



**HAL**  
open science

## Long-Range and Targeted Ectopic Recombination between the Two Homeologous Chromosomes 11 and 12 in *Oryza* Species

J. Jacquemin, C. Chaparro, M. Laudie, Arjan Berger, Frédérick F. Gavory, J.  
Goicoechea, A. Wing, R. Cooke

► **To cite this version:**

J. Jacquemin, C. Chaparro, M. Laudie, Arjan Berger, Frédérick F. Gavory, et al.. Long-Range and Targeted Ectopic Recombination between the Two Homeologous Chromosomes 11 and 12 in *Oryza* Species. *Molecular Biology and Evolution*, 2011, 28 (11), pp.3139-3150. 10.1093/molbev/msr144 . hal-02114343

**HAL Id: hal-02114343**

**<https://hal.science/hal-02114343>**

Submitted on 25 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Long-Range and Targeted Ectopic Recombination between the Two Homeologous Chromosomes 11 and 12 in *Oryza* Species

J. Jacquemin,<sup>\*‡</sup><sup>1</sup> C. Chaparro,<sup>1</sup> M. Laudí,<sup>1</sup> A. Berger,<sup>2</sup> F. Gavory,<sup>2</sup> J. L. Goicoechea,<sup>3</sup> R. A. Wing,<sup>3</sup> and R. Cooke<sup>1</sup>

<sup>1</sup>Laboratoire Génome et Développement des Plantes, Unité Mixte de Recherche Centre National de la Recherche Scientifique/ Institut de Recherche pour le Développement/Université de Perpignan Via Domitia, Université de Perpignan, 66860 Perpignan-Cedex, France

<sup>2</sup>Genoscope-Centre National de Séquençage, Evry, France

<sup>3</sup>Arizona Genomics Institute, The University of Arizona

‡Present address: Arizona Genomics Institute, The University of Arizona

\*Corresponding author: E-mail: juliej@cals.arizona.edu.

Associate editor: Manolo Gouy

## Abstract

Whole genome duplication (WGD) and subsequent evolution of gene pairs have been shown to have shaped the present day genomes of most, if not all, plants and to have played an essential role in the evolution of many eukaryotic genomes. Analysis of the rice (*Oryza sativa* ssp. *japonica*) genome sequence suggested an ancestral WGD ~50–70 Ma common to all cereals and a segmental duplication between chromosomes 11 and 12 as recently as 5 Ma. More recent studies based on coding sequences have demonstrated that gene conversion is responsible for the high sequence conservation which suggested such a recent duplication. We previously showed that gene conversion has been a recurrent process throughout the *Oryza* genus and in closely related species and that orthologous duplicated regions are also highly conserved in other cereal genomes. We have extended these studies to compare megabase regions of genomic (coding and noncoding) sequences between two cultivated (*O. sativa*, *Oryza glaberrima*) and one wild (*Oryza brachyantha*) rice species using a novel approach of topological incongruity. The high levels of intraspecies conservation of both gene and nongene sequences, particularly in *O. brachyantha*, indicate long-range conversion events less than 4 Ma in all three species. These observations demonstrate megabase-scale conversion initiated within a highly rearranged region located at ~2.1 Mb from the chromosome termini and emphasize the importance of gene conversion in cereal genome evolution.

**Key words:** comparative genomics, duplication, gene conversion, *Oryza* genus, recombination hot spot.

## Introduction

The availability of genome sequences from closely related species, such as yeasts (reviewed in Dujon 2010) or *Drosophila* (Hahn et al. 2007), has led to considerable advances in our understanding of genome evolution. In plants, the *Oryza* Map Alignment Project (OMAP, Wing et al. 2005), articulated around the reference *Oryza sativa* ssp. *japonica* c.v. Nipponbare genome sequence (hereafter RefSeq) has developed resources aimed at characterizing rice genome evolution. In a genus containing two cultivated and 22 wild species, molecular resources have been created representing the ten genome types and which provide the means of studying short-term evolutionary dynamics in plants. This has allowed deep comparative analysis of these closely related species at specific loci (Lu et al. 2009; Sanyal et al. 2010).

The importance of duplications in the evolution of plant genomes has been emphasized by the analysis of several complete genome sequences (Van de Peer et al. 2009). Preliminary analysis of the rice RefSeq suggested a whole genome duplication (WGD), probably common to all grasses

and a more recent segmental duplication of ~2–3 Mb in the distal region of the short arms of chromosomes 11 and 12 (Yu et al. 2005; Rice Chromosomes 11 and 12 Sequencing Consortia 2005). More recent studies by ourselves (Jacquemin et al. 2009) and others (Paterson et al. 2009) demonstrated that this duplicated block is not specific to the *Oryza* genus, as its presumed age suggested, and this is confirmed by its presence in two other model cereal genomes, *Sorghum bicolor* and *Brachypodium distachyon*. As chromosomes 11 and 12 result from the WGD at the base of the Poaceae, this strongly suggests that this duplication has the same origin. Wang et al. (2007), comparing 278 gene pairs along the whole 11–12 block in the RefSeq and the *indica* subspecies sequence, proposed a stochastic evolution of gene pairs in this region, in which gene conversion acts as an occasional, sometimes frequent interruption to independent evolution of paralogs. Our study (Jacquemin et al. 2009) on a wider sampling of species within and closely related to the *Oryza* genus rather indicated recurrent concerted evolution affecting the same gene pairs in all species, at least in the immediate

subtelomeric region and suggested a breakpoint in colinearity at ~2 Mb from the telomeres.

Gene conversion is the nonreciprocal transfer of genetic information between homologous sequences, leading to homogenization during meiotic or mitotic recombination (Szostak et al. 1983). Four pathways to repair DNA double strand breaks (DSBs) through homologous recombination (HR) are generally grouped under the term of gene conversion (reviewed in Chen et al. 2007; Llorente et al. 2008; De Muyt et al. 2009; Duret and Galtier 2009): double-strand break repair (DSBR), double-Holliday junction (HJ) dissolution, synthesis-dependent strand annealing (SDSA), and break-induced replication (BIR). Ectopic gene conversion involves dispersed duplicated sequences rather than sister chromatids or homologous loci. As this process has mainly been described for multigene families and tandemly duplicated genes (Gao and Innan 2004; Xu et al. 2008; Hogan and Bettencourt 2009; Yang et al. 2009; Ezawa et al. 2010), the long-term conservation of large genomic regions in rice and other cereals was unexpected and raises questions on the extent and pattern of gene conversion in plant genome evolution, as well as the recombination mechanisms involved.

Previous studies on the evolution of the region duplicated between chromosomes 11 and 12 were carried out either on the two very closely related *O. sativa* subspecies (Wang et al. 2007) or widely divergent species (rice, sorghum, and *B. distachyon*), largely concentrating on protein-coding sequences. We chose three species from the *Oryza* genus to carry out a deep comparative study of the duplication at the genome level. In addition to the RefSeq, we selected two annual African species, *Oryza glaberrima* S. ( $2n = 24$ , AA genome) and *Oryza brachyantha* Chev. Et Roehr. ( $2n = 24$ , FF genome). The former has the same genome type as Asian domesticated rice (Linares 2002), whereas the latter, which diverged from the AA lineage ~15 Ma (Tang et al. 2010), has the smallest genome in the genus (340 Mb) (Uozu et al. 1997) and may display a faster evolution rate (Zou et al. 2008). The divergence of *O. sativa* and *O. glaberrima* genomes is estimated between 0.6 and 1 Ma (Ge et al. 2005; Zhu and Ge 2005; Roulin et al. 2010).

Here, we studied the extent and pattern of paralogous conversion between chromosomes 11 and 12 since the *O. brachyantha*/genome AA and *O. glaberrima*/*O. sativa* divergences, focusing on a region from 1.5 to 2.5 Mb overlapping the colinearity breakpoint. We show recent long-range conversion, particularly in *O. brachyantha*, involving both coding and noncoding sequences. The breakpoint is located in syntenic positions in all three species and we discuss the mechanisms that could explain these observations.

## Materials and Methods

A detailed version of all Methods is available in [supplementary text S1 \(Supplementary Material online\)](#).

### Sequencing, Assembly, and Contig Annotation

BAC contigs were defined using SyMAP (Soderlund et al. 2006) and refined manually. Lengths of assembled contigs

are reported in [table 1](#). Annotation was carried out using available tools and in-house Perl scripts, gene models being refined in Artemis (Rutherford et al. 2000). Overall statistics are presented in [table 1](#).

### Comparative Structural Analysis

Sequence conservation and rearrangement was analyzed with Dotter (Sonnhammer and Durbin 1995) using default parameters and with the Artemis comparison tool (ACT, Carver et al. 2008) for small rearrangements.

### Inference of Paralogous Pairs and Homologous Sextets

BlastN (Altschul et al. 1990) alignment was used to identify paralogous pairs for each species, with a cutoff  $e$  value of  $1 \times e^{-10}$ , and homologous sextets using *O. glaberrima* chromosome 11 CDSs as query sequences and retaining the best hit on each chromosome with minima of 60% identity and 10% length coverage. These criteria were defined empirically to take into account widely divergent genes and potential anomalies in annotation of poorly supported gene models. Corresponding CDSs were translated, amino acid sequences aligned with ClustalW (Thompson et al. 1994) and CDS aligned with bp\_mrtrans (Stajich J, [jasonatbioperl.org](http://jasonatbioperl.org)).

### Whole Contig Alignments

Finished contigs were aligned with Mauve (Darling et al. 2004), using minimum Locally Colinear Block (LCB) weight and backbone size at 100 and 50, respectively. Homologous colinear sequence blocks were aligned with ClustalW, as were intervening sequences. These data set were joined together and the resulting alignment split into 500 bp segments (including gaps). In total, 1539 blocks with six homologous sequences were analyzed. Gap information was coded with the simple indel coding (SIC) method (Simmons and Ochoterena 2000) using IndelCoder (Ogden and Rosenberg 2007).

### Evolutionary Distances, Phylogenetic, and GENECONV Analysis

For all paralogous gene pairs, pairwise synonymous (dS) and nonsynonymous (dN) substitution rates and nonsynonymous/synonymous ( $\omega$ ) substitution ratios were calculated with the basic maximum likelihood (ML) method of Goldman and Yang (1994). In order to detect functional constraint on both copies in paralogous gene pairs, we determined if the  $\omega$  values were significantly lower than 0.5 using the likelihood ratio test (LRT, Yang 1998; Betran et al. 2002). For genes in homologous sextets, random-site codon substitution models (Nielsen and Yang 1998), which allow the  $\omega$  to vary among codons, were implemented in CODEML (PAML 4.3, Yang 2007) and tested with the LRT (M0 vs. M3, M1 vs. M2, M7 vs. M8). Phylogenetic trees were reconstructed by ML and Bayesian inference (BI) methods. The DNA substitution model was selected using the Data-monkey webserver (Kosakovsky Pond and Frost 2005), with all sequences fitting the Hasegawa–Kishino–Yano (HKY85) model. ML was implemented with PhyML 3.0

**Table 1.** General Features of Contigs of *Oryza glaberrima* and *Oryza brachyantha* and Orthologous Segments on the MSU Rice Genome Annotation v6.1 Pseudomolecules of *Oryza sativa* ssp. *japonica* (RefSeq).

	RefSeq		<i>O. glaberrima</i>		<i>O. brachyantha</i>	
	Ch11	Ch12	Ch11	Ch12	Ch11	Ch12
Length (bp)	1,090,000	1,200,000	874,636	971,932	857,170	628,505
Genbank accession	/	/	FQ378034	FQ377974	FQ378032	FQ378033
Coordinates (kb) <sup>a</sup>	1.42–2.51	1.34–2.54	1.44–2.40	1.52–2.53	1.43–2.50	1.57–2.53
Number of genes	180	168	116	104	116	74
Density (genes/kb)	0.165	0.139	0.132	0.107	0.135	0.117
%GC	42.76	43.41	42.77	42.99	41.07	40.49
Coding %	37.4	32.4	33.6	29.8	37.1	30.4
TE%	15.3	33	10.2	8.5	1.8	3.7
Class I TE	18	47	18	18	6	4
Class II TE	43	36	32	24	3	5
MITES	153	136	115	101	74	117
Other	2	5	0	0	0	0

NOTE.—Numbers of genes do not include alternative splicing forms and CDS with TE-related annotations.

<sup>a</sup> Coordinates are relative to the RefSeq.

(Guindon and Gascuel 2003) and BI with MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001), with Nst = 2, Rates = Invgamma. For the 500 bp blocks of the whole contig analysis, nucleotide distances were inferred by BI. The data were partitioned according to data type: DNA (HKY85 model) and binary gap information. Statistical analysis and graph construction was performed using the R software (R Development Core Team 2011). GENECONV (Sawyer 1989) was used with the default settings.

## Results

### Global Structural Analysis

Despite the high level of conservation between each 1 Mb paralogous segment pair, comparison of orthologous chromosomes shows the extensive divergence of this region. For similar Blast minimal criteria, an ACT comparison emphasizes the strong divergence in the distal region between the orthologs in AA and FF genomes compared with the paralogous pairs (fig. 1). *Oryza sativa* and *O. glaberrima* orthologous contigs display weaker divergence than with the FF species. A complete list of large structural variants (>5 kb) is presented in detail in supplementary table S1 (Supplementary Material online). Indels involve both genes and repeat elements. The most striking rearrangement is

a specific inversion at the 5' end on *O. glaberrima* chromosome 12 covering at least 82 kb. Overall, contigs from *O. glaberrima* and *O. brachyantha* are shorter compared with the RefSeq *O. sativa* chromosomes (table 2). The expansion of the chromosome 11 segment in *O. sativa* compared with *O. glaberrima* results from eight insertions/deletions (indels) for a total of 83.4 kb (table 2). Four events, three indels, and one tandem duplication (~14 kb, 1,832,000–1,846,500 bp) contribute to the size expansion of both AA genomes compared with *O. brachyantha* (supplementary table S1, Supplementary Material online and table 2). *Oryza brachyantha* chromosome 11 also displays a specific inverted duplication of 42 kb.

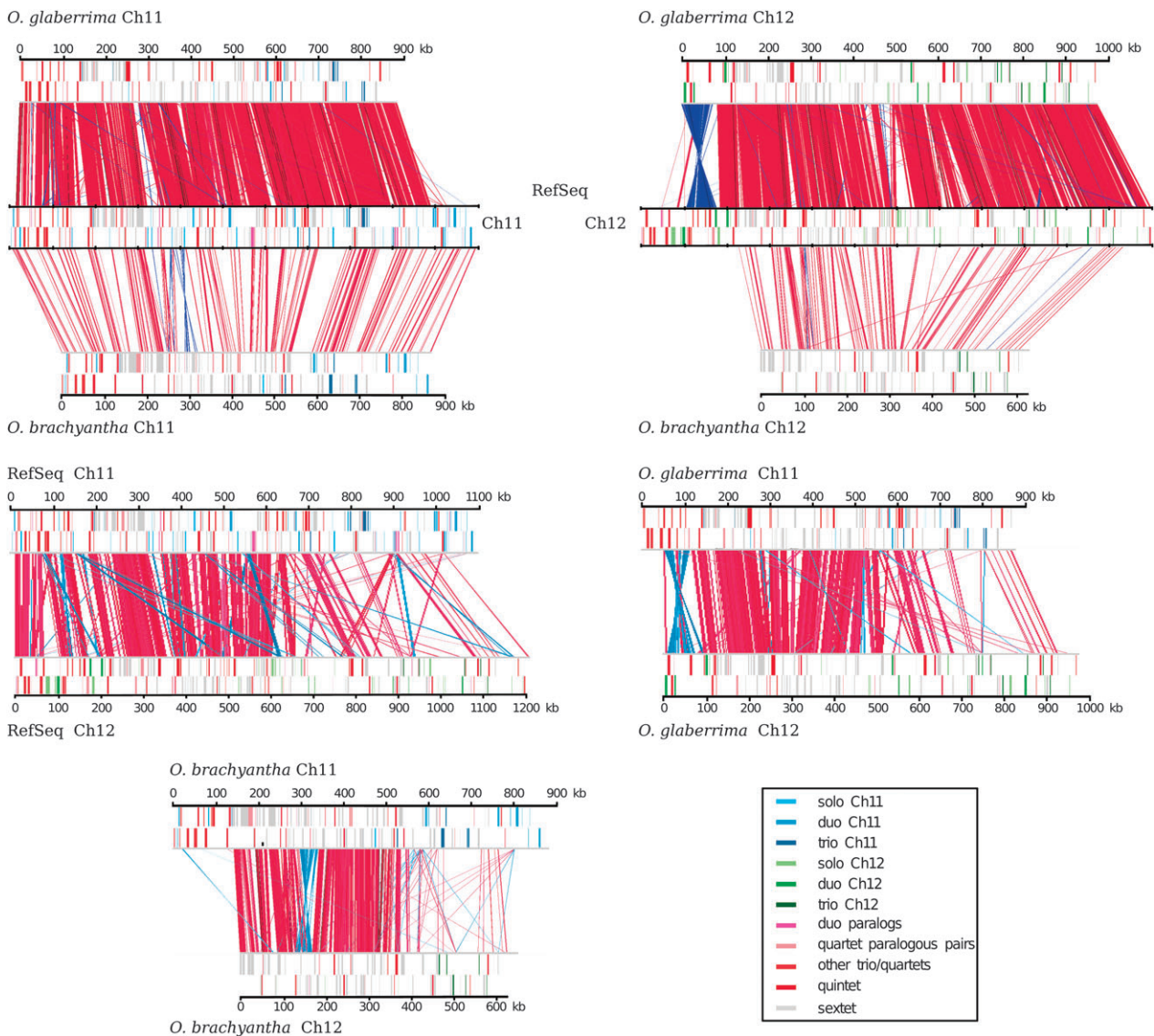
The expansion of the chromosome 12 segment in *O. sativa* compared with *O. glaberrima* is explained by five insertions, but the Og12 region also displays two large insertions and a tandem duplication (table 2). The size difference is particularly striking for *O. brachyantha* chromosome 12 (628,505 bp compared with 966,580 bp for the RefSeq orthologous region). Comparison with the two AA genomes identifies three large indels for a total contraction of 272.2 kb in *O. brachyantha*. On the first half of the largest insertion (~160 kb, RefSeq coordinates: 1,925,654–2,117,228), seven genes were annotated on the RefSeq (between Os12g04720 and Os12g04850), of which at least three are expressed, and four on *O. glaberrima*. The proximal region is composed of transposable elements in the RefSeq and is reshuffled in *O. glaberrima* (supplementary table S1, Supplementary Material online). Sequence analysis of the non-TE region showed significant nucleotide conservation only with sequences from AA genome species, suggesting that the genes may be de novo genes specific to the AA complex.

Of particular interest in the context of potential conversion are species-specific rearrangements shared by chromosomes 11 and 12 (supplementary table S1, Supplementary Material online). We observed four large events shared by paralogous chromosome pairs or in syntenic positions. *Oryza brachyantha* chromosomes 11 and 12 have insertions of ~20 kb on chromosome 11 and ~32 kb on

**Table 2.** Summary of Expansion Events between Analyzed Orthologous Segments on Chromosomes 11 and 12 in *Oryza sativa* (Os), *Oryza glaberrima* (Og), and *Oryza brachyantha* (Ob).

Segment	Segment Size Difference	Number of Expansions	Size Range	Total Size	
Os11/Og11	85.4	Os11	8	5.5–20.3	83.4
		Og11	1	9.6	9.6
Os11/Ob11	212.8	Os11	4	7.4–24.7	56.1
		Ob11	2	7.3–42.3	49.6
Os11/Og12	38.1	Os12	5	6.5–33.5	68.3
		Og12	3	11.6–15.5	39.7
Os11/Ob12	331.5	Os12	3	19–158	272.2
		Ob12	1	8.8	8.8

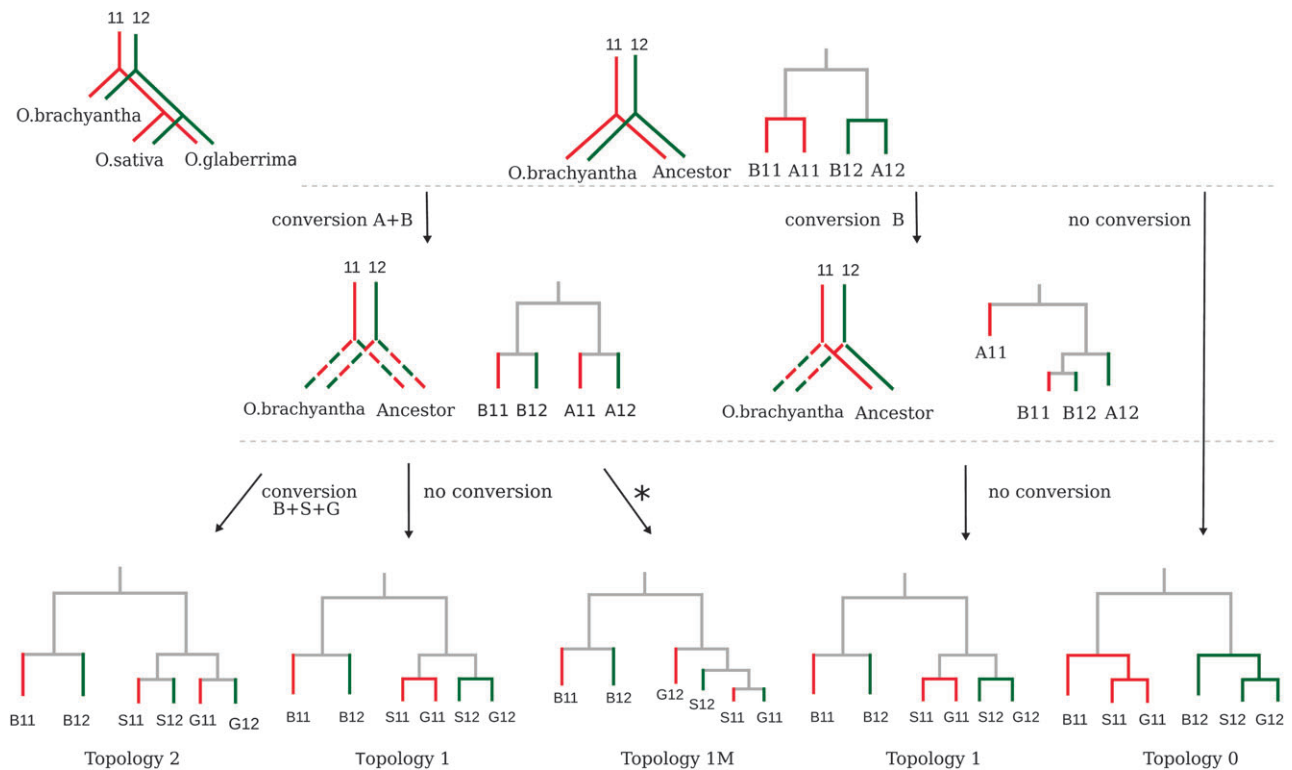
NOTE.—Both indels and tandem duplications more than 5 kb long are considered. Sizes are indicated in kilobases.



**Fig. 1.** Graphical representation of synteny between the orthologous and paralogous 11 and 12 contigs in the RefSeq, *Oryza glaberrima*, and *Oryza brachyantha*. Coordinates are indicated in kilobases. The segments for the RefSeq correspond to 1.42–2.51 Mb on chromosome 11 and 1.34–2.54 Mb on chromosome 12. Lines represent sequence similarity comparison by BlastN, with blue lines representing inverted matches. The minimum score and size of matches are 300 and 300 bp, respectively. The CDS composition of each contig is shown, with a color code indicating their presence/absence on the six homologous chromosomes.

chromosome 12 in common and a tandem duplication spanning  $\sim 16$  kb. The latter contains two pairs of annotated genes. Construction of phylogenetic trees of the coding sequences using the AA genome sequences as out-group (supplementary fig. S1, Supplementary Material online) clearly shows a topology of (Ob11-1, Ob12-1), (Ob11-2, Ob12-2), indicative of gene conversion rather than independent duplication. The four AA lineage chromosomes share two expansions compared with *O. brachyantha*, the first varying from 10 to 38 kb and the second covering approximately 29 kb (supplementary table S1, Supplementary Material online). The most parsimonious explanation for these rearrangements conserved between paralogous chromosomes, but which are specific to the two lineages, is concerted evolution since their divergence at the time of the WGD, after speciation events.

We found 65 CDSs conserved on all chromosomes (sextets: see fig. 1 and supplementary table S2, Supplementary Material online). A further 20 were absent only on *O. brachyantha* chromosome 12, consistent with the observed deletions. Six were observed in the AA genomes, but not in *O. brachyantha*, whereas one was absent only in *O. glaberrima*. Only seven CDSs were specific to orthologous chromosomes 11 and six to chromosomes 12, all except one located at the proximal end, confirming the widespread homogenization of the distal ends of the duplicated blocks. Three, 9, 3, 4, 29, and 27 genes are specific to Og11, Og12, Ob11, Ob12, Os11, and Os12, respectively. The greater number for the RefSeq sequences can be explained by our stringent annotation for the wild species, as at least nine and six of the CDSs on Os11 and Os12, respectively, are TE related, although they are not annotated as such.



**Fig. 2.** Evolutionary scheme of the 11–12 duplicated block in the *Oryza* genus, as a function of conversion events in the FF and AA lineages. A = Ancestor of AA lineage, B = *Oryza brachyantha*, G = *Oryza glaberrima*, S = *Oryza sativa*. Conversion is inferred based on topological incongruity with the topology 0. \*Only one example of topology 1M is shown as we group several trees in this class: The first have only one orthologous pair, S11–G11 or S12–G12, clustered in a terminal branch, whereas the two remaining genes form intermediate branches between this cluster and the *O. brachyantha* node. The second have only one paralogous pair, S11–S12 or G11–G12, clustered in a terminal branch, whereas the two remaining genes form intermediate branches between this cluster and the *O. brachyantha* node. This topology is ambiguous as it could reveal 1) too weak divergence of the four AA genes to resolve their phylogenetic relationships, 2) the strong divergence of one of these genes blurring their true relationships, and 3) conversion in one of the AA lineages after their divergence.

### Gene Conversion between Paralogous Coding Sequences

We applied a topological incongruity approach (Gao and Innan 2004; Lin et al. 2006) to the sextets. Fifteen contained redundant sequences, resulting from local duplication on one or several of the six chromosomes and were excluded from the analysis. Figure 2 shows the topologies expected under different evolutionary schemes. Topology 0 is the null hypothesis, indicative of no conversion events. Topology 2, where all paralogous pairs are grouped together, is expected if gene conversion has occurred separately in all lineages since their divergence. Topology 1, in which *O. sativa* and *O. glaberrima* orthologs group together and *O. brachyantha* copies form their own clade, indicates conversion specific to *O. brachyantha*. In topology 1M, one orthologous *O. sativa*/*O. glaberrima* pair (11 or 12) forms a terminal node with one of the paralogous genes, whereas the other is more distant in the tree. This topology, indicative of conversion in *O. brachyantha*, is not informative on the relationships between *O. sativa* and *O. glaberrima*, as several hypotheses can explain it.

Using BI methods, 24 of 50 sextets present topology 1 and 15 topology 1M (table 3 and supplementary table S2, Supplementary Material online). For two 1M sextets (Os11g04200 and Os11g04500), the distances between the four *O. sativa*

and *O. glaberrima* sequences are too weak to distinguish the relationships clearly and for four (Os11g04274, Os11g04360, Os11g4570, and Os11g04650), one of the sequences is highly divergent, putatively indicative of pseudogenization. For the last nine 1M sextets, the topology and distances observed could indicate conversion of one of the two paralogous pairs or a greater divergence in one pair.

We found no topology 2 trees and only seven sextets indicated lack of conversion (topology 0), all located in the proximal region of the contigs, after sextet Os11g04980. However, this region also contains three sextets showing conversion in *O. brachyantha*. Finally, four showed uninterpretable topology 3. Eight trees were incongruent between Bayesian and ML methods, most moving between topologies 1 and 1M. These results suggest widespread conversion in *O. brachyantha* since its divergence from the AA lineage, notably in the distal region.

### Nongenetic Conversion

Recombination is not exclusively observed in intragenic regions (Mézard 2006). The availability of megabase-sized sequences from closely related species allows the identification of conversion on a large scale, in both gene and non-gene regions. We first tested the frequently used program

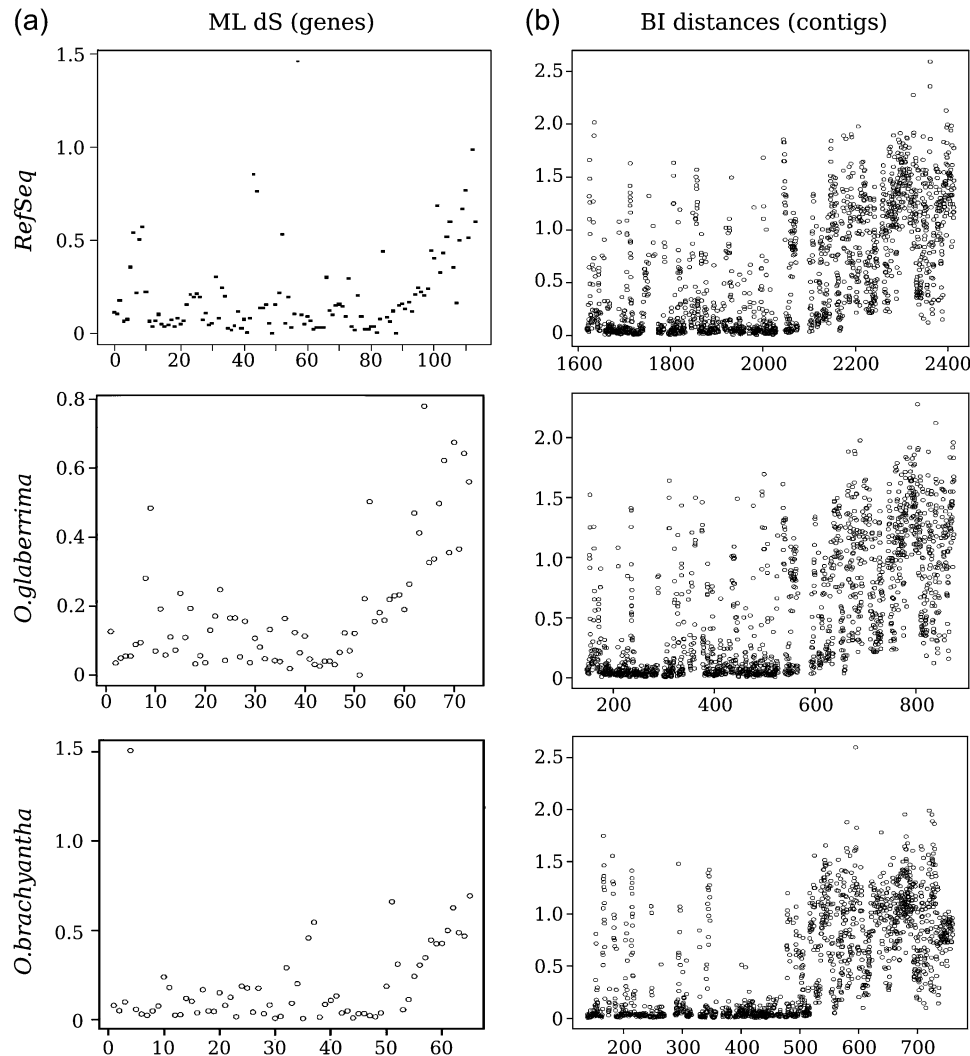
**Table 3.** Topology Data for Sextets of CDS and Whole Contig Blocks (Divided into Three Zones).

	Topology					Total
	1	1M	0	2	3	
CDS sextets	24	15	7	0	4	50
Whole sequence						
Zone 1	515	136	1	1	147	800
Intermediate zone	32	1	30	0	135	198
Zone 2	5	0	404	0	132	541
Total	552	137	435	1	414	1539
% Zone 1	93.3	99.3	0.2	100	35.5	
% Intermediate zone	5.8	0.7	6.9	0	32.6	
% Zone 2	0.9	0	92.9	0	31.9	

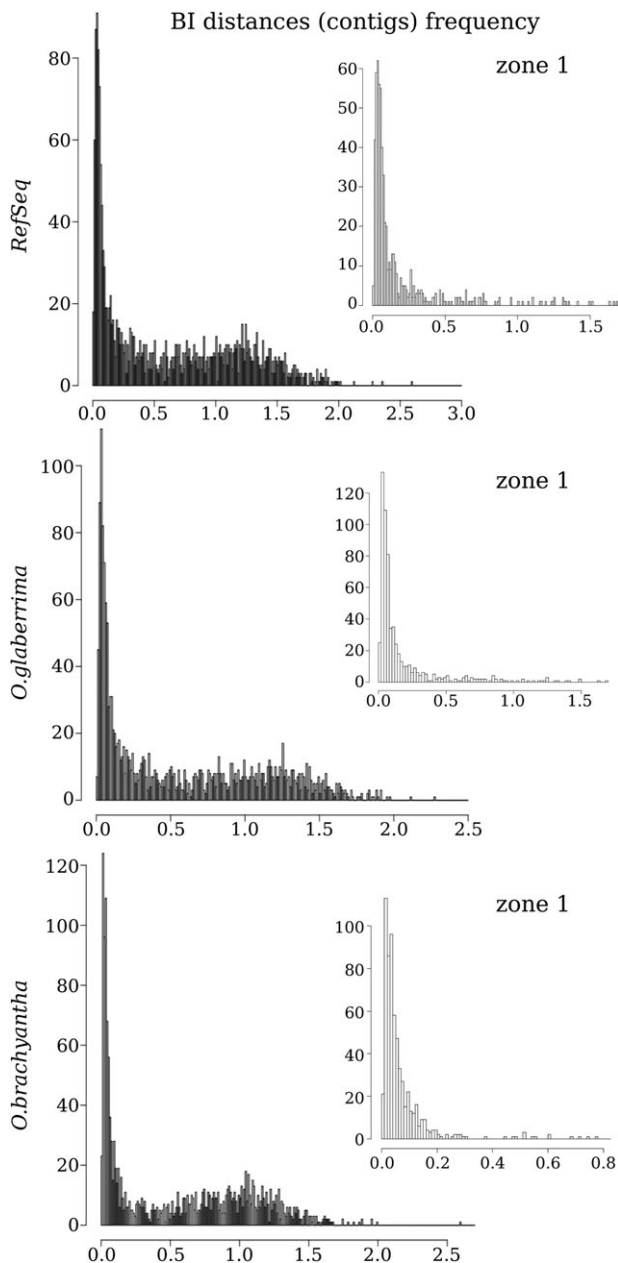
GENECONV on the CDS sextet data set (see Results in [supplementary table S2, Supplementary Material](#) online). Among the 27 sextets where conversion tracts were detected, 19 display topology 1 or 1M. For seven of these, GENECONV found converted fragments only for *O. sativa* and *O. glaberrima* pairs, although we also expected conversion for *O. brachyantha* copies. More surprisingly, GENE-

CONV did not detect conversion tracts for *O. brachyantha* in the remaining 21 sextets with topology 1 and 1M. This apparent contradiction with the topological incongruity analysis may be explained by the failure of GENECONV to detect conversion events when the duplicated region is highly homogenized (McGrath et al. 2009). This confirms the prediction of Mansai and Innan (2010) that GENECONV detects few regions in the case of large-scale gene conversion and can only give indications on events which are both local and relatively recent.

As GENECONV proved to be an unsatisfactory tool, we adapted a topological approach, incorporating indel coding, to look for evidence of conversion throughout the 1 Mb region. Mauve alignment was used to identify conserved blocks between the six genomic sequences, choosing 500 bp segments for topological analysis as gene conversion tracts described in the literature range from a few base pairs to 3 kb (Kuang et al. 2004; Mondragon-Palomino and Gaut 2005; Chen et al. 2007; Xu et al. 2008; Benovoy and Drouin 2009). This approach inevitably



**Fig. 3.** Spatial distribution of synonymous substitution rates (dS) between paralogous gene pairs computed with the basic ML codon model, plotted against the number of pairs (a) and BI nucleotide distances between paralogous 500 bp fragments of the whole contig alignment, plotted against the chromosome 11 coordinates (kb) for the three species (b).



**FIG. 4.** Frequency distribution of BI nucleotide distances between paralogous 500 bp fragments of the whole contig alignment. The inset histograms show distance distributions in converted zone 1 only.

produces a number of uninformative alignments and, among the 1539 trees examined, those with strongly divergent branches were classified as topology 3 (table 3).

The distribution of the tree topologies along the 1 Mb sequence is not random, defining three regions (table 3 and supplementary table S2, Supplementary Material online). The distal region (zone 1), where more than 80% of the trees display topologies 1 (515) or 1M (136), extends to block 1,047,501–1,048,000 (total 800 blocks), corresponding to 2,108,257 and 2,151,747 bp on *O. sativa* chromosomes 11 and 12, respectively. As for the CDS, topologies 1M are mainly indicative of very weak distances between the four *O. sativa* and *O. glaberrima* contigs. Only

one block in this region has topology 0 (856,501–857,000) and only one (663,001–663,500) suggests independent conversion in both *O. sativa* and *O. glaberrima* (topology 2). For the 652 blocks displaying topologies 1, 1M, and 2 in the first zone, 337 (~106,000 bp), 414 (135,000 bp), and 415 (126,000 bp) are located in intergenic regions for *O. sativa*, *O. glaberrima*, and *O. brachyantha*, respectively, whereas 315 (120,000 bp), 238 (94,000 bp), and 237 (93,000 bp) overlap protein-coding sequences.

The proximal region (zone 2) extends from block 1182001–1182500 to the end and covers RefSeq chromosome 11 from 2,195,478 bp and chromosome 12 from 2,214,633 bp. Most trees in this region show topology 0 (404; 75%) with only five isolated topology 1 alignments. Nine CDS sextets were found in this area (beginning after sextet Os11g05050), all classified as topology 0 except for two showing topology 1 (Os11g05320 and Os11g05370). However, we did not find topology 1 in the 500 bp blocks corresponding to these two loci (1511501\_1512000 to 1516501\_1517000 and 1551501\_1552000 to 1552001\_1552500). This could be explained by the presence of introns and coding of gaps in the whole contig analysis, suggesting rather local conversion events limited to CDSs. The intermediate zone displays a balanced ratio of topologies 1 and 0, and a high percentage of topologies 3 (135, 68%), indicating considerable rearrangement.

The uniformity of conservation of large tracts of both coding and noncoding sequences in the distal regions is indicative of long-range mechanisms rather than small and repetitive recombination events. Nonetheless, our GENECONV analysis and observations of topologies 1M in phylogenetic analysis confirm that regular small-scale conversion may have occurred since the divergence of the AA species but no extensive homogenization. In the proximal regions, we found 23, 16, and 9 paralogs in *O. sativa*, *O. glaberrima*, and *O. brachyantha*. This conservation of isolated coding sequences after the breakpoint of conservation could be due to local conversion events but may simply reflect slowly diverging gene pairs, generated by older conversion events.

### Finding the Limits and Dating the Conversion Events

Figure 3 displays the synonymous substitution rates (dS) resulting from ML analysis for all paralogous gene pairs and the nucleotide distances inferred by the BI between pairs of fragments from the whole contig analysis, plotted against their positions on the contigs. There is a clear rupture in the distribution in all three species, values being low in the first two-thirds of the region, increasing clearly in the proximal region. The breakpoint in the whole contig analysis is located between 2,100,000 and 2,106,000 bp on *O. sativa* chromosome 11, corresponding to 2,120,000–2,128,000 bp on *O. sativa* chromosome 12, in agreement with the topological analysis on sextets. It is at syntenic locations in *O. glaberrima*, between 591,500–597,000 bp and 599,500–606,000 bp on *O. glaberrima* contigs 11 and 12, respectively. The *O. brachyantha* breakpoint is slightly more proximal, between 518,000–519,000 bp and 374,000–375,000 bp on contigs 11 and 12



(2,118,000 and 2,155,000 bp on RefSeq chromosome 11 and 12, respectively). These breakpoints all map to the intermediate region as described above.

The distributions of nucleotide distance values for the paired 500 bp fragments show a bimodal distribution, with the first peak corresponding to zone 1 (fig. 4). Distributions of distance values for zone 1 (fig. 4, small histograms) indicate that these regions of ~0.6 Mb were homogenized at the same time either by one unique conversion event or by several concomitant long-range events. The first peak is at 0.03–0.04 for the AA species and 0.01–0.02 for *O. brachyantha*, indicative of more recent conversion in the FF genome. Furthermore, the mean distance between *O. brachyantha* contig pairs (0.07) is lower than that of the AA pairs (0.17) (supplementary table S2, Supplementary Material online). The second peak represents the distances between the sequences in the nonconverted contig ends (1.25–1.26 for the AA species and 1.04–1.05 for *O. brachyantha*). Distributions of dS rate for the paired genes display a unimodal distribution with peaks at 0.02–0.04, 0.04–0.06, and 0.02–0.04 for *O. sativa*, *O. glaberrima*, and *O. brachyantha*, respectively, consistent with the whole contig analysis (results not shown).

Based on a divergence time of 15 Ma for *O. brachyantha* in the genus and ~0.8 Ma for the divergence of *O. sativa* and *O. glaberrima*, we estimated the relative time of the last conversion event for each paralogous pair using the median dS and nucleotide distance values among the orthologs and paralogs (supplementary table S2, Supplementary Material online) using the formulas:

$$x(p_{11}, p_{12}) = (\text{median}(d(p_{11}, p_{12})) \times 0.8) / \text{mean}(\text{median}(d(Os_{11}, Og_{11})), \text{median}(d(Os_{12}, Og_{12})))$$

$$x(p_{11}, p_{12}) = (\text{median}(d(p_{11}, p_{12})) \times 15) / \text{mean}(\text{median}(d(Os_{11}, Ob_{11})), \text{median}(d(Os_{12}, Ob_{12})))$$

where p11 and p12 are the paralogous pair considered and d(a,b) either the dS or the BI distance.

Considering only zone 1, the last conversion events were dated between 2.5 and 4.0 Ma for the AA species and 1.5–3.5 Ma for *O. brachyantha*, much lower than previous estimations, from 5 to 21 Ma, given for the whole region in *O. sativa* but based only on coding sequences (Goff et al. 2002; Rice Chromosomes 11 and 12 Sequencing Consortia 2005; Wang et al. 2005; Salse et al. 2008). Using pairs from zone 2, we calculate 15–55 Ma for the AA species and 20–50 Ma for *O. brachyantha*. Age estimations for the WGD event are somewhat greater (50–90 Ma, Chaw et al. 2004; Yu et al. 2005) but the difference is easily explained by the small size of the region, local conversion events since the duplication or traces of older conversion events.

### Paralog Divergence after Conversion

Large-scale conversion events as described here reset the evolutionary clock and harmonize both coding and essential noncoding regions. We have analyzed the divergence and selection pressure on the 11 and 12 paralogous copies, because we thought that could indicate, indirectly, the role of this recurrent homogenization. If paralogous func-

tionally redundant copies are conserved identically, we should see purifying selection, whereas if the copies are evolving toward pseudogenization, subfunctionalization, or neofunctionalization, we would expect to observe signals of neutral evolution or positive selection (Innan and Kondrashov 2010). Studies using tiling arrays (Li et al. 2005) or microarrays (Throude et al. 2009) did not detect significantly different expression patterns between gene pairs in the 11–12 duplication. However, Yim et al. (2009) observed that between 50.9% and 67.3% of 55 gene pairs in the block may have diverged in their expression, so no clear conclusion can be drawn. We compared the nonsynonymous/synonymous ratios ( $\omega$ ) for paralogs in the three species and tested for selection pressures.

We found 122, 76, and 67 paralogous pairs in the RefSeq, *O. glaberrima* and *O. brachyantha* sequences, respectively and eliminated those with null dS values. The  $\omega$  ratio, calculated by the method of Goldman and Yang (1994), ranged from 0.001 to 1.042 (mean  $0.3 \pm 0.02$ ), 0.001 to 1.282 ( $0.25 \pm 0.02$ ), and 0.001 to 1.560 ( $0.34 \pm 0.03$ ), in Os, Og, and Ob, respectively. Only two pairs in *O. sativa* displayed  $\omega = 1$  (neutrality level) and one pair for each other species displayed  $\omega > 1$  (indicator of positive selection). Under the LRT, among 245 paralogous pairs, 112 showed an  $\omega$  value that was significantly lower than 0.5 with  $P < 0.05$  (71 pairs with  $P < 0.001$ ), indicating that duplicated copies are both under purifying selection. The Benjamini–Hochberg procedure for controlling the false discovery rate in multiple comparisons was implemented at the  $\alpha = 0.05$  level, and ratios for 103 paralogous pairs were still significantly  $< 0.5$  at  $P < 0.05$  (45 for the RefSeq, 38 for *O. glaberrima*, and 20 for *O. brachyantha*).

Random-site codon substitution models were applied to sextets in order to test the presence of positive Darwinian selection at individual sites. The one-ratio model (M0) gives the average  $\omega$  over all sites and branches for each data set and this ranged from 0.004 to 0.57, still indicating the overwhelming role of purifying selection. The LRT indicates that M3 fits the data significantly better than M0 for 36 sextets (degrees of freedom [df] = 4,  $P = 0.05$ ), indicating significant variation in selective constraints among sites. For 22 sextets, both models M2 and M8, which allow the  $\omega$  ratio to exceed 1, fit the data significantly better (df = 2,  $P = 0.05$ ) than models M1 and M7 (supplementary table S2, Supplementary Material online). The number of sites with  $\omega > 1$  varied from 5 to 142.

Thus, a certain fraction of duplicated pairs (42%, 52%, and 32% in Os, Og, and Ob, respectively) are under purifying selection in the region under study suggesting they could tend to diverge slowly after conversion, whereas only 22 pairs common to all three species display positive selection on a fraction of codons.

### Discussion

We have demonstrated that the duplicated blocks between 1.5 and 2.1 Mb on the RefSeq chromosomes 11 and 12, and orthologous regions in *O. glaberrima* and *O. brachyantha*,

are uniformly homogenized by long-range recombination mechanisms. Our observation of syntenic breakpoints of conservation in the AA (*O. sativa* and *O. glaberrima*) and FF (*O. brachyantha*) lineages suggests that conversion is recurrently initiated around this point (2.1 Mb on the RefSeq), indicative of a putative hot spot of recombination. This is coherent with the fact that, in Poaceae, recombination increases with relative distance from the centromere (Wu et al. 2003; Anderson et al. 2004; Kao et al. 2006) and is greater in gene-dense regions near the telomeres (Mézard 2006). Two studies provide estimations of recombination rates along the 12 chromosomes in rice, and both support our hypothesis (Rizzon et al. 2006; Tian et al. 2009). Indeed, both chromosomes 11 and 12 display a high recombination rate ( $\sim 12$  cM/Mb and  $>12$  cM/Mb, respectively in Tian et al. 2009) between 2 and 3 Mb from the short arm telomere. The peak is more striking for chromosome 12 compared with the surrounding regions.

The extent of gene conversion depends on the recombination process involved, but we have no evidence allowing us to favor one particular mechanism. Nonetheless, we can exclude noncrossover DSB and SDSA as they generally yield small conversion tracts, less than a few kilobases (Mancera et al. 2008). Two mechanisms could potentially explain the large conversion tracts observed. A DSB event associated with half crossing over between the short arm ends of these two chromosomes would lead to reciprocal exchange between the two chromatids. This could generate gametes with conversion tracts depending on how the chromatids segregate. The second process is BIR, which is initiated as DSB, following a DSB where just one of the two ends can undergo homology-dependent strand invasion (Llorente et al. 2008). It continues with a processive replication fork, and DNA synthesis proceeds to the end of the donor chromosome (Llorente et al. 2008). BIR have been implicated in homogenization of subtelomeric regions in yeast (Bosco and Haber 1998) and their relative frequency increases toward telomeric regions, in which their consequences are less deleterious than in other regions of the chromosomes (Ricchetti et al. 2003). The 11–12 duplicated block extends beyond the limit of the subtelomeric regions ( $\sim 500$  kb from the distal end, Fan et al. 2008), but the underlying mechanisms of BIR (reviewed for the yeast model in Lydeard et al. 2007; Llorente et al. 2008) do not limit the size of the fragment which is reconstructed. These two mechanisms are described as putative models of formation of segmental duplications (Kozul and Fischer 2009), which was the first hypothesis proposed for the 11/12 duplication (Goff et al. 2002).

We propose that conversion events have recurrently replaced large segments of one chromosome with homologous sequences from another, which implies the recurrence of meiotic pairing of nonhomologous chromosomes 11 and 12 since their formation by polyploidization, certainly facilitated by the maintenance of redundancy in their telomeric and subtelomeric regions which obscure true homologous relationships.

Whatever the mechanism leading to this duplication, it has not occurred independently in the two AA species since their divergence. This extends the observations of Wang et al. (2007) on the *O. sativa* subspecies who found very few partial-gene conversion events and only two whole-gene conversions, both in *O. sativa* ssp. *japonica*. To our knowledge, the 11–12 duplication and its orthologs in sorghum and Brachypodium (Wang et al. 2011) represent the first described example of such long-term conservation of two duplicated segments in plants.

Based on our calculation of selective pressure on paralogous gene pairs, we can not exclude the possibility that the presence and maintenance of the recombination hot spot and long-range gene conversion are selected themselves for the benefits of buffering crucial functionality. However, no particular class of genes have been identified in the segments. The Rice Chromosomes 11 and 12 Sequencing Consortia (2005) came to the conclusion that chromosomes 11 and 12 are enriched in disease resistance gene clusters, but these are not preferentially located to the distal 2 Mb of the chromosomes and are rather known for their variability. No significant bias of Pfam domain composition or GO categories was found in the converted genes in rice and sorghum genomes (Wang et al. 2009 and our unpublished observations).

Our comparative study highlights considerable divergence, not only between the AA and FF genomes but also between the two AA genomes, including de novo gene formation. If we consider only interspecific rearrangements larger than 10 kb with genes involved, we observe one insertion (five genes) specific to the RefSeq, one expansion for the AA lineage compared with *O. brachyantha* (four genes), and two tandem duplication, one for *O. brachyantha* (involving ten genes) and one for the AA species (two genes), all on the chromosome 11 1 Mb segment. On chromosome 12, we observed one inversion (seven genes), one expansion (two genes) specific to *O. glaberrima*, and one expansion (five genes) on *O. sativa*. Contractions compared with the RefSeq (6, 2, and 19 genes) were particularly striking on *O. brachyantha* chromosome 12. Genome expansions and contractions in the 11–12 duplicated region (15 and 12, respectively) in a short evolutionary time frame, involving up to one-third of the genome sequence, are strikingly different from the highly conserved gene colinearity observed in the comparative studies of MONOCULM1-orthologous regions (2.4 Mb, chr6) in 14 *Oryza* genomes (Lu et al. 2009). This latter region is disrupted by only three rearrangements (a three-gene segment translocation in *O. coarctata*, a three-gene segment insertion in *O. sativa*, and a single gene tandem duplication in *O. granulata*).

Wang et al. (2011) recently showed that ectopic concerted evolution acting on the duplicated blocks in rice chromosomes 11 and 12 and homologous sorghum chromosomes 5 and 8 has significantly increased gene divergence between lineages compared with the genome-wide average, particularly in the more distal ends of these blocks which show the greatest intragenomic similarity. Whereas these studies concerned gene content and

divergence, our studies on structural rearrangements lead to the same conclusion. Two segments derived from the initial duplication event will diverge independently and accumulate structural variants. Subsequent interspecies divergence will depend on the timing of speciation and conversion events, as well as on the direction of conversion. After speciation (species A and B), if conversion occurs from chromosome 11 to 12 in A, and from chromosome 12 to 11 in B, the comparison between A11–B11 or A12–B12 represents the divergence since the duplication and not since the speciation. Repetitive cycles of divergence and alternative conversion will increase the distance between orthologous pairs.

Gene-scale conversion is already incorporated in the classical models of the evolution of duplicated genes (Teshima and Innan 2004; Gay et al. 2007; Innan 2009; Innan and Kondrashov 2010) and the occurrence of conversion between homeologous genes during polyploid formation and divergence (Udall et al. 2005; Salmon et al. 2009), or between the two LTR of a retrotransposon (Kijima and Innan 2010) have also been discussed. However, the story of conversion in the 11–12 distal ends is currently unique in genome evolution. Further comparative genomic and genetic studies within and outside the *Oryza* genus will be useful to confirm our hypothesis and clear up the mystery of possible functionality and benefits of this genome redundancy.

## Supplementary Material

Supplementary text S1, tables S1 and S2, and figure S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank Michael Rosenberg for supplying his program IndelCoder, and Yves Desdèves for his helpful advice on the Bayesian method. This work was supported by the Centre National de la Recherche Scientifique (Cooke laboratory). Sequencing was financed by Génoscope CNS.

## References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Anderson LK, Salameh N, Bass HW, Harper LC, Cande WZ, Weber G, Stack SM. 2004. Integrating genetic linkage maps with pachytene chromosome structure in maize. *Genetics* 166:1923–1933.

Benovoy D, Drouin G. 2009. Ectopic gene conversions in the human genome. *Genomics* 93:27–32.

Betran E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res.* 12:1854–1859.

Bosco G, Haber JE. 1998. Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. *Genetics* 150:1037–1047.

Carver T, Berriman M, Tivey A, Patel C, Böhme U, Barrell BG, Parkhill J, Rajandream M. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24:2672–2676.

Chaw SM, Chang CC, Chen HL, Li WH. 2004. Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J Mol Evol.* 58:424–441.

Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet.* 8:762–775.

Darling ACE, Mau B, Blatter FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394–1403.

De Muyst A, Mercier R, Mézard C, Grelon M. 2009. Meiotic recombination and crossovers in plants. *Genome Dyn.* 5:14–25.

Dujon B. 2010. Yeast evolutionary genomics. *Nat Rev Genet.* 11:512–524.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Ann Rev Genomics Hum Genet.* 10:285–311.

Ezawa K, Ikeo K, Gojobori T, Saitou N. 2010. Evolutionary pattern of gene homogenization between primate-specific paralogs after human and macaque speciation using the 4-2-4 method. *Mol Biol Evol.* 27:2152–2171.

Fan C, Zhang Y, Yu Y, Rounsley S, Long M, Wing RA. 2008. The subtelomere of *Oryza sativa* chromosome 3 short arm as a hot bed of new gene origination in rice. *Mol Plant.* 1:839–850.

Gao L, Innan H. 2004. Very low gene duplication rate in the yeast genome. *Science* 306:1367–1370.

Gay J, Myers S, McVean G. 2007. Estimating meiotic gene conversion rates from population genetic data. *Genetics* 177:881–894.

Ge S, Guo Y, Zhu Q. 2005. Molecular phylogeny and divergence of the rice tribe Oryzaceae, with special reference to the origin of the genus *Oryza*. In: Toriyama K, Heong KL, Hardy B, editors. Rice is life: scientific perspectives for the 21st century. Los Banos, Philippines: International Rice Research Institute Publications. p. 40–44.

Goff SA, Ricke D, Lan TH, et al. (55 co-authors). 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296:92–100.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11:725–736.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.

Hahn MW, Han MV, Han S. 2007. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* 3:e197.

Hogan CC, Bettencourt BR. 2009. Duplicate gene evolution toward multiple fates at the *Drosophila melanogaster* HIP/HIP-replacement locus. *J Mol Evol.* 68:337–350.

Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.

Innan H. 2009. Population genetic models of duplicated genes. *Genetica* 137:19–37.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet.* 11:97–108.

Jacquemin J, Laudie M, Cooke R. 2009. A recent duplication revisited: phylogenetic analysis reveals an ancestral duplication highly-conserved throughout the *Oryza* genus and beyond. *BMC Plant Biol.* 9:146.

Kao FI, Cheng YY, Chow TY, Chen HH, Liu SM, Cheng CH, Chung MC. 2006. An integrated map of *Oryza sativa* L chromosome 5. *Theor Appl Genet.* 112:891–902.

Kijima TE, Innan H. 2010. On the estimation of the insertion time of LTR retrotransposable elements. *Mol Biol Evol.* 27:896–904.

Kosakovsky Pond SL, Frost SD. 2005. DATAMONKEY: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.

Kozul R, Fischer G. 2009. A prominent role for segmental duplication in modeling eukaryotic genomes. *C R Biol.* 332:254–266.

- Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *The Plant Cell*. 16:2870–2894.
- Li WH, Yang J, Gu X. 2005. Expression divergence between duplicate genes. *Trends Genet*. 21:602–607.
- Lin YS, Byrnes JK, Hwang JK, Li WH. 2006. Codon-usage bias versus gene conversion in the evolution of yeast duplicate genes. *Proc Natl Acad Sci U S A*. 103:14412–14416.
- Linares OF. 2002. African rice (*Oryza glaberrima*): history and future potential. *Proc Natl Acad Sci U S A*. 99:16360–16365.
- Llorente B, Smith CE, Symington LS. 2008. Break-induced replication. *Cell Cycle*. 7:859–864.
- Lu F, Ammiraju JS, Sanyal A, et al. (15 co-authors). 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci U S A*. 106:2071–2076.
- Lydeard JR, Jain S, Yamaguchi M, Haber JE. 2007. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* 448:820–823.
- Mancera E, Bourgon R, Brozzi A, Huber W, Steinmetz LM. 2008. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454:479–485.
- Mansai SP, Innan H. 2010. The power of the methods for detecting interlocus gene conversion. *Genetics* 184:512–527.
- McGrath CL, Casola C, Hahn MW. 2009. Minimal effect of ectopic gene conversion among recent duplicates in four mammalian genomes. *Genetics* 182:615–622.
- Mézard C. 2006. Meiotic recombination hotspots in plants. *Biochem Soc Trans*. 34:531–534.
- Mondragon-Palomino M, Gaut BS. 2005. Gene conversion and the evolution of three leucine-rich repeat gene families in *Arabidopsis thaliana*. *Mol Biol Evol*. 22:2444–2456.
- Nielsen R, Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.
- Ogden TH, Rosenberg MS. 2007. How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenet Evol*. 42:817–826.
- Paterson AH, Bowers JE, Bruggmann R, et al. (45 co-authors). 2009. The Sorghum bicolor genome and the diversification of grasses. *Nature* 457:551–556.
- R Development Core Team. 2011. R: a language and environment for statistical computing, reference index version 2.13.0. Vienna, Austria: R Foundation for Statistical Computing. [cited 2011 June 7]. Available from: <http://www.R-project.org>.
- Ricchetti M, Dujon B, Fairhead C. 2003. Distance from the chromosome end determine the efficiency of double strand break repair in subtelomeres of haploid yeast. *J Mol Biol*. 328:847–862.
- Rice Chromosomes 11 and 12 Sequencing Consortia. 2005. The sequence of rice chromosomes 11 and 12, rich in disease resistance genes and recent gene duplications. *BMC Biol*. 3:20.
- Rizzon C, Ponger L, Gaut BS. 2006. Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Comput Biol*. 9:e115.
- Roulin A, Chaparro C, Piégu B, Jackson S, Panaud O. 2010. Paleogenomic analysis of the short arm of chromosome 3 reveals the history of the African and Asian progenitors of cultivated rices. *Genome Biol Evol*. 2010:132–139.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
- Salmon A, Flagel L, Ying B, Udall JA, Wendel JF. 2009. Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol*. 186:123–134.
- Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C. 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. *Plant Cell*. 20:11–24.
- Sanyal A, Jetty AS, Lu F, et al. (14 co-authors). 2010. Orthologous comparisons of the Hd1 region across genera reveal Hd1 gene lability within diploid *Oryza* species and disruptions to micro-synteny in sorghum. *Mol Biol Evol*. 27(11):2487–2506.
- Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol Biol Evol*. 6:526–538.
- Simmons MP, Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol*. 49:369–381.
- Soderlund C, Nelson W, Shoemaker A, Paterson A. 2006. SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res*. 16:1159–1168.
- Sonnhammer EL, Durbin R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167:GC1–G10.
- Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. 1983. The double-strand-break repair model for recombination. *Cell* 33:25–35.
- Tang L, Zou X, Achoundong G, Potgieter C, Second G, Zhang D, Ge S. 2010. Phylogeny and biogeography of the rice tribe (Oryzaceae): evidence from combined analysis of 20 chloroplast fragments. *Mol Phylogenet Evol*. 54:266–277.
- Teshima KM, Innan H. 2004. The effect of gene conversion on the divergence between duplicated genes. *Genetics* 166:1553–1560.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 22:4673–4680.
- Throude M, Bolot S, Bosio M, et al. (14 co-authors). 2009. Structure and expression analysis of rice paleo duplications. *Nucleic Acids Res*. 37:1248–1259.
- Tian Z, Rizzon C, Du J, Zhu L, Bennetzen JL, Jackson SA, Gaut BS, Ma J. 2009. Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Res*. 19:2221–2230.
- Udall JA, Quijada PA, Osborn TC. 2005. Detection of chromosomal rearrangements derived from homeologous recombination in four mapping populations of *Brassica napus* L. *Genetics* 169:967–979.
- Uozu S, Ikehashi H, Ohmido N, Ohtsubo H, Ohtsubo E, Fukui K. 1997. Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol Biol*. 35:791–799.
- Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet*. 10:725–732.
- Wang X, Shi X, Hao B, Ge S, Luo J. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol*. 165:937–946.
- Wang X, Tang H, Bowers JE, Feltus FA, Paterson AH. 2007. Extensive concerted evolution of rice paralogs and the road to regaining independence. *Genetics* 177:1753–1763.
- Wang X, Tang H, Bowers JE, Paterson AH. 2009. Comparative inference of illegitimate recombination between rice and sorghum duplicated genes produced by polyploidization. *Genome Res*. 19:1026–1032.
- Wang X, Tang H, Paterson AH. 2011. Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. *Plant Cell* 23:27–37.
- Wing RA, Ammiraju JS, Luo M, et al. (17 co-authors). 2005. The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol*. 59:53–62.
- Wu J, Mizuno H, Hayashi-Tsugane M, et al. (22 co-authors). 2003. Physical maps and recombination frequency of six rice chromosomes. *Plant J*. 36:720–730.

- Xu S, Clark T, Zheng H, Vang S, Li R, Wong GK, Wang J, Zheng X. 2008. Gene conversion in the rice genome. *BMC Genomics*. 9:93.
- Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. 15(5):568–573.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yang Z, Gao Q, Sun C, Li W, Gu S, Xu C. 2009. Molecular evolution and functional divergence of HAK potassium transporter gene family in rice (*Oryza sativa* L.). *J Genet Genomics*. 36:161–172.
- Yim WC, Lee BM, Jang CS. 2009. Expression diversity and evolutionary dynamics of rice duplicate genes. *Mol Genet Genomics*. 281:483–493.
- Yu J, Wang J, Lin W, et al. (117 co-authors). 2005. The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol*. 3:e38.
- Zhu Q, Ge S. 2005. Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytol*. 167:249–265.
- Zou X, Zhang F, Zhang J, Zang L, Tang L, Wang J, Sang T, Ge S. 2008. Analysis of 142 genes resolves the rapid diversification of the rice genus. *Genome Biol*. 9:R49.