.** Identification and Tagging a new gene for resistance to bacterial blight (*Xanthomonas oryzae* pv. *oryzae*) from *O. rufipogon*. *Rice Genetics Newsletter* **15**: 138-142.
- Zhang G., Lu Y. 1996.** Genetics of F₁ pollen sterility in *Oryza sativa*. En: Khush G. S. ed. *Rice Genetics III*. Manila, Philippines: IRRI, 418-422.
- Zhang X., Zhou S. X., Fu Y. C., Su Z., Wang X. K., Sun C. Q. 2006.** Identification of a drought tolerant introgression line derived from dongxiang common wild rice (*O. rufipogon* Griff.). *Plant Molecular Biology* **62**(1-2): 247-259.
- Zhang Z., Xu P., Hu F., Zhou J., Li J., Deng X., Ren G., Li F., Tao D. 2005.** A new sterile gene from *Oryza glaberrima* on chromosome 3. *Rice Genetics Newsletter* **22**: 26-29.
- Zhao X., Yang L., Zheng Y., Xu Z., Wu W. 2009.** Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza rufipogon* L.) in relation to the domestication of cultivated rice (*O. sativa* L.). *Journal of Genetics and Genomics* **36**(7): 435-442.

- Zhao Z., Jiang L., Zhang W., Yu C., Zhu S., Xie K., Tian H., Liu L., Ikehashi H., Wan J. 2007.** Fine mapping of *S31*, a gene responsible for hybrid embryo-sac abortion in rice (*Oryza sativa* L.). *Planta* **226**(5): 1087-1096.
- Zhao Z., Wang C., Jiang L., Zhu S., Ikehashi H., Wan J. 2006.** Identification of a new hybrid sterility gene in rice (*Oryza sativa* L.). *Euphytica* **151**(3): 331-337.
- Zheng N., Schulman B. A., Song L., Miller J. J., Jeffrey P. D., Wang P., Chu C., Koepp D. M., Elledge S. J., Pagano M., Conaway R. C., et al. 2002.** Structure of the Cull1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**(6882): 703-709.
- Zhiping S., Baorong L., Yingguo Z., Jiakuan C. 2002.** Pollen competition between cultivated and wild rice species (*Oryza sativa* and *O. rufipogon*). *New Phytologist* **153**(2): 289-296.
- Zhou Y.-L., Xu J.-L., Zhou S.-C., Yu J., Xie X.-W., Xu M.-R., Sun Y., Zhu L.-H., Fu B.-Y., Gao Y.-M., Li Z.-K. 2009.** Pyramiding *Xa23* and *Rxo1* for resistance to two bacterial diseases into an elite indica rice variety using molecular approaches. *Molecular Breeding* **23**(2): 279-287.
- Zhu Q., Zheng X., Luo J., Gaut B. S., Ge S. 2007.** Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. *Molecular Biology and Evolution* **24**(3): 875-888.
- Zhu Q. H., Ge S. 2005.** Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist* **167**(1): 249-265.
- Zhu S., Jiang L., Wang C., Zhai H., Li D., Wan J. 2005a.** The origin of weedy rice Ludao in China deduced by genome wide analysis of its hybrid sterility genes. *Breeding Science* **55**(4): 409-414.
- Zhu S., Wang C., Zheng T., Zhao Z., Ikehashi H., Wan J. 2005b.** A new gene located on chromosome 2 causing hybrid sterility in a remote cross of rice. *Plant Breeding* **124**(5): 440-445.