Supporting Information

Lu *et al.* 10.1073/pnas.0812798106

SI Text

BAC Clone Identification and Sequencing. High-density filters of *Oryza* BAC libraries (1) were screened with *MOC1* probes. To maximize overlapping orthologous regions, 191 positive clones from 12 *Oryza* genomes were digested with HindIII, run on an agarose gel, and transferred to nylon filters. Southern blots were performed with 5 additional probes from upstream and downstream regions of *MOC1*. Sixteen BAC clones were selected for sequencing. In each tetraploid species, 2 BAC clones with distinct hybridization patterns were selected for sequencing, each representing one subgenome. Each BAC clone was shotgun sequenced to more than 8× coverage with BigDye Terminator (Applied Biosystems) on an ABI 3730 DNA Analyzer. Trace files were analyzed by Phred and assembled with Phrap and Consed. The *indica* sequence was downloaded from the Beijing Genomics Institute Rice Information System (http://rise.genomics.org.cn/). Primer walking was performed to fill gaps in the *indica* sequence. The *japonica MOC1* region was downloaded from the International Rice Genome Sequencing Project Build 4.0 (2).

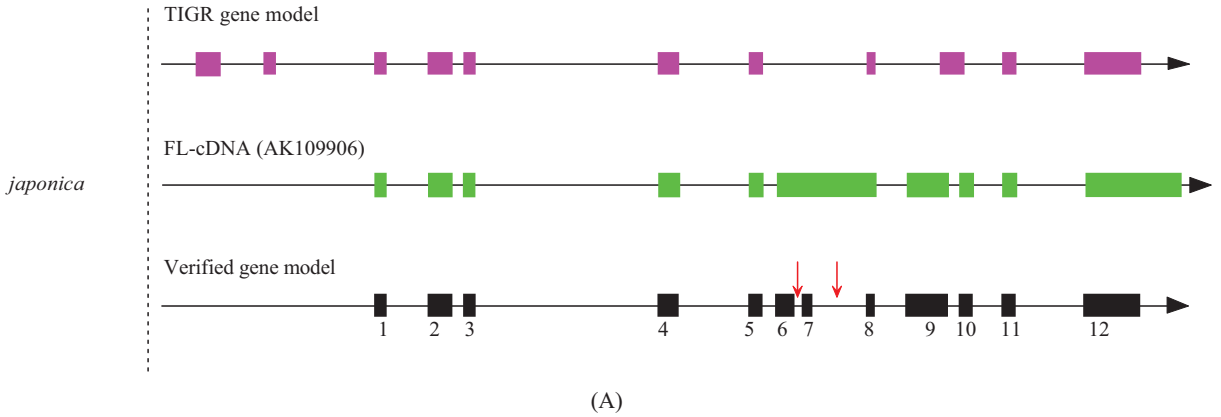
Gene and TE Annotation. Genes were annotated through a combination of ab initio gene prediction, gene conservation in comparative analysis, and experimental verification by cDNA, EST, and RT-PCR analyses. Because *japonica* was the annotation template for other *Oryza* genomes, *japonica* annotation had to be clarified and refined. Current *japonica* gene annotation was downloaded from the Institute for Genomic Research (3) and was compared with the gene annotation of the Rice Annotation Project (4). The exon/intron structure of gene models were verified by aligning the genomic sequence with cDNAs or ESTs.

To annotate genes in non-*japonica* genomes, 3 steps were undertaken. First, 5 gene prediction programs (GENSCAN, FGENESH, GeneMark.hmm, GlimmerM, and RiceHMM) were evaluated for their accuracy in gene prediction. Of these 5 programs, FGENESH was found to be the most accurate on rice genomic sequence and thus was used to predict non-*japonica* gene models. Second, the predicted non-*japonica* gene models were BLASTed against the rice cDNA database (5) and the nonredundant database in the National Center for Biotechnology Information with a cutoff value of 1e-10. Only gene models that were supported by cDNA, EST, or known protein homology in other species, or that were conserved in multiple *Oryza* species, were kept in the non-*japonica* annotations. Predictions without such support or that were TE related were removed. Third, the *japonica* annotation was used to annotate other *Oryza* genomes. The *japonica* gene models were aligned with *Oryza* orthologs with ClustalW. The non-*japonica* gene structures were adjusted on the basis of exon/intron structure in *japonica*. If an ORF existed, the gene structure was retained. If no ORF was found, the gene structure was tested by RT-PCR. If RT-PCR failed to amplify a gene fragment, gene structures were considered ambiguous.

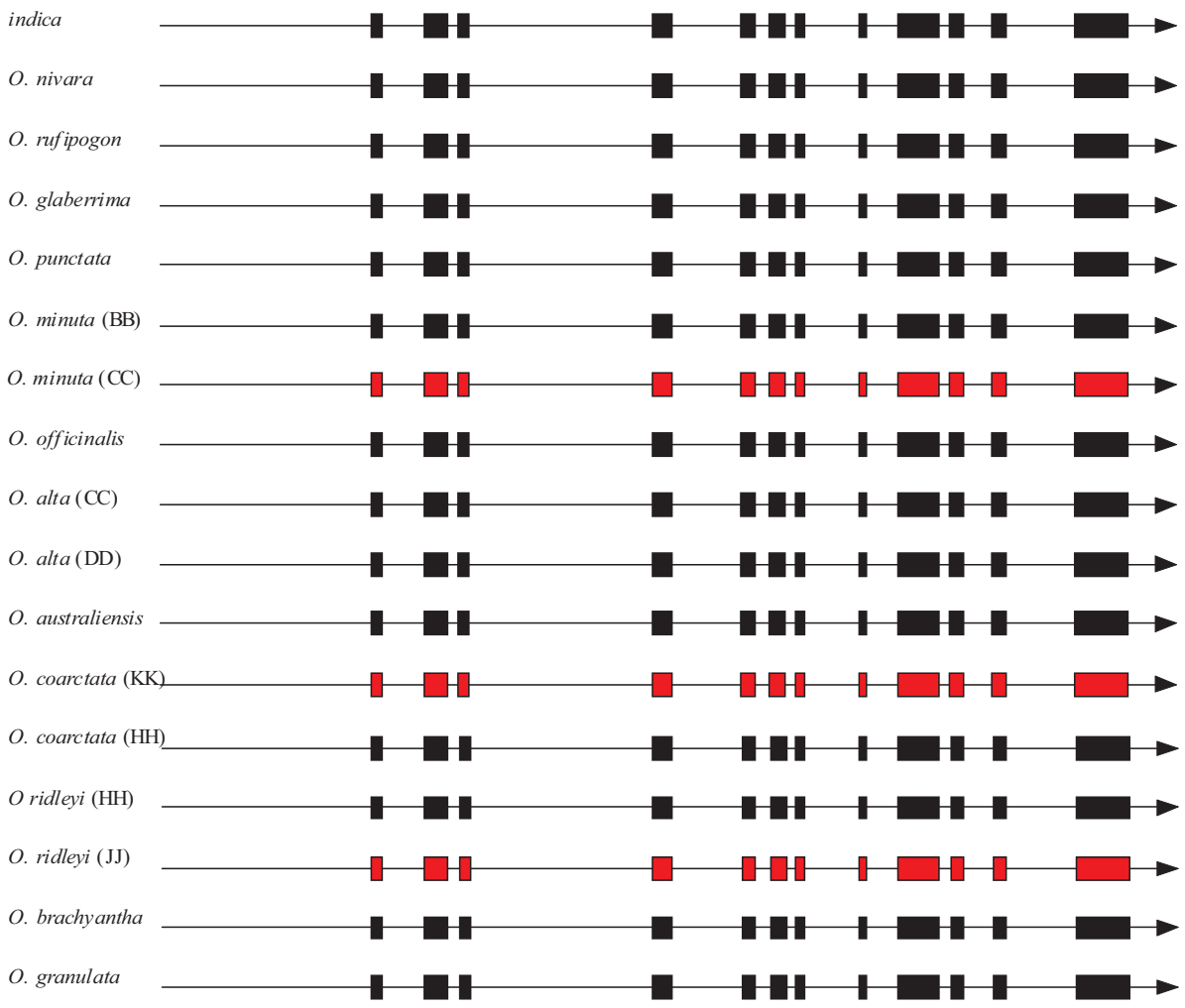
To annotate TEs, RepeatMasker was used initially and followed by manual annotation. Simple repeats or repeats below a cutoff value of 250 were not considered to be TEs. Alignment of orthologous regions between different *Oryza* genomes revealed TE boundaries. The Artemis Comparison Tool (6) was used to graphically recognize TE features, such as direct or inverted terminal repeats. If adjacent TE fragments matched consecutive positions of the same TE in Repbase, they were annotated as a single TE. A TE was "intact" if its length was >90% complete compared with the longest element with this name in the Repbase. Otherwise, the TE was "truncated." Some insertion sequences with clear boundaries and typical TE features not masked by RepeatMasker were also annotated as intact.

1. Ammiraju JS, *et al.* (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res* 16:140–147.
2. IRGSP (2005) The map-based sequence of the rice genome. *Nature* 436:793–800.
3. Zhu W, Buell CR (2007) Improvement of whole-genome annotation of cereals through comparative analyses. *Genome Res* 17:299–310.
4. Tanaka T, Antonio BA, Kikuchi S, Matsumoto T, Nagamura Y, *et al.* (2008) The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res* 36:D1028–1033.
5. Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, *et al.* (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. *Science* 301:376–379.
6. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, *et al.* (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21:3422–3423.
7. Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc Natl Acad Sci USA* 101:12404–12410.

Tecan Ex. 2174, Page 1 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029



(A)



(B)

Fig. S1. Verification of exon/intron structure of gene 23 and its orthologs in *Oryza* genomes. (*A*) Exon/intron structure of gene 23 in *japonica*. The TIGR gene model is partially supported by full-length cDNA (FL-cDNA) data. No read-through ORF can be identified in the FL-cDNA. RT-PCR experiments identified 2 unspliced introns (red arrows) in the FL-cDNA. (*B*) Exon/intron structure of gene 23 orthologs in other *Oryza* genomes. Genes with predicted nonsense or frameshift mutations are shown in red.

Tecan Ex. 2174, Page 2 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

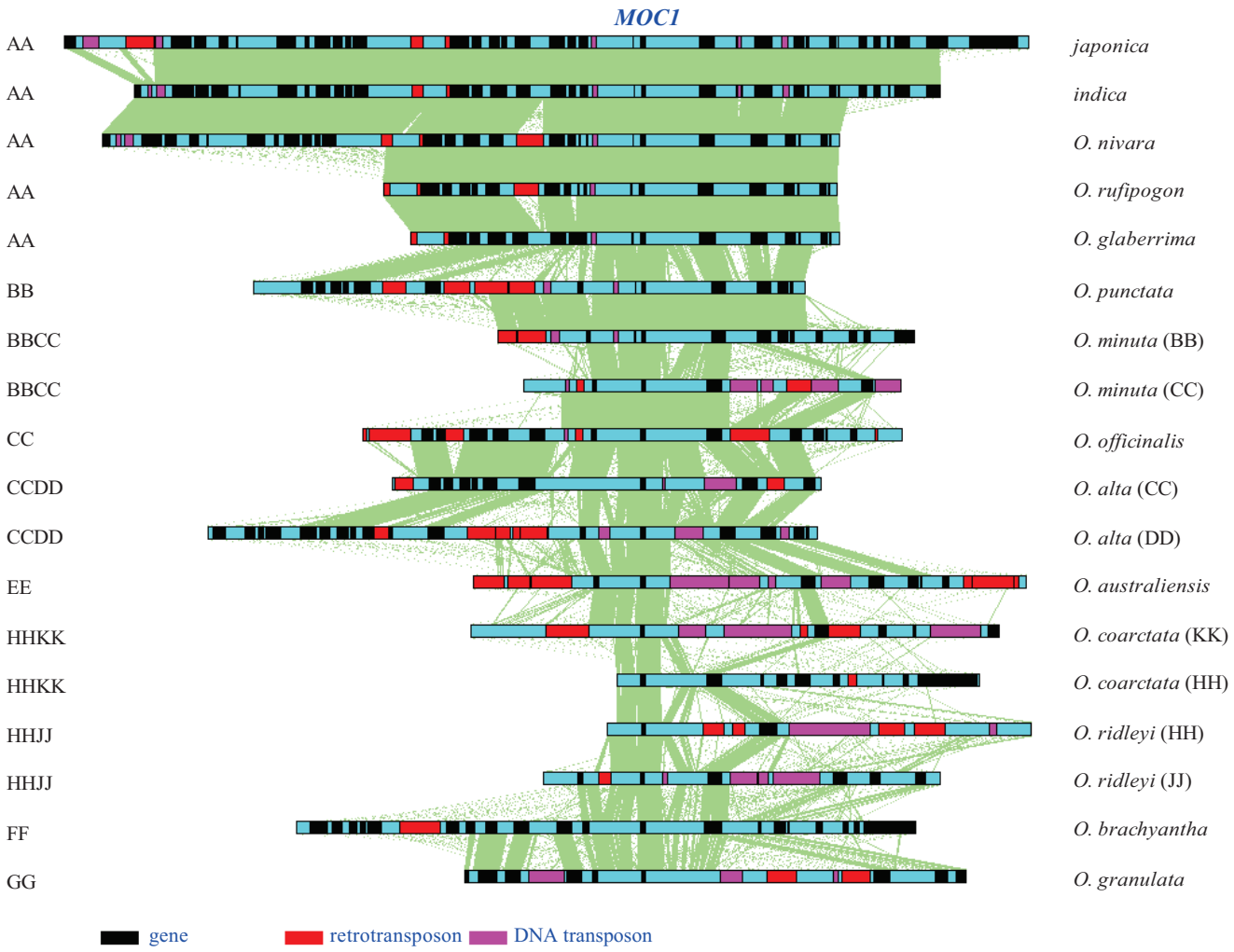
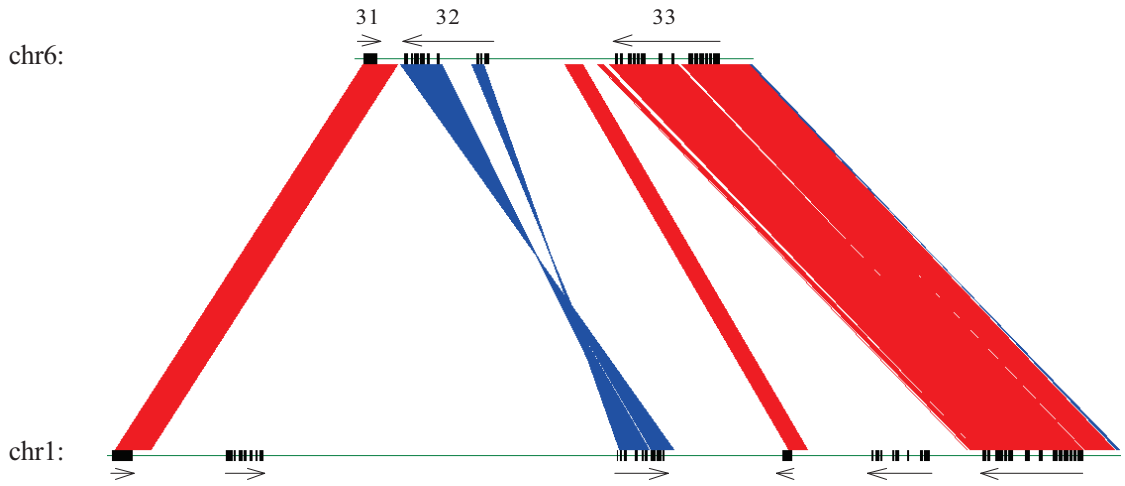
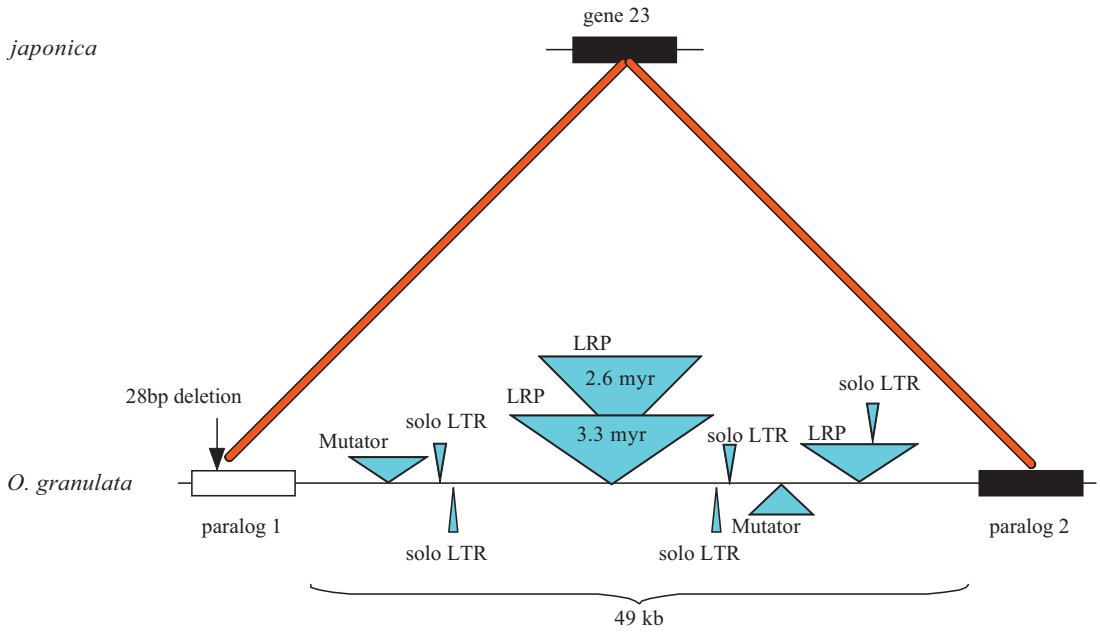


Fig. S2. Sequence homology in *MOC1* regions across the genus *Oryza*. Horizontal bars stand for the genomic sequences of the *MOC1* regions. Homologous sequences are connected in green.

Tecan Ex. 2174, Page 3 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029



(A)



(B)

Fig. S3. Disruption of microcolinearity. (*A*) Genes 31, 32, and 33 were identified in the *MOC1* regions of *japonica* and *indica* but in no other studied *Oryza* genome (Fig. 1). A homology search identified a paralogous region on chromosome 1 of *japonica*. Paralogs for all 3 genes were detected in the paralogous region with a colinear gene order, suggesting that gene movement occurred simultaneously as a single intact segment of DNA. The transcription orientation of gene 32's paralog is reversed relative to that of gene 32. By analyzing the divergence of these 3 genes from those on chromosome 1, this DNA segment was estimated to have inserted into the *MOC1* region of a common *japonica* and *indica* ancestor 1–2 Mya, based on a molecular clock of 6.5×10^{-9} synonymous substitutions per site per year. No signature for Pack-MULE, retrotransposon, or *Helitron*-mediated gene movement was identified in the flanking regions. The mechanism of the genomic fragment movement remains unknown. (*B*) A tandem duplication of gene 23 was detected in *O. granulata*. The 2 paralogs are separated by LTR retrotransposons (LRPs) and *Mutator* elements, some in nested structures. A 28-bp deletion was observed in one paralog; this deletion led to a frameshift mutation. RT-PCR experiments showed the other paralog has the same exon/intron structure as gene 23 in *japonica*. Based on a substitution rate of 1.3×10^{-8} substitutions per site per year (7), both intact LTR retrotransposons were estimated to be have inserted ≈3.3 and ≈2.6 Mya.

Tecan Ex. 2174, Page 4 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

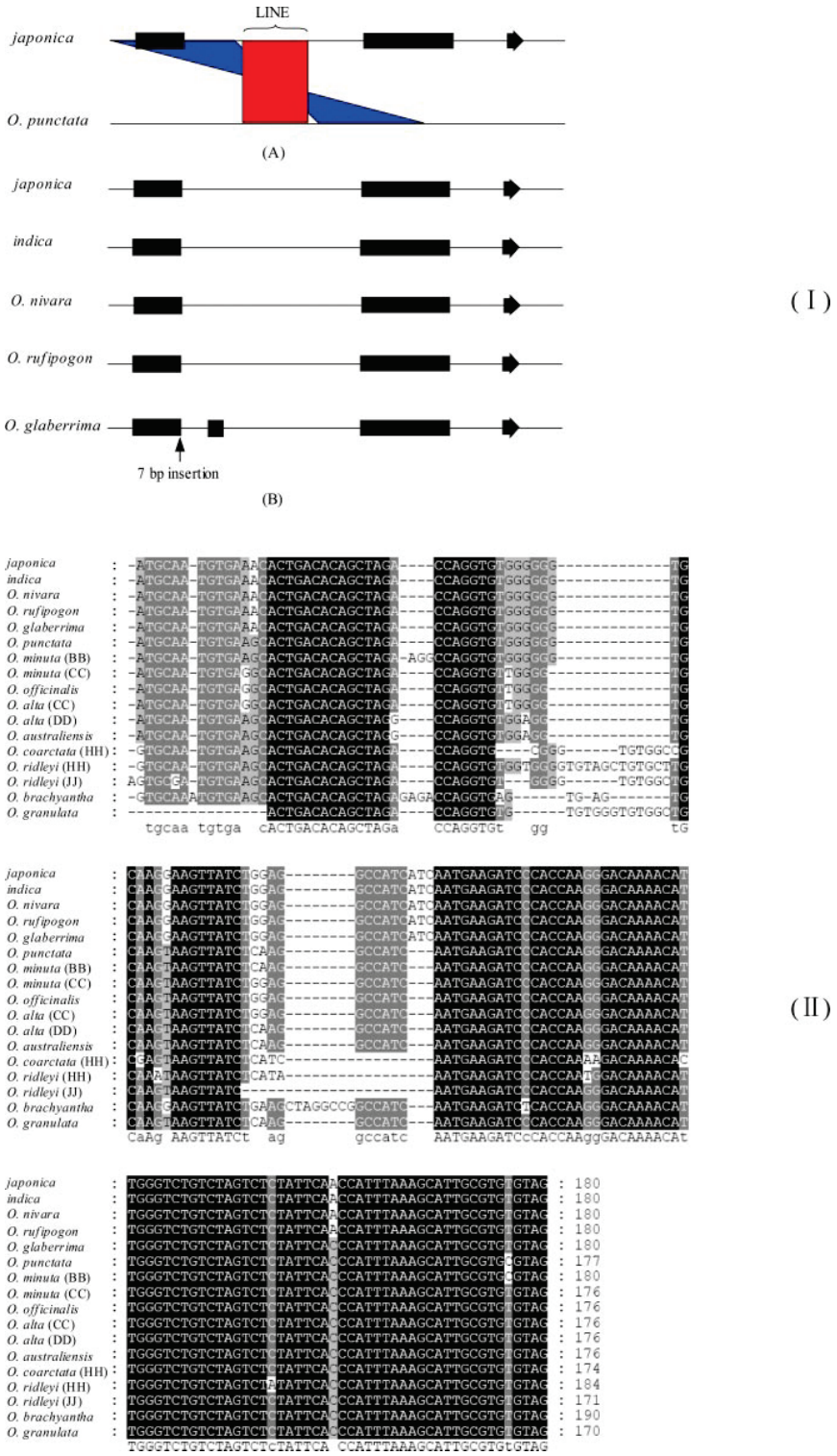


Fig. S4. Candidate de novo genes. (I) Exon/intron structure of gene 18 orthologs. Black rectangles represent exons. (*A*) Sequence homology between *japonica* and *O. punctata*. The red rectangle shows homologous sequence; the blue shape shows inverse homologous sequence. (*B*) Structural variation of gene 18 in AA genomes. A 7-bp insertion was detected in *O. glaberrima*. (II) Sequence alignment of gene 21 orthologous regions in *Oryza* species.

Tecan Ex. 2174, Page 5 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

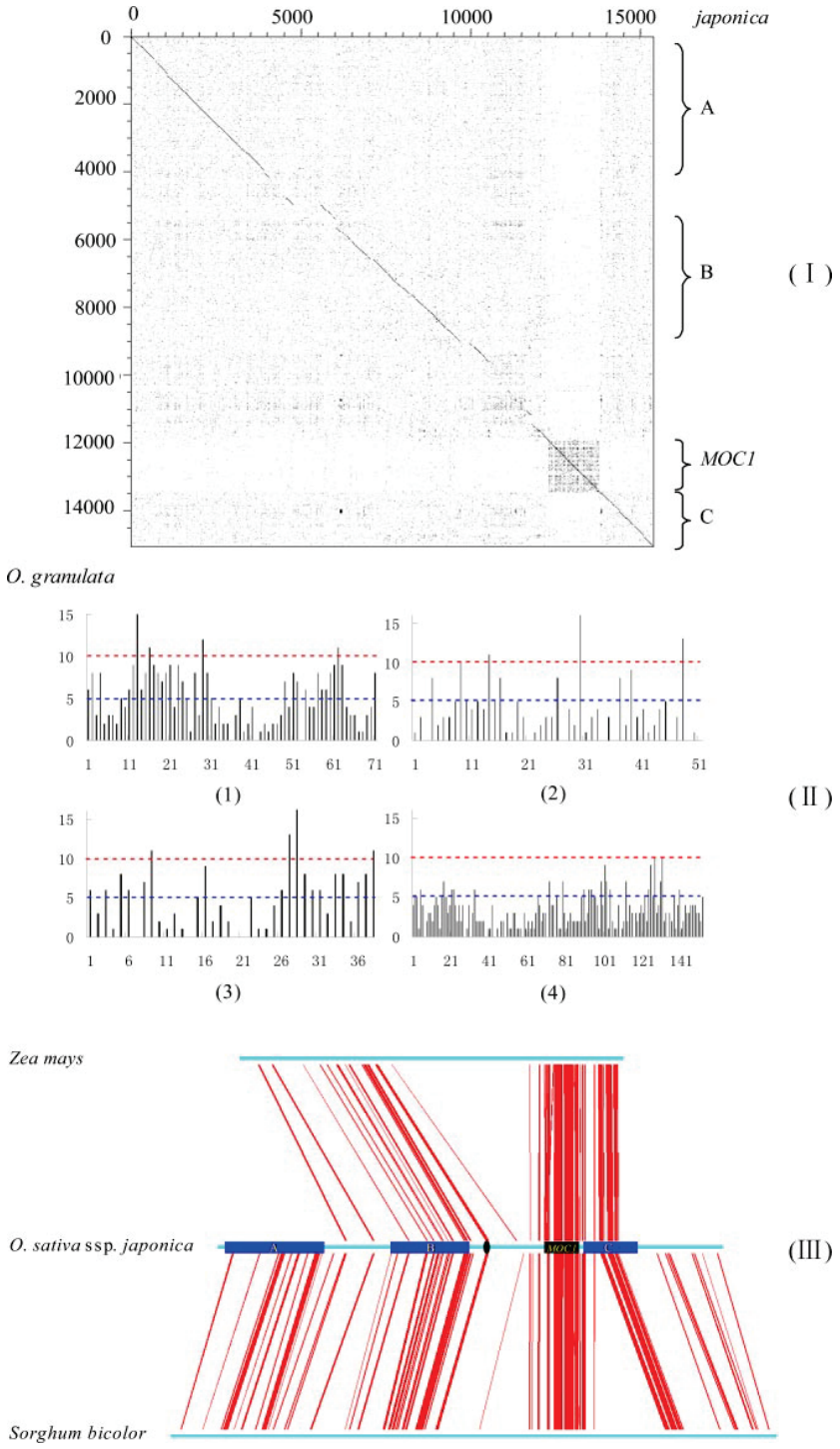
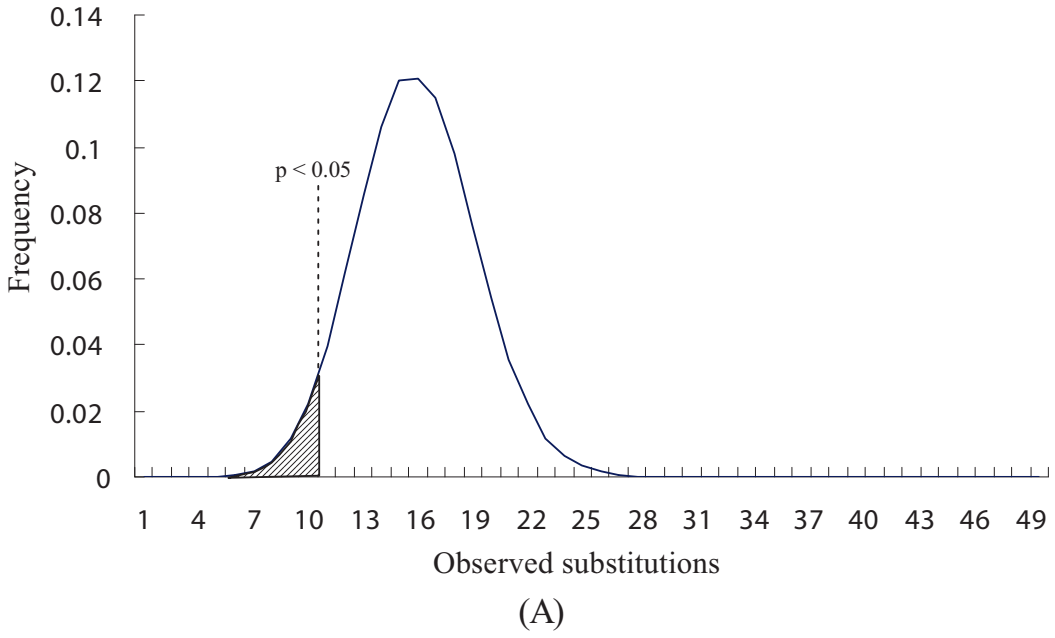


Fig. S5. The CNS regions flanking the *MOC1* gene. (I) Dotplot of the CNS regions between *japonica* and *O. granulata*. The CNS regions A, B, C, and *MOC1* are labeled. The horizontal bar represents the *japonica* sequence in base pairs. The vertical bar stands for the *O. granulata* sequence in base pairs. (II) The CNS regions deviate from a neutral substitution model. The bars illustrate the substitution frequency per 50 bp between *japonica* and *O. granulata*. (1–3) The substitution rate between *japonica* and *O. granulate* in the 3 CNS regions A (1), B (2), and C (3). (4) Genes 14, 15, 16, 17, 22, 23, and 24 were combined to compute the substitutions of coding sequence. The red and blue dashed lines represent significance ($P < 0.05$) for TEs and introns, respectively. The numbers on the *x* axis indicate the sequential 50-bp segments analyzed for each sequence. (III) The CNS regions in rice, maize, and sorghum. The position of gene 21 is indicated as a black oval. The genomic sequences of *Zea mays* and *Sorghum bicolor* orthologous to rice *MOC1* region were downloaded from GenBank (accession no. AC211683) and the Joint Genome Institute (http://genome.jgi-psf.org). Red lines connect identical sequences.

Tecan Ex. 2174, Page 6 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029



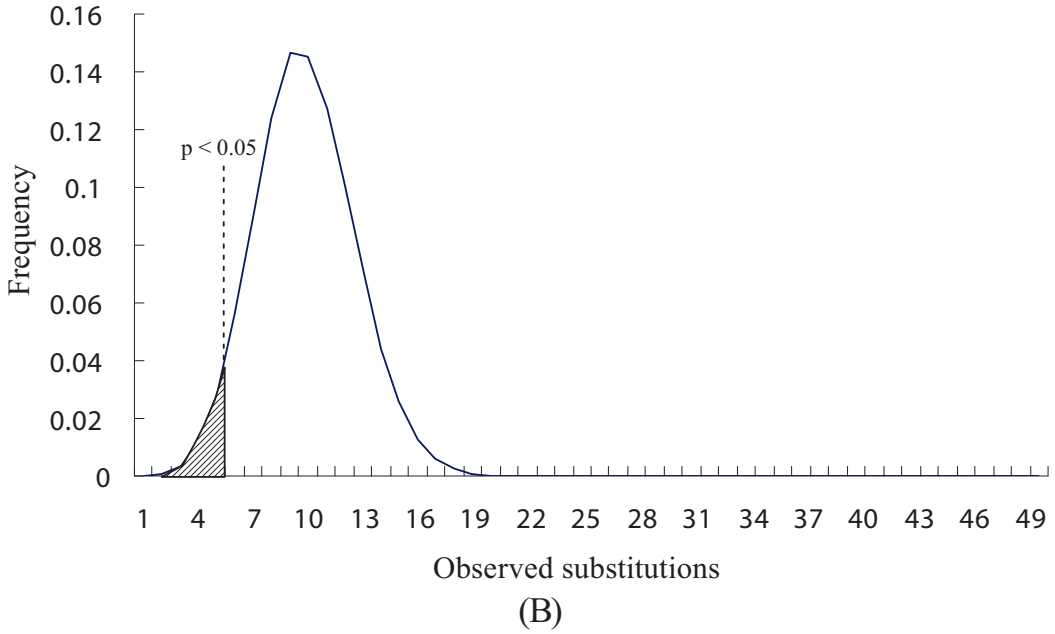


Fig. S6. Theoretical distribution of observed neutral substitutions between *japonica* and *O. granulata*. According to our estimated TE and intron substitution rates, the ratios of different base pairs of TE and intron sequences in a 50-bp window between *japonica* and *O. granulata* are 0.31 and 0.17, respectively. Based on these average values, the different base pairs in the 50-bp window were sampled 20,000 times to produce TE (*A*) and intron (*B*) distribution curves. Sequences with fewer than 10 or 5 base pair differences were considered to have evolved more slowly than TEs and introns, respectively.

Tecan Ex. 2174, Page 7 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Table S1. BAC clones sequenced

Species	Genome type	Sequenced BAC	Sequence length (bp)	GenBank accession no.
O. nivara	AA	a0074E19	218473	FJ032625
O. rufipogon	AA	a0138I22	134695	FJ032626
O. glaberrima	AA	a0079B18	127107	FJ032627
O. punctata	BB	a0082J04	163589	FJ032628
O. minuta	BBCC (BB)	a0185F11	123532	FJ032629
O. minuta	BBCC (CC)	a0318C03	111968	FJ032630
O. officinalis	CC	a0134O03	160114	FJ032631
O. alta	CCDD (CC)	a0212A21	127351	FJ032632
O. alta	CCDD(DD)	a0036L17	180962	FJ032633
O. australiensis	EE	a0096N06	164214	FJ032634
O. coarctata	HHKK (KK)	a0226N01	156698	FJ032635
O. coarctata	HHKK(HH)	a0295K14	107571	FJ032636
O. ridleyi	HHJJ(HH)	a0301G20	125771	FJ032637
O. ridleyi	HHJJ (JJ)	a0085C11	117951	FJ032638
O. brachyantha	FF	a0054N20	183535	FJ032639
O. granulata	GG	a0047L22	148590	FJ032640

Tecan Ex. 2174, Page 8 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Table S2. Gene and repeat annotation of the *MOC1*-orthologous regions

Species		*jap*	*ind*	*niv*	*ruf*	*gla*	*pun*	*min*		*off*	*alt*		*aus*	*coa*		*rid*		*bra*	*gra*
Genome type		AA	AA	AA	AA	AA	BB	BB	CC	CC	CC	DD	EE	KK	HH	HH	JJ	FF	GG
Gene no.		34	33	28	17	17	14	12	5	14	10	16	7	5	10	2	7	21	12
Gene density (kb/gene)		8.4	7.2	7.8	7.9	7.5	11.7	10.3	22.3	11.4	12.7	11.3	23.6	31.3	10.8	62.9	16.9	8.8	11.4
Class I TE																			
gypsy	Intact	1		1	1		4	2	1	1	1		1	1		2	1	1	1
	Truncated						1		1	3	1	2	3	1		2			1
copia	Intact									2	1	3	1			1			
	Truncated	1	1							1	1		3		1				1
Solo LTR	Intact	5	5	4	2	2	6	3	2	4	2	4	1			1	3		3
	Truncated	1	1				3	4	2	3	1	5			3	8	9	1	9
LINE	Intact																		
	Truncated	5	5	5	4	4	2	2				1	1				1	1	1
SINE	Intact	4	5	4	2	2	1	1	1	1									
	Truncated	3	3	3	2	3				1	1	2	1						
Class II TE																			
Mutator	Intact	4	4	4	2	1	1		3		2	2	1	1			1	1	
	Truncated	2	2			1	1	1	1	2		1				1		5	1
hAT	Intact	3	1						2	1	1	2			1		1		1
	Truncated	1	1	1	1	1	1	1	1	2	2	1	1			1	5	3	2
CACTA	Intact	1					1	2	1		1	1	2	2		1	1		
	Truncated	2	1	1			1						1			1			
Tourist	Intact	18	13	12	10	9	4	2	1	2	1	4				3	1	17	2
	Truncated	10	8	6	5	6	3	4	3	6	4	3				1	3	11	8
Stowaway	Intact	25	23	24	12	11	4	7	2	4	4	3	2			3	8	18	6
	Truncated	9	8	8	5	5	3	2		6	4	5	1		1		1	10	4
Other	Intact	15	13	10	6	6	5	4	2	5	6	3		1	1	1	2	11	
	Truncated	15	13	12	10	8	6	6	9	8	8	8	1		4	2	2	11	6
Total		125	107	95	62	59	47	41	32	52	41	50	20	6	11	28	39	90	46

jap: *japonica*; *ind*: *indica*; *niv*: *O. nivara*; *ruf*: *O. rufipogon*; *gla*: *O. glaberrima*; *pun*: *O. punctata*; *min*: *O. minuta*; *off*: *O. officinalis*; *alt*: *O. alta*; *aus*: *O. australiensis*; *coa*: *O. coarctata*; *rid*: *O. ridleyi*; *bra*: *O. brachyantha*; *gra*: *O. granulata.*

Tecan Ex. 2174, Page 9 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Table S3. K_S value and divergence time between different diploid species calculated from genes 22, 23, 24, and 25

Gene	Species	*japonica* (AA)	*O. puncata* (BB)	*O. officinalis* (CC)	*O. australiensis* (EE)	*O. granulata* (GG)
22 (*MOC1*)	*japonica* (AA)		**0.1298**	**0.1427**	**0.1894**	**0.2579**
	O. puncata (BB)	*9.984*		**0.0897**	**0.1538**	**0.2314**
	O. officinalis (CC)	*10.976*	*6.9*		**0.132**	**0.2158**
	O. australiensis (EE)	*14.569*	*11.83*	*10.153*		**0.2405**
	O. granulata (GG)	*19.838*	*17.8*	*16.6*	*18.5*	
23	*japonica* (AA)		**0.0725**	**0.0666**	**0.0969**	**0.1954**
	O. puncata (BB)	*5.576*		**0.065**	**0.0927**	**0.1793**
	O. officinalis (CC)	*5.123*	*5*		**0.0878**	**0.2018**
	O. australiensis (EE)	*7.453*	*7.13*	*6.753*		**0.1964**
	O. granulata (GG)	*15.03*	*13.792*	*15.523*	*15.107*	
24	*japonica* (AA)		**0.0933**	**0.1145**	**0.1434**	**0.1787**
	O. puncata (BB)	*7.176*		**0.0889**	**0.1199**	**0.1715**
	O. officinalis (CC)	*8.807*	*6.838*		**0.098**	**0.1522**
	O. australiensis (EE)	*11.03*	*9.223*	*7.538*		**0.1502**
	O. granulata (GG)	*13.746*	*13.192*	*11.707*	*11.553*	
25	*japonica* (AA)		**0.095**	**0.0756**	**0.0866**	**0.1432**
	O. puncata (BB)	*7.307*		**0.1094**	**0.098**	**0.1507**
	O. officinalis (CC)	*5.815*	*8.415*		**0.0839**	**0.1181**
	O. australiensis (EE)	*6.661*	*7.538*	*6.453*		**0.1397**
	O. granulata (GG)	*11.015*	*11.592*	*9.084*	*10.746*	

For each gene pair, the K_S value (in bold) was translated into divergence time in million years (in italics) based on a substitution rate of 6.5×10^{-9} substitutions per site per year.

Tecan Ex. 2174, Page 10 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

Table S4. Synonymous and nonsynonymous substitutions identified in *Oryza* allotetraploids and their diploid "progenitors"

		12[a]		13		15		16		17		21		22		23		24		25		26	
Group	Genomes	s	n	s	n	s	n	s	n	s	n	s	n	s	n	s	n	s	n	s	n	s	n
BBCC	*O. punctata*											0	0	0	1	0	0	0	0	0	1	0	0
	O. minuta (BB)											0	1	1	6	1	1	2	0	0	2	0	0
	O. minuta (CC)													4	3			0	2				
	O. officinalis													0	0			0	2				
CCDD	*O. officinalis*	9	3	4	1	5	1	4	1	1	4			6	1	1	5	1	3				
	O. alta (CC)	6	8	3	1	1	1	1	1	4	3			3	3	2	2	4	7				

The *s* and *n* represent synonymous and nonsynonymous substitutions, respectively. The base substitutions were deduced by comparing orthologous gene pairs between the subgenome in the allotetraploid and its corresponding diploid progenitor genome, using *japonica* as outgroup. Gene 23 in *O. minuta* (CC) and *O. officinalis* is not listed because gene 23 in *O. minuta* (CC) is a pseudogene. Open reading frames for predicted genes 18–21 are missing from *O. alta*. [a]Numerical designation of the gene

Tecan Ex. 2174, Page 11 of 11
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029