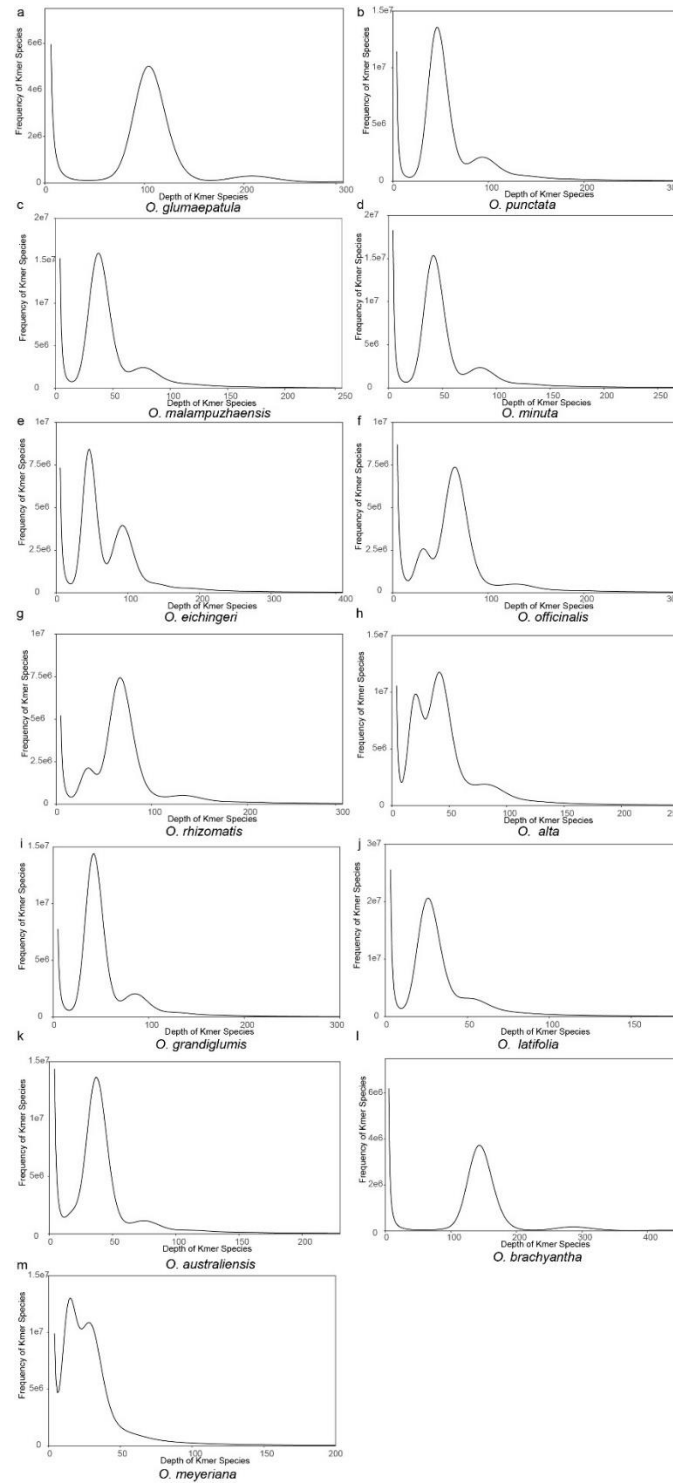
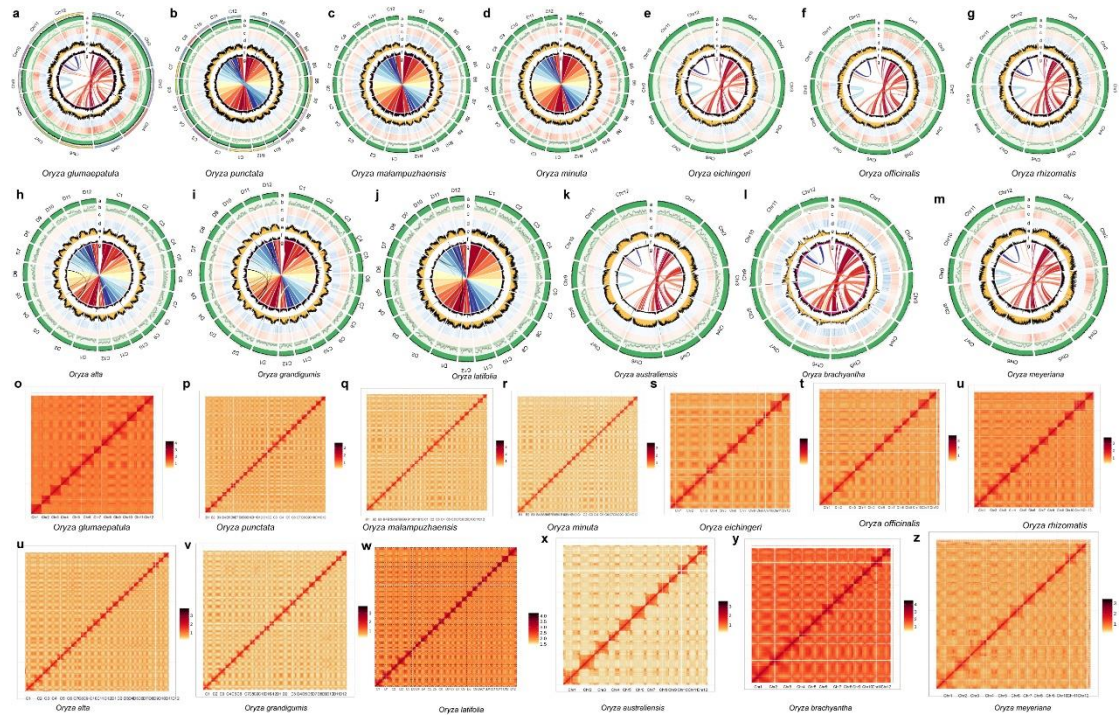


Genome evolution and diversity of wild and cultivated rice species

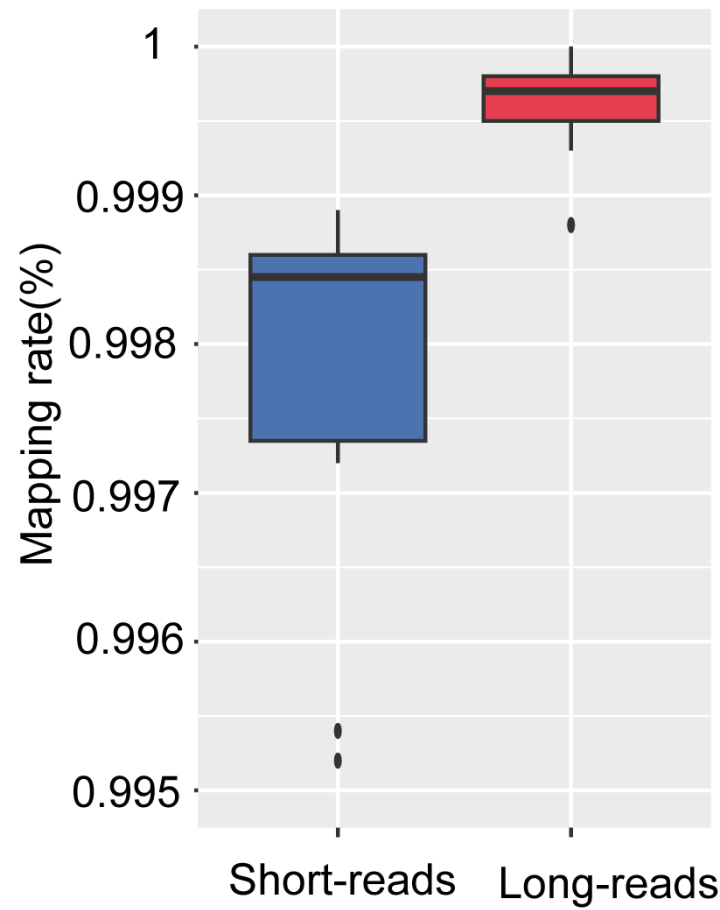
Long *et al.*



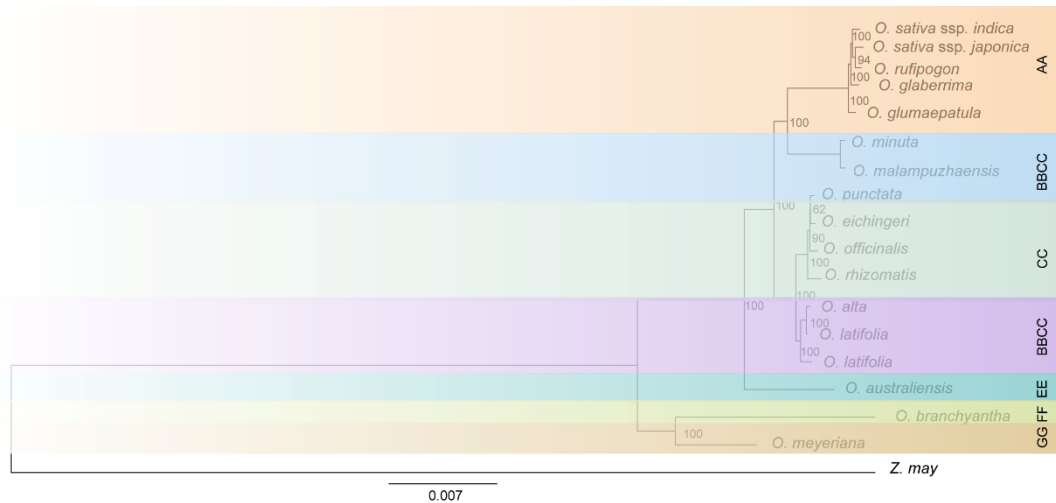
Supplementary Fig. 1. K-mer based genome size estimation. K-mer based estimation of the 13 wild rice species genome size using resequencing data. **a-m** represent the k-mer distribution of *O. glumaepatula*, *O. punctata*, *O. malampuzhaensis*, *O. minuta*, *O. eichingeri*, *O. officinalis*, *O. rhizomatis*, *O. alta*, *O. grandiglumis*, *O. latifolia*, *O. australiensis*, *O. brachyantha*, and *O. meyeriana*. Source data are provided as a Source Data file.



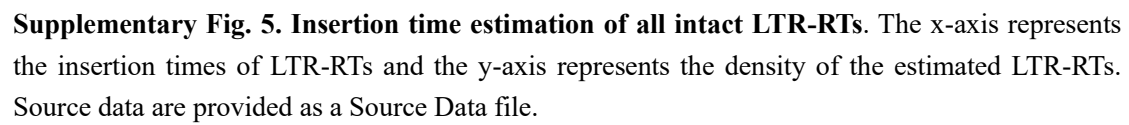
Supplementary Fig. 2. The genome feature, Hi-C map of thirteen chromosome-scale assembled species. (a). Distribution of genomic features of the assemble genomes, (a). Distribution of GC content, (b). Distribution of LAI. (c). Distribution of gene expression, (d). Density of genes. (e). Distribution of transposable elements (TEs) (f). the number of gene, (g). Syntentic blocks within the genome. The different colored lines represent gene collinear connections between each chromosome and itself or other chromosomes (each line in collinearity indicates the syntenic gene). **a-m** represent the genomic features of *O. glumaepatula*, *O. punctata*, *O. malampuzhaensis*, *O. minuta*, *O. eichingeri*, *O. officinalis*, *O. rhizomatis*, *O. alta*, *O. grandiglumis*, *O. latifolia*, *O. australiensis*, *O. brachyantha*, and *O. meyeriana*. **n-z** Chromatin interaction matrices heatmaps of *O. glumaepatula*, *O. punctata*, *O. malampuzhaensis*, *O. minuta*, *O. eichingeri*, *O. officinalis*, *O. rhizomatis*, *O. alta*, *O. grandiglumis*, *O. latifolia*, *O. australiensis*, *O. brachyantha*, and *O. meyeriana*. Source data are provided as a Source Data file.

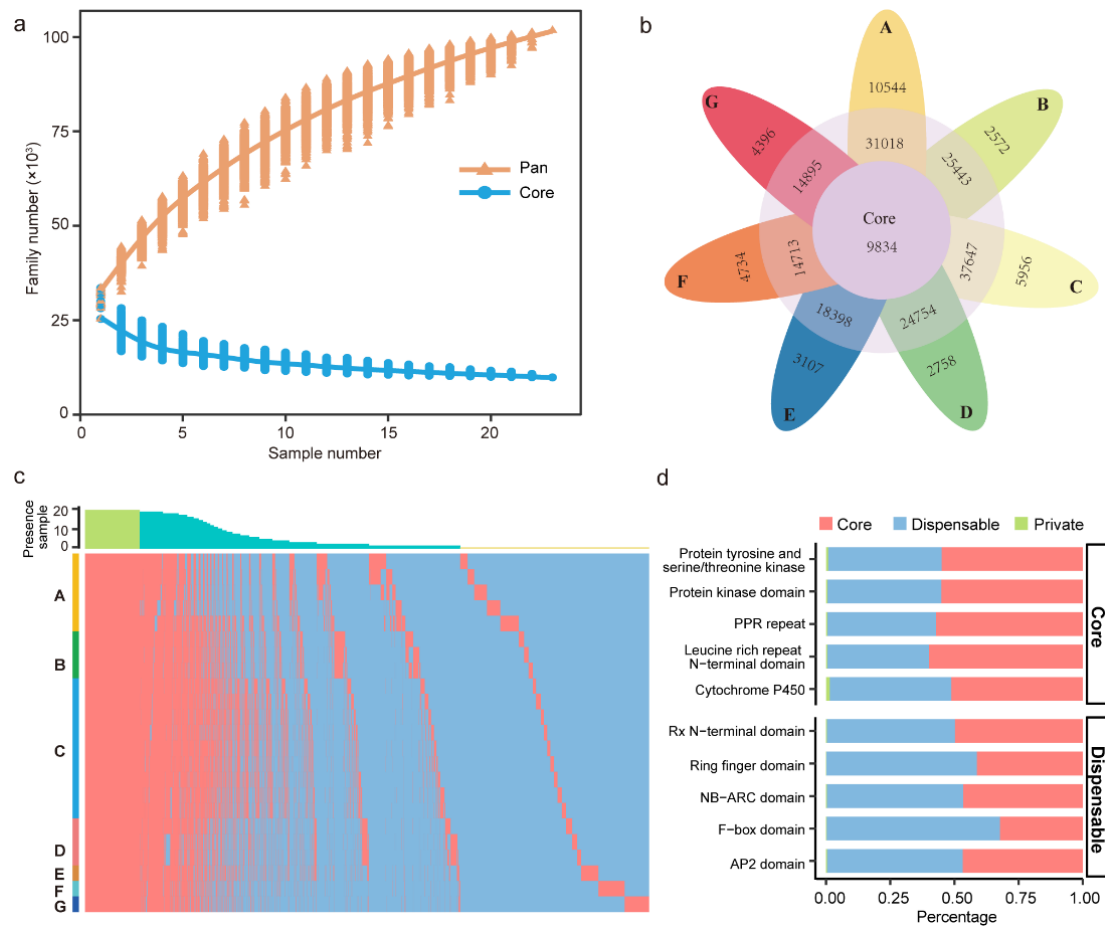


Supplementary Fig. 3. The mapping rate to responding assemble genomes with illumina reads and HIFI reads. Source data are provided as a Source Data file.

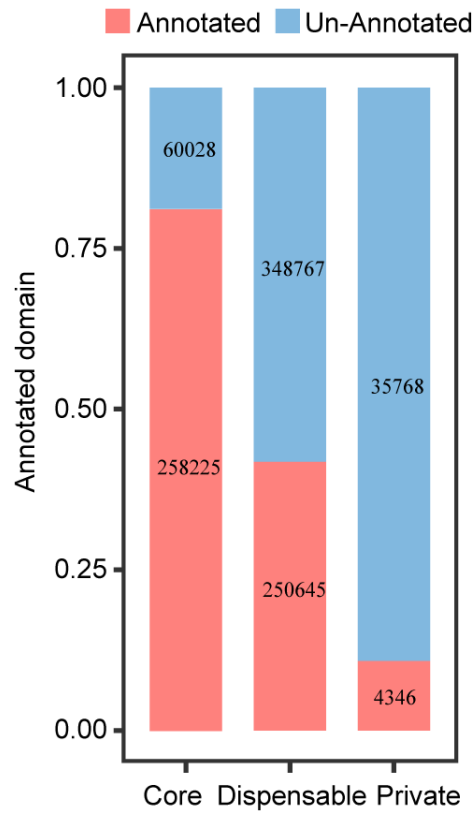


Supplementary Fig. 4. *Oryza* phylogenetic tree. The phylogenetic tree of *Oryza* genotypes was constructed using the chloroplast full sequence obtained from our resequencing data and NCBI. The RAxML program was utilized to generate a Maximum Likelihood (ML) tree with 1000 bootstraps. The numbers beside branches indicate the frequency (%) with which the branch was found in 1,000 bootstraps. Source data are provided as a Source Data file.

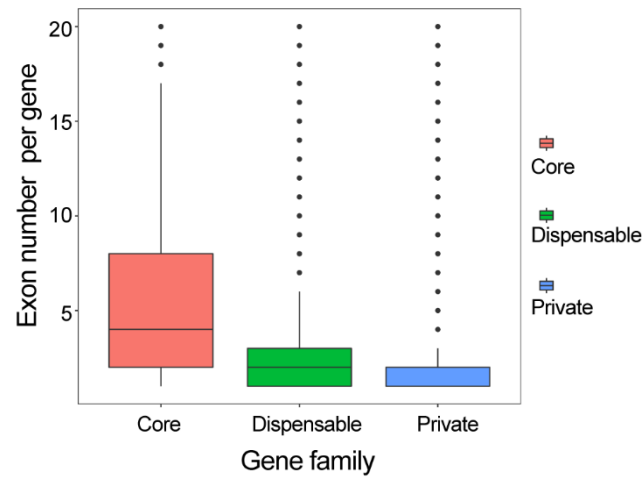




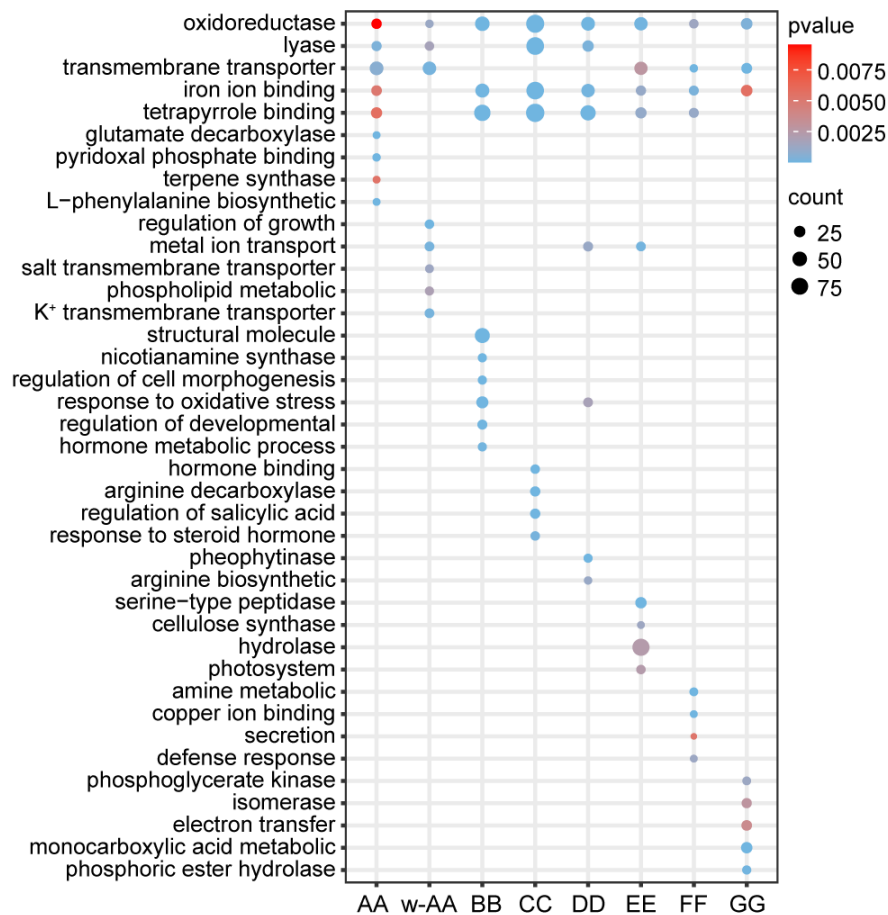
Supplementary Fig. 6. Pan-genome of the all genome type of rice genus. **a** The number of pan-genome and core genome gene families of *Oryza* with the rice diploid genome increases. **b** The gene families of the core, dispensable, and private genomes in terms of different genome types rice. **c** Presence and absence information of pan gene families relative to NIP in the 22 diploid rice genomes. **d** Proportion of the top five genes with interPro domains in the core and dispensable genomes. Source data are provided as a Source Data file.



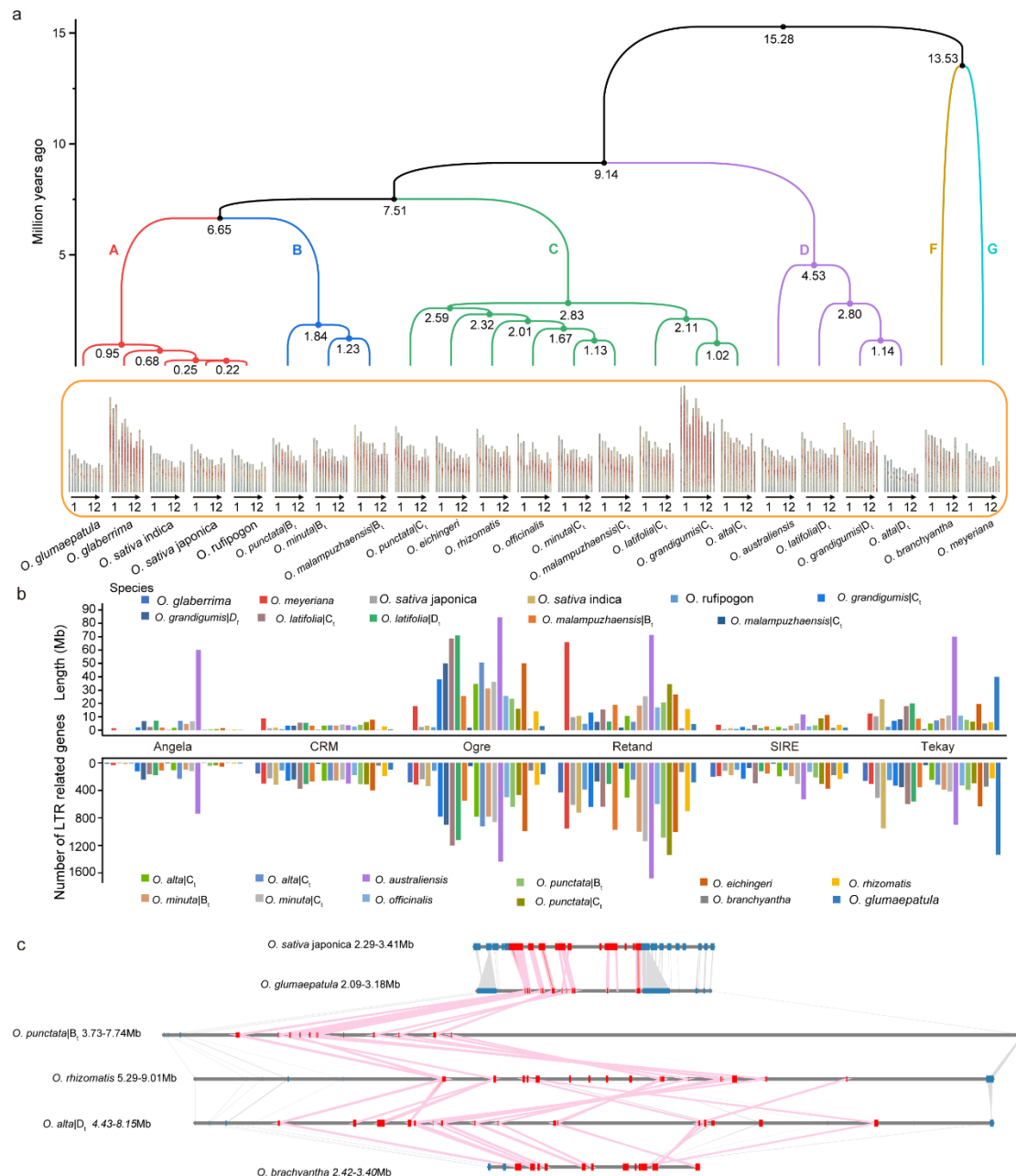
Supplementary Fig. 7. The percentage of genes with annotated InterPro domains in each component of the super pangenome. Red: annotated InterPro domains, blue: un-annotated InterPro domains. Source data are provided as a Source Data file.



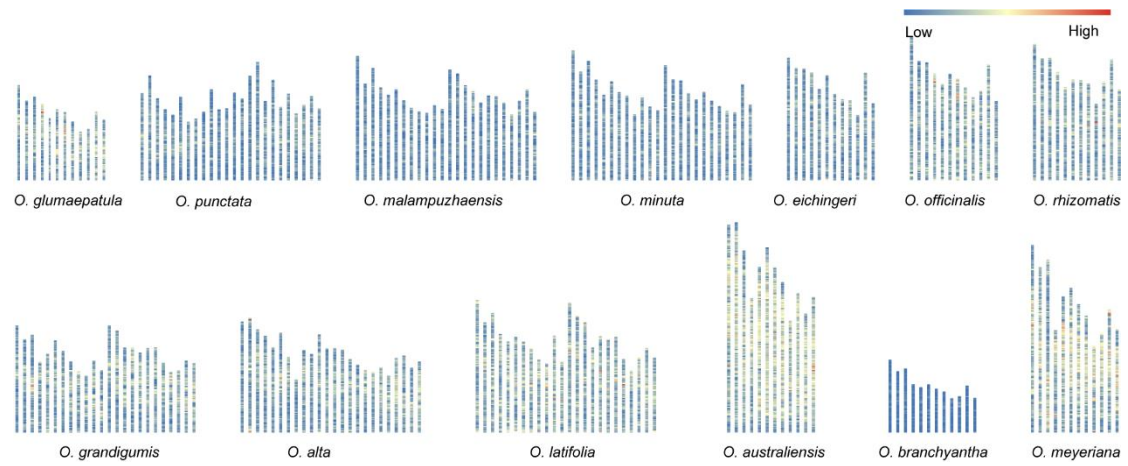
Supplementary Fig. 8. The exon number of per gene in core, dispensable and private genome. Source data are provided as a Source Data file.



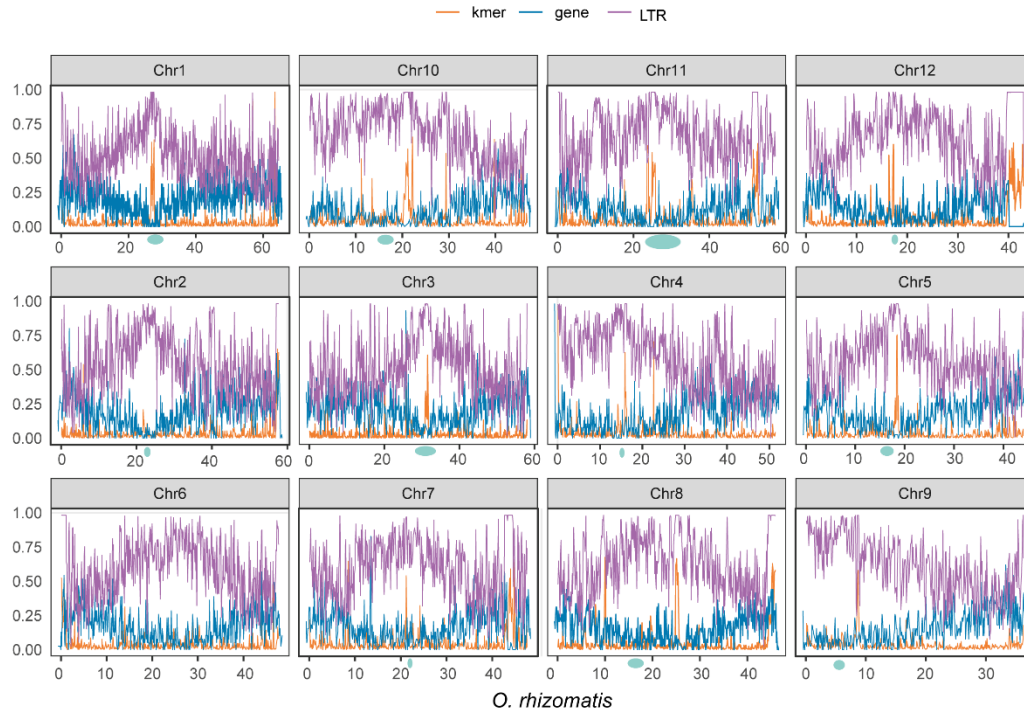
Supplementary Fig. 9. GO enrichment of private gene derived from wild and cultivated rice species with different genome types. w-AA, wild rice with AA genome type, AA, cultivated rice. The circle color shows the statistical significance (P -value), and the size represents the number of hits. Significance was tested by two tailed Fisher's exact method. Source data are provided as a Source Data file.



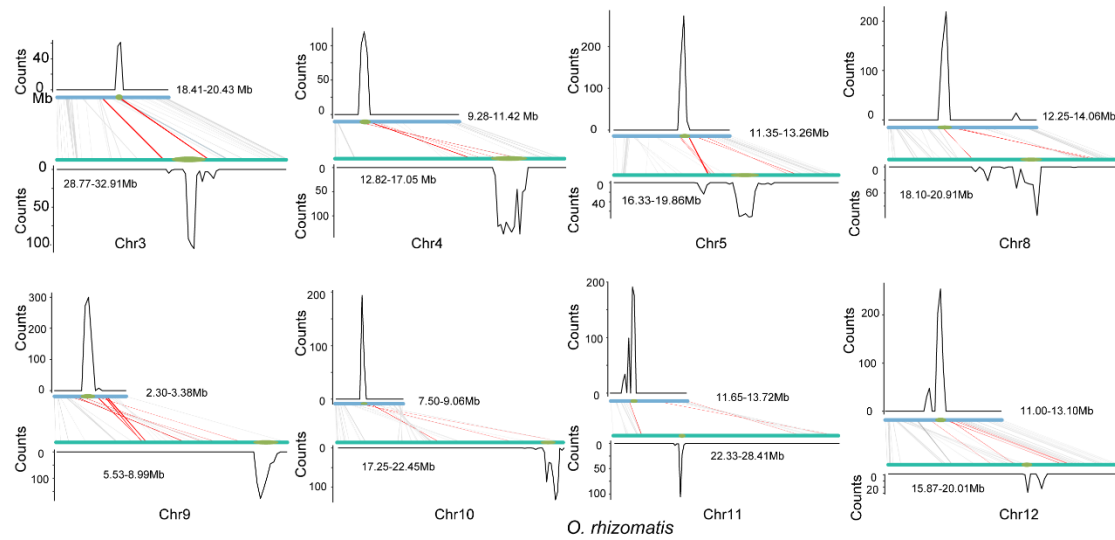
Supplementary Fig. 10. Distribution and impact of transposable elements. **a** The distribution of transposable elements in various rice species within the genus *Oryza* is presented in a panoramic view. **b** The figure illustrates the distribution of each rice species in the top six subfamilies of repeated sequences, along with the proportion of genes they impact across the entire genome. **c** A 1.1-Mb region on chromosome 9 containing functional centromeric domains in the Nipponbare genome is compared with the corresponding homologous region in wild rice species. Syntenic and conserved sequences are visually connected by lines, with genes located in the centromeric region highlighted by red rectangles. Source data are provided as a Source Data file.



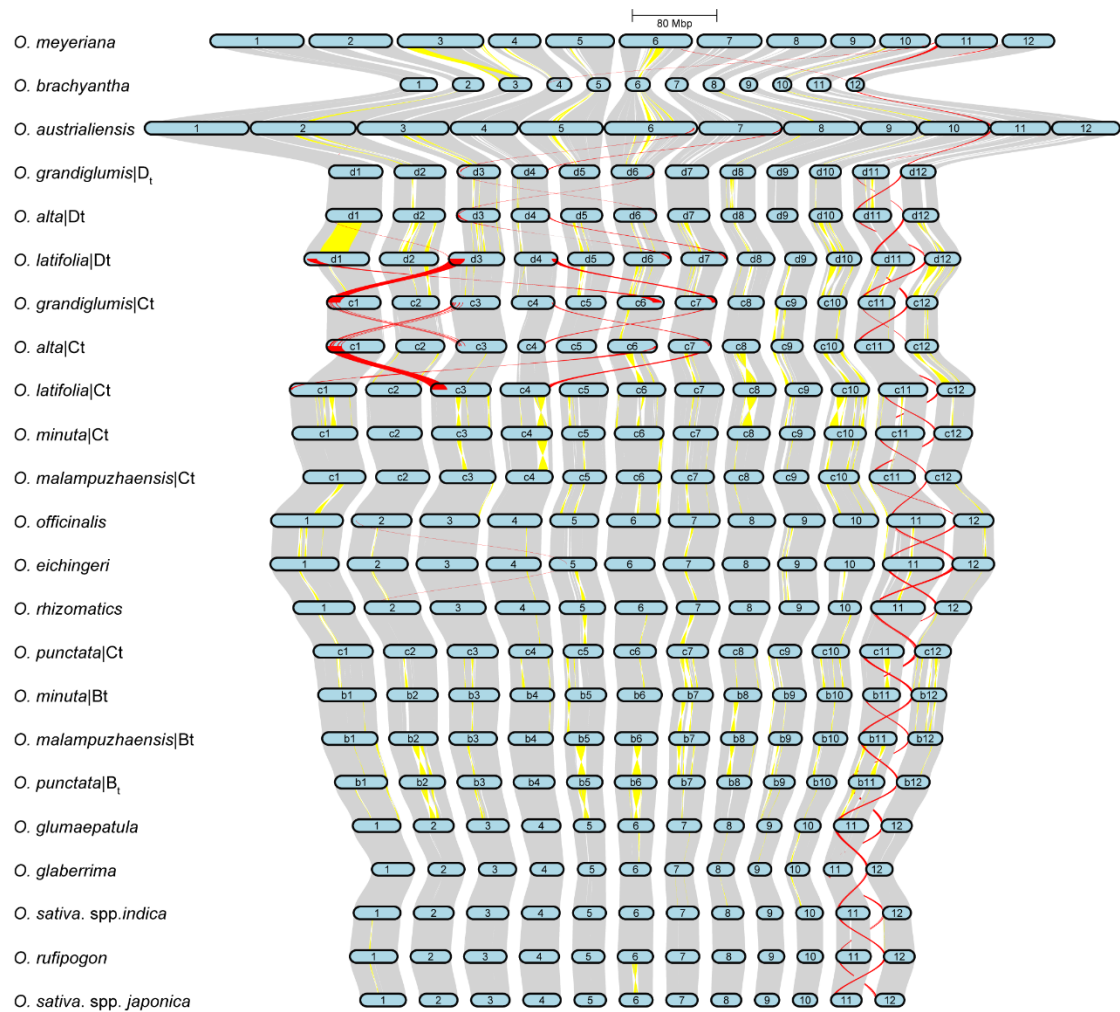
Supplementary Fig. 11. LTR density in wild rice. The density from low to high was showed from blue to red. Source data are provided as a Source Data file.



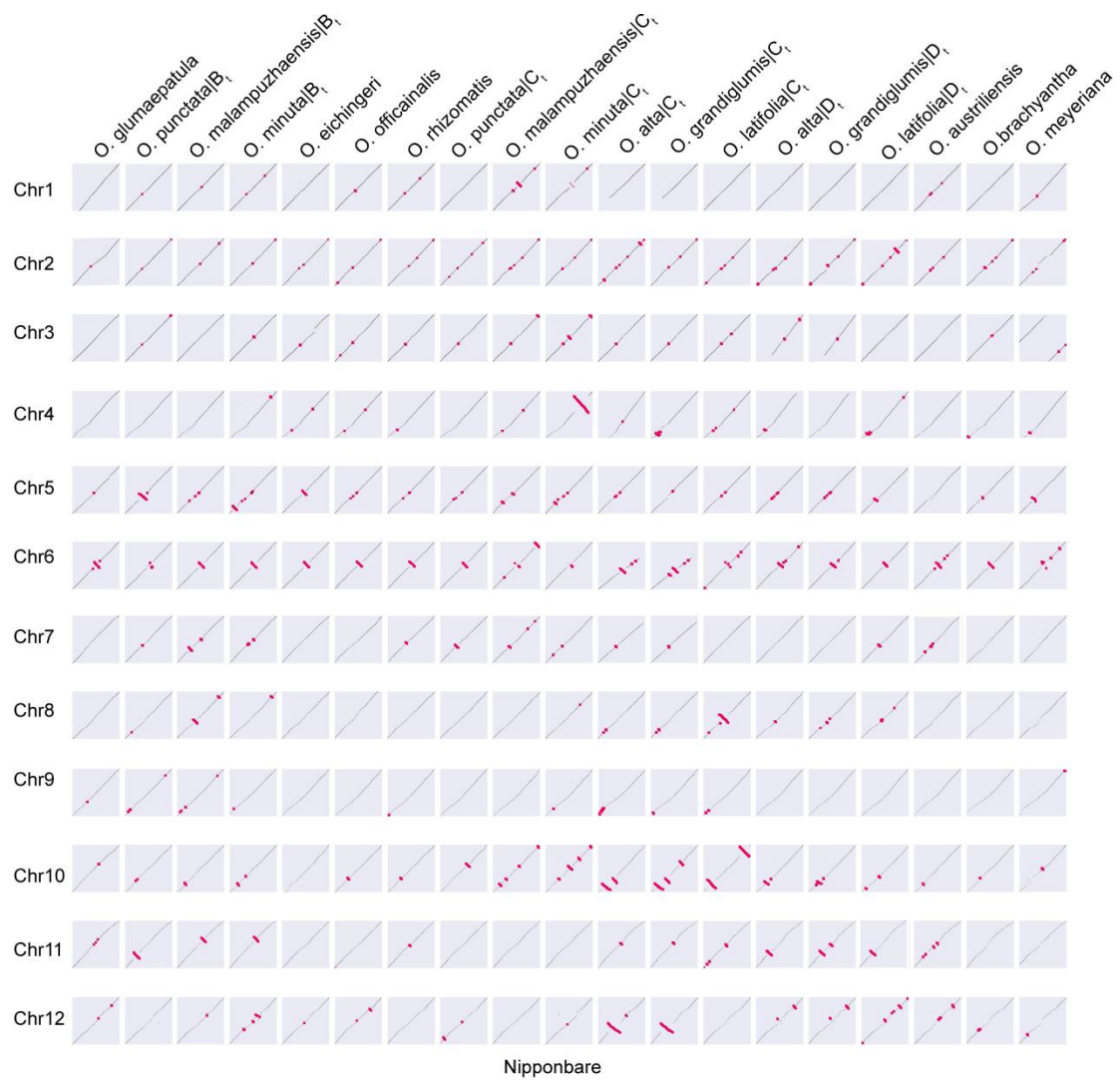
Supplementary Fig. 12. Feature of the centromere in wild rice. 100kb was used as a window length. Yellow: k-mer, green: gene, purple: LTR. Source data are provided as a Source Data file.



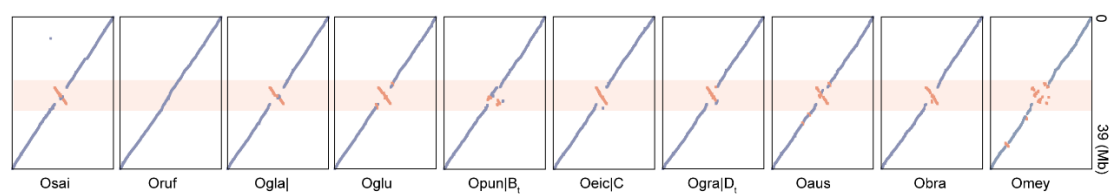
Supplementary Fig 13. The identification and synteny analysis centromeric between cultivated rice and wild rice *O. rhizomatis*. The distribution of centromeric repeats along with the wild rice genome. X-axis represents the chromosome, Y-axis indicates the number of repeats of microsatellite repeat units. Blue shows the centromeric. Red lines indicate the collinear genes on both sides of the centromere between *O. sativa* and *O. rhizomatis*. Source data are provided as a Source Data file.



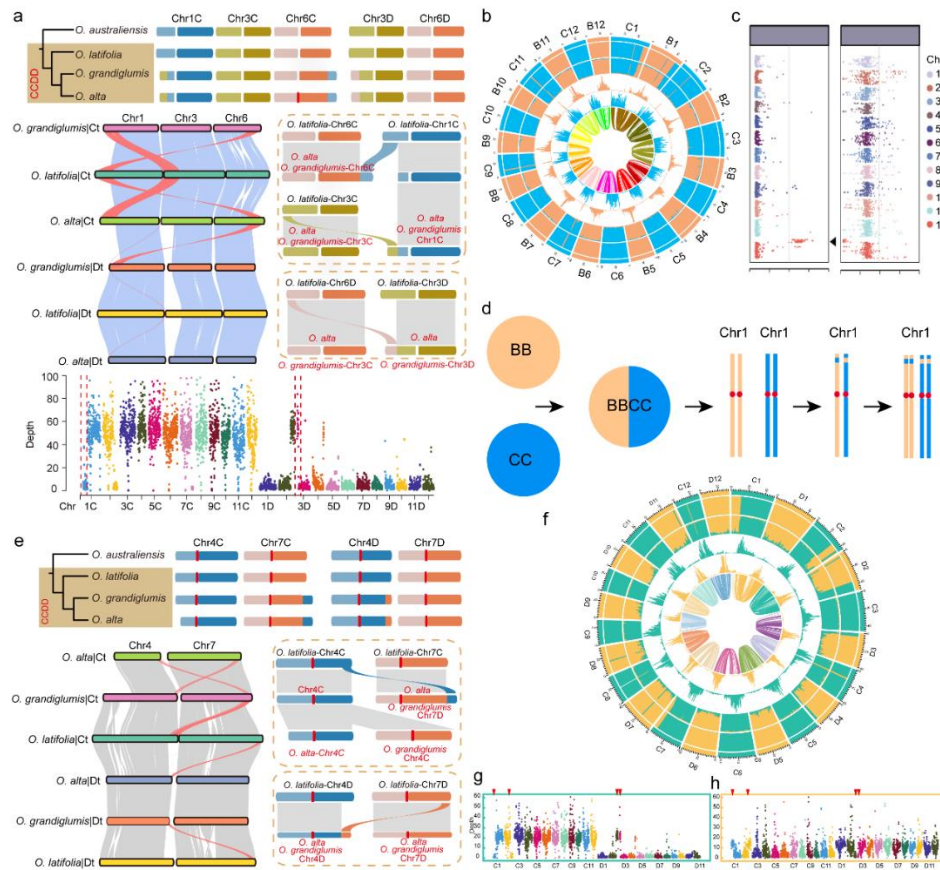
Supplementary Fig. 14. Landscape of the synteny among the rice genus. The genespace syntenic map of orthologous region among the rice genus (sub)genomes. (Sub)genomes were shown from G to A genome types, corresponding to the phylogenetic order within the *Oryza* genus. The length of each chromosome was scaled to represent its appropriate length relative to the other chromosomes. Bar length: 80 Mb. Source data are provided as a Source Data file.



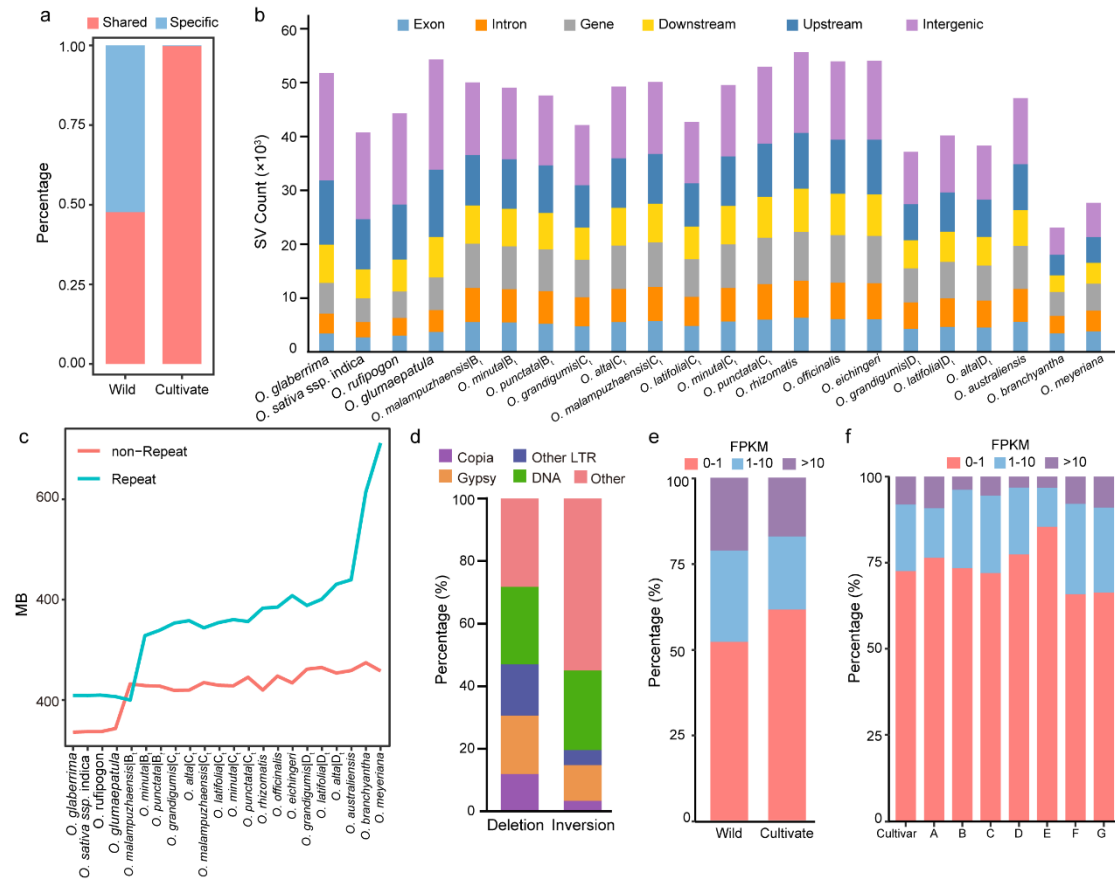
Supplementary Fig. 15. Whole genome alignments of the assemblies of 19 diploid rice genome (13 wild rice species) to the Nipponbare reference genome. Source data are provided as a Source Data file.



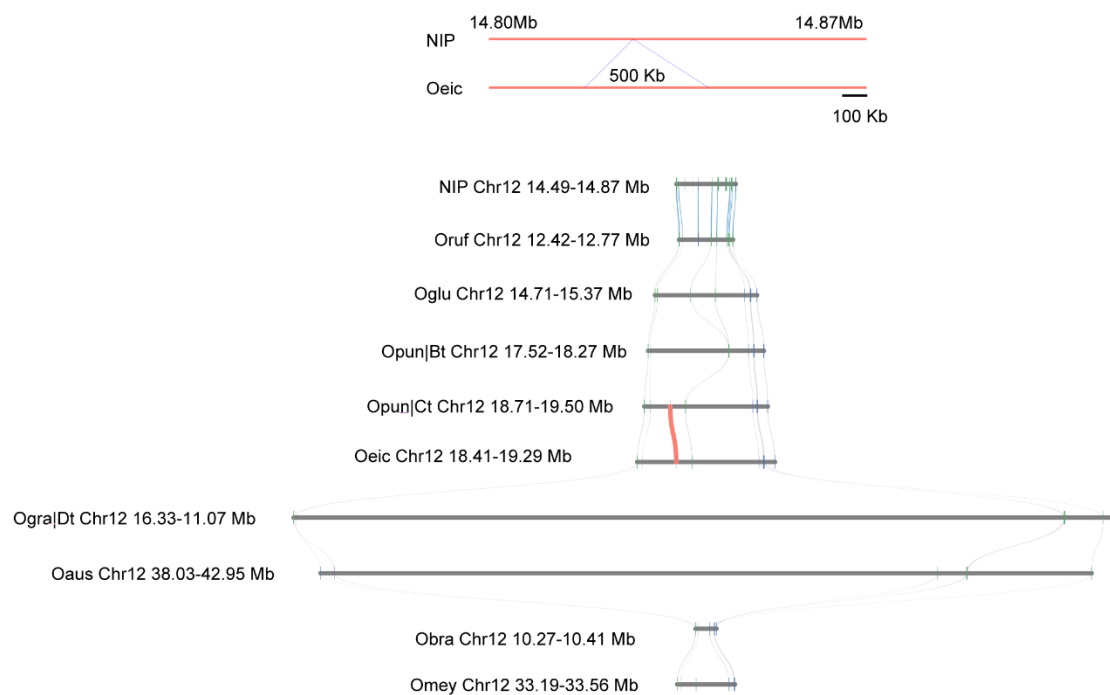
Supplementary Fig. 16. Chromosome inversion among the genus *Oryza*. The orange shade indicates the inversion. Source data are provided as a Source Data file.



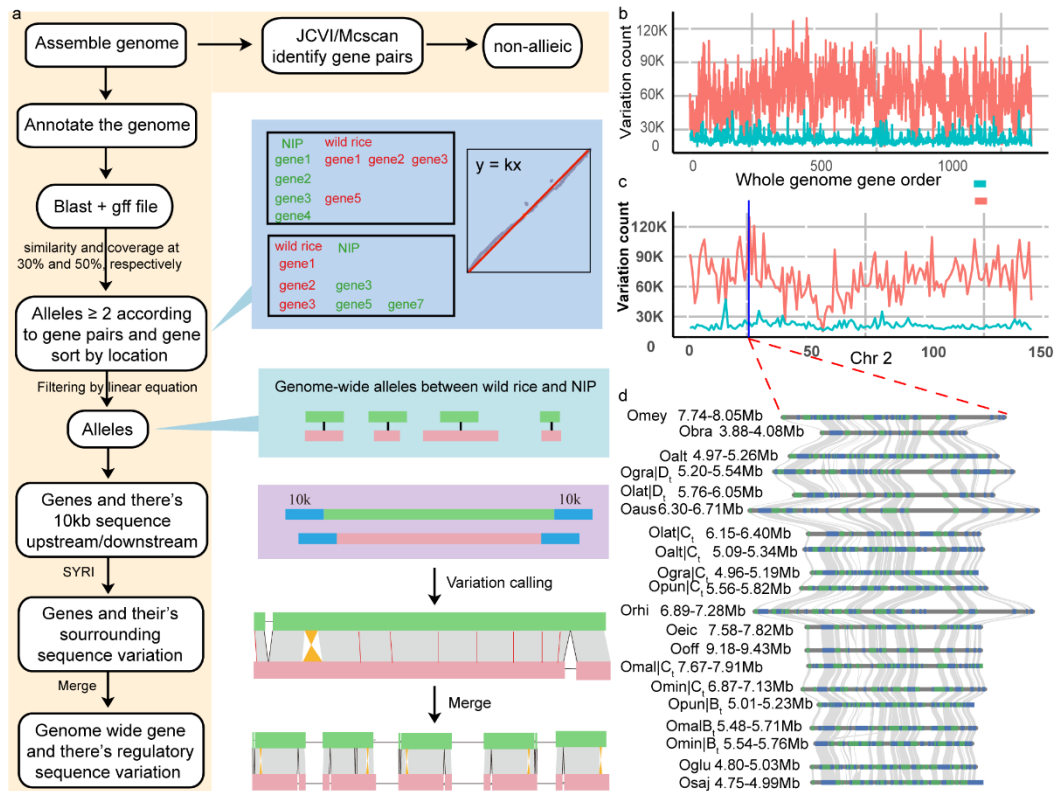
Supplementary Fig. 17. Large scale chromosome rearrangement within wild rice species. **a** Syntenic blocks between the D genome of *O. alta*, *O. latifolia* and *O. grandiglumis* and C genome of *O. alta*, *O. latifolia* and *O. grandiglumis*. Blue lines indicate syntenic regions, red lines indicate the representative chromosome rearrangements between C_t and D_t subgenome. **b** Circos plot of chromosomal characteristics. From outer to inner circles (1–6): (1) Subgenome assignments based on k-means algorithm; (2) Significant enrichment of subgenome-specific k-mers—the same color as the subgenome indicates significant enrichment for those subgenome-specific k-mers; white areas are not significantly enriched; (3) Normalized proportion (relative) of subgenome-specific k-mers; (4–5) count (absolute) of each subgenome-specific k-mer set; (6) homoeologous blocks. All statistics (2–5) are computed in sliding windows of 1 Mb. **c** The mapping depth of Illumina sequencing reads from *O. punctata* (BB) and *O. officinalis* (CC) to *O. punctata* (BBCC) subgenomes. Arrows indicate the breakpoints. **d** Schematic of HE junctions in B and C chromosomes. **e** Whole genome alignment shows an example of large chromosomal inversion on Chr6 within the *Oryza* genus. A Syntenic blocks between the C genome of *O. latifolia* and D genome of *O. alta* and *O. grandiglumis*. Grey lines indicate syntenic regions, Orange lines indicate the representative chromosome rearrangements of Chr4 and Chr7 between *O. lati* and *O. alta*, *O. grandiglumis*. Proposed evolutionary model of the two translocation events leading to the current structure of *O. latifolia*, *O. alta*, and *O. grandiglumis*. **f** The HE junction in C and D chromosomes in *O. alta*. Circos plot of chromosomal characteristics. Orange color indicates D chromosome, cyan represents C chromosome. **g** The tracks indicated the pattern of reads mapping depth in C and D chromosomes using the resequencing data of *O. officinalis*. **h** The depth of mapping to C and D subgenomes by using *O. australiensis* resequencing reads. Source data are provided as a Source Data file.



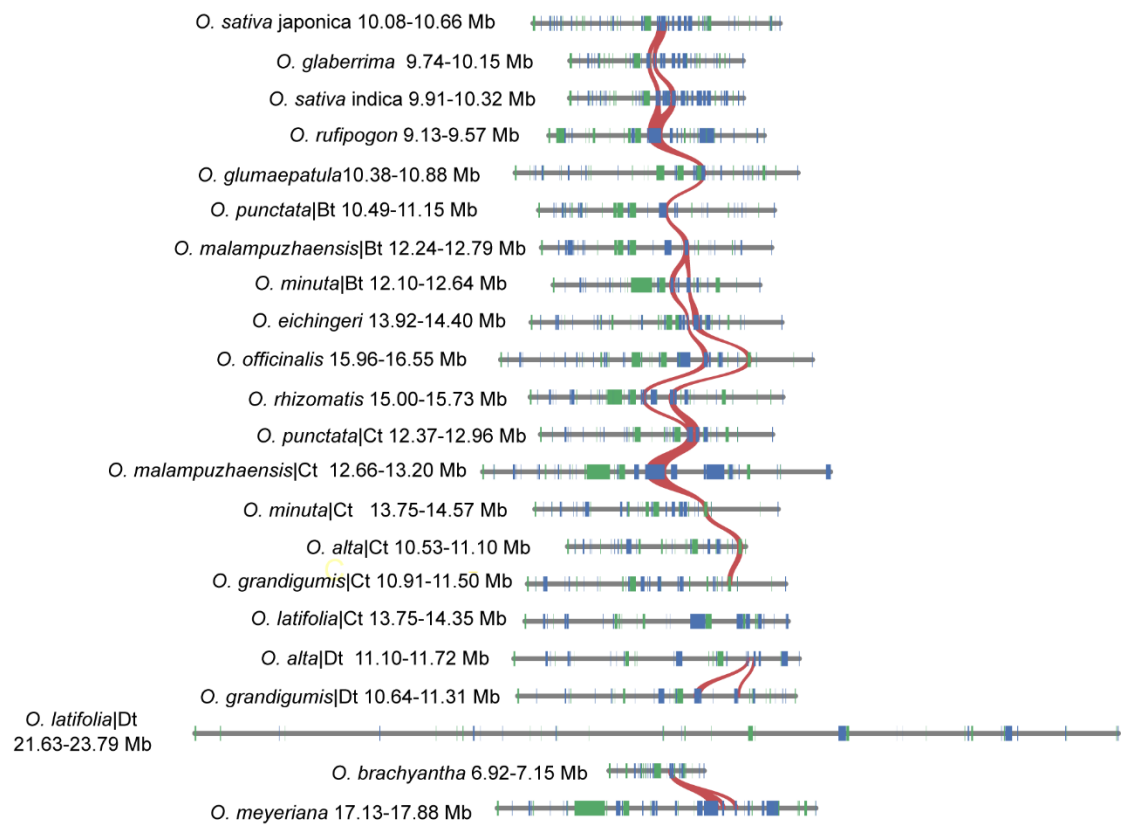
Supplementary Fig. 18. Characteristics of structural variation identified in the rice super pangenome. **a** Proportion of the share SVs and private SVs in wild and cultivated rice. while these variants overlap in the Nipponbare reference genome position, they are regarded as shared. Red color represents shared variations, green indicates the private SVs. **b** The number of SVs sharing overlap with different genomic regions in cultivated and wild rice species. **c** The number of SVs per 1Mb repeat sequences and non-repeat sequences in wild and cultivated rice. **d** Proportion of the TE subfamilies in deletion and inversion variations. **e** Proportion of the number of variations affect gene expression in wild and cultivated rice. **f** Proportion of gene with different expression in wild and cultivated rice with different genome types. **e-f** Red color represents expression value less than 1, green color indicates expression value ranged from 1 to 10, purple color shows the expression value large than 10. Source data are provided as a Source Data file.



Supplementary Fig. 19. Microsynteny at *PstII* loci among *Oryza* species. Red color indicates the collinear gene pairs. Source data are provided as a Source Data file.



Supplementary Fig. 20. Variations in alleles and their regulatory sequences between wild rice and cultivated rice species. **a** Flowchart showing the process of variation between wild rice and cultivated rice alleles and their upstream and downstream 10kb sequences. **b** The number of SVs between wild and cultivated rice on the whole genome. **c** The number of SVs between wild and cultivated rice on chromosome 3. (b-c) 30 genes were used as a window. **d** The syntenic analyses of the largest variation region. Source data are provided as a Source Data file.



Supplementary Fig. 21. Microsynteny at *Pi9* loci among *Oryza* species. Red lines mark the syntenic region. Source data are provided as a Source Data file.

Supplementary Table 1. The basic information of the 13 wild rice species.

#ID	Sequencing number	Accession number	Basic information				
			Species	Genome type	Group	Source/Country	Data type
1	W-18	105133	<i>O. alta</i>	CCDD	Wild	Central and Southern America	HiFi/Hi-C/RNA
2	W-26	86535	<i>O. australiensis</i>	EE	Wild	Australia	HiFi/Hi-C/RNA
3	W-59	104157	<i>O. brachyantha</i>	FF	Wild	Africa	HiFi/Hi-C/RNA
4	W-71	105162	<i>O. eichingeri</i>	CC	Wild	Sourth Asia,East Africa	HiFi/Hi-C/RNA
5	W-97	105666	<i>O. glumaepatula</i>	AA	Wild	Central and Southern America	HiFi/Hi-C/RNA
6	W-100	101405	<i>O. grandiglumis</i>	CCDD	Wild	Central and Southern America	HiFi/Hi-C/RNA
7	W-128	105139	<i>O. latifolia</i>	CCDD	Wild	Central and Southern America	HiFi/Hi-C/RNA
8	W-171	80765	<i>O. malampuzhaensis</i>	BBCC	Wild	India	HiFi/Hi-C/RNA
9	W-190	93256	<i>O. minuta</i>	BBCC	Wild	Philippines, Papua New Guinea	HiFi/Hi-C/RNA
10	W-323	100180	<i>O. officinalis</i>	CC	Wild	Tropical Asia, Tropical Australia	HiFi/Hi-C/RNA
11	W-363	105128	<i>O. punctata</i>	BBCC	Wild	Africa	HiFi/Hi-C/RNA
12	W-375	105450	<i>O. rhizomatis</i>	CC	Wild	Sri Lanka	HiFi/Hi-C/RNA
13	M-6	unknown	<i>O. meyeriana</i>	GG	Wild	China	HiFi/Hi-C/RNA

Supplementary Table 2. Estimated Genome size and Heterozygosity.

Rice species	Genome size(Mb)	Genome heterozygosity(%)
<i>O. alta</i>	961.09	1.04
<i>O. australiensis</i>	967.54	0.31
<i>O. brachyantha</i>	321.76	0.07
<i>O. eichingeri</i>	1139.47	0.11
<i>O. glumaepatula</i>	419.38	0.11
<i>O. grandiglumis</i>	960.69	0.15
<i>O. malampuzhaensis</i>	985.64	0.20
<i>O. minuta</i>	1043.39	0.16
<i>O. officinalis</i>	617.66	0.59
<i>O. punctata</i> (BBCC)	974.93	0.13
<i>O. rhizomatis</i>	629.52	0.51
<i>O. meyeriana</i>	1706.03	0.30
<i>O. latifolia</i>	1191.14	0.19

Supplementary Table 3. Summary of PacBio and HiC sequencing data for 13 accessions.

Number	Species	Abbreviate	Base number	HIFI	Subreads N50(bp)	Depth(X)	HiC	
				subreads N90(bp)			Raw reads	Raw bases
W-18	<i>O. alta</i>	Oalt	36,323,731,817	14255	11674	39.47	195,469,770	58,640,931,000
W-26	<i>O. australiensis</i>	Oaus	33,684,979,685	17233	15178	37.29	133,911,587	40,173,476,100
W-59	<i>O. brachyantha</i>	Obra	19,772,212,199	15540	13400	71.57	138,436,714	41,531,014,200
W-71	<i>O. eichingeri</i>	Oeic	28,906,052,350	15355	13177	76.99	133,933,475	40,180,042,500
W-97	<i>O. glumaepatula</i>	Oglu	8,857,329,157	19802	17631	22.49	145,433,913	43,630,173,900
W-100	<i>O. grandiglumis</i>	Ogra	20,589,374,280	16023	13332	27.69	143,561,168	43,068,350,400
W-128	<i>O. latifolia</i>	Olat	39,820,057,262	16505	14441	37.61	141,403,044	42,420,913,200
W-171	<i>O. malampuzhaensis</i>	Omal	30,489,018,308	15963	13400	31.04	139,281,814	41,784,544,200
W-190	<i>O. minuta</i>	Omin	27,577,693,417	17502	14627	27.09	144,363,193	43,308,957,900
W-323	<i>O. officinalis</i>	Ooff	17,198,616,358	15422	13845	28.39	156,821,813	47,046,543,900
W-363	<i>O. punctata</i>	Opun	28,915,913,659	16132	13823	31.25	162,161,262	48,648,378,600
W-375	<i>O. rhizomatis</i>	Orhi	16,309,508,398	15425	11070	27.09	203,097,632	60,929,289,600
M-6	<i>O. meyeriana</i>	Omey	22,898,547,922	15631	13278	30.09	20,001,828	6,000,548,400

Supplementary Table 4. BUSCO values of the 13 wild rice genome and their diploid genomes.

Accession	Completeness (%)	Single completeness (%)	Duplicated completeness (%)	Fragment (%)	Missing (%)	Total number (%)
<i>O. alta</i>	99.4	12.6	86.8	0.2	0.4	1614
<i>O. australiensis</i>	97.5	96.0	1.5	1.1	1.4	1614
<i>O. brachyantha</i>	97.2	96.0	1.2	1.2	1.6	1614
<i>O. rhizomatis</i>	95.2	93.3	1.9	2.2	2.6	1614
<i>O. glumaepatula</i>	95.0	93.4	1.6	0.9	4.1	1614
<i>O. grandiglumis</i>	99.0	11.5	87.5	0.4	0.6	1614
<i>O. latifolia</i>	99.2	7.5	91.7	0.3	0.5	1614
<i>O. rhizomatis</i>	99.5	13.3	86.2	0.1	0.4	1614
<i>O. minuta</i>	99.0	12.4	86.6	0.3	0.7	1614
<i>O. officinalis</i>	96.6	94.1	2.5	0.9	2.5	1614
<i>O. punctata</i>	99.3	8.6	90.7	0.4	0.3	1614
<i>O. eichingeri</i>	97.1	94.7	2.4	1.0	1.9	1614
<i>O. meyeriana</i>	95.8	92.9	2.9	1.4	2.8	1614
<i>O. malampuzhaensis</i>	99.5	13.3	86.2	0.1	0.4	1614
<i>O. alta</i> Ct	94.3	85.4	8.9	0.6	5.1	1614
<i>O. alta</i> Dt	89.5	86.5	3.0	0.9	9.6	1614
<i>O. grandiglumis</i> Ct	95.1	85.9	9.2	0.7	4.2	1614
<i>O. grandiglumis</i> Dt	89.7	86.7	3.0	0.7	9.6	1614
<i>O. latifolia</i> Ct	98.1	96.1	2.0	0.3	1.6	1614
<i>O. latifolia</i> Dt	97.9	96.1	1.8	0.6	1.5	1614
<i>O. minuta</i> Bt	94.7	93.0	1.7	2.6	2.7	1614
<i>O. minuta</i> Ct	94.1	91.9	2.2	2.8	3.1	1614
<i>O. punctata</i> Bt	96.1	94.9	1.2	1.9	2.0	1614
<i>O. punctata</i> Ct	96.8	94.4	2.4	1.4	1.8	1614
<i>O. malampuzhaensis</i> Bt	95.7	93.9	1.8	2.0	2.3	1614
<i>O. malampuzhaensis</i> Ct	97.4	94.5	2.9	0.9	1.7	1614

Supplementary Table 5. Quality value (QV) and completeness of the 13 wild rice genome assembly.

Species	Completeness	QV
<i>O. meyeriana</i>	75.15	49.49
<i>O. grandigumis</i>	99.09	53.58
<i>O. latifolia</i>	98.68	55.08
<i>O. malampuzhaensis</i>	99.13	54.14
<i>O. alta</i>	81.89	53.78
<i>O. minuta</i>	99.24	54.95
<i>O. australiensis</i>	95.60	57.62
<i>O. officinalis</i>	90.12	50.33
<i>O. punctata</i>	99.59	55.94
<i>O. rhizomatis</i>	91.95	52.11
<i>O. branchyantha</i>	99.52	49.51
<i>O. eichingeri</i>	68.62	54.48
<i>O. glumaepatula</i>	99.24	48.31

Supplementary Table 6. Summary of the *Oryza* genus super pangenome.

Components	Content	Number	Percentage (%)
Core genome	Gene families	9,834	9.66
	Genes	318,253	33.22
Dispensable genome	Gene families	57,822	56.84
	Genes	599,412	62.58
Private genome	Gene families	34,067	33.48
	Genes	40,115	4.18
Total	Gene families	101,723	
	Genes	957,779	

Supplementary Table 7. Summary of the *Oryza* genus super pangenome with different genome types.

Genome Type	Core gene families	Dispensable gene families	Private gene families
A	9,834	31,018	10,544
B	9,834	25,443	2,572
C	9,834	37,647	5,956
D	9,834	24,754	2,758
E	9,834	18,398	3,107
F	9,834	14,713	4,734
G	9,834	14,895	4,396

Supplementary Table 8. The mummer results of wild rice dioploid genome assemblies to NIP reference genome.

Species	Collinear size(bp)	Genome size(bp)	Percentage(%)
<i>O. glaberrima</i>	260,445,301	344,750,569	75.55
<i>O. meyeriana</i>	6,507,903	742,528,311	0.88
<i>O. sativa indica</i>	298,447,364	3903,22,188	76.46
<i>O. grandigumis</i> C _t	45,994,257	454,884,074	10.11
<i>O. grandigumis</i> D _t	34,392,218	450,070,459	7.64
<i>O. latifolia</i> C _t	51,273,306	546,625,894	9.38
<i>O. latifolia</i> D _t	37,487,562	510,736,461	7.34
<i>O. malampuzhaensis</i> B _t	49,870,973	466,791,938	10.68
<i>O. malampuzhaensis</i> C _t	50,207,554	514,842,920	9.75
<i>O. alta</i> C _t	45,869,727	456,354,181	10.05
<i>O. alta</i> D _t	34,667,141	457,173,059	7.58
<i>O. minuta</i> B _t	49,705,034	476,5714,96	10.43
<i>O. minuta</i> C _t	50,197,677	540,844,746	9.28
<i>O. australiensis</i>	33,474,046	902,713,972	3.71
<i>O. officinalis</i>	50,728,336	594,295,013	8.54
<i>O. punctata</i> B _t	50,963,310	435,384,930	11.71
<i>O. punctata</i> C _t	52633381	489,208,212	10.76
<i>O. rhizomatis</i>	50,909,395	601,074,595	8.47
<i>O. branchyantha</i>	3,546,208	274,400,564	1.29
<i>O. eichingeri</i>	51,905,892	544,745,050	9.53
<i>O. glumaepatula</i>	244,215,195	391,712,539	62.35

Supplementary Table 9. Number of collinear genes between wild and cultivated genomes corresponding to the reference genome.

Type	Species/Cultivated name	Number
Wild rice	<i>O. meyeriana</i>	18,463
	<i>O. rufipogon</i>	23,812
	<i>O. grandiglumis</i> Ct	19,018
	<i>O. grandiglumis</i> Dt	18,510
	<i>O. latifolia</i> Ct	20,753
	<i>O. latifolia</i> Dt	20,049
	<i>O. malampuzhaensis</i> Bt	20,781
	<i>O. malampuzhaensis</i> Ct	20,524
	<i>O. alta</i> Ct	18,949
	<i>O. alta</i> Dt	18,644
	<i>O. minuta</i> Bt	20,536
	<i>O. minuta</i> Ct	20,554
	<i>O. officinalis</i>	19,865
	<i>O. australiensis</i>	20,463
	<i>O. punctata</i> Bt	20,838
	<i>O. punctata</i> Ct	20,960
	<i>O. rhizomatis</i>	21,013
	<i>O. brachyantha</i>	18,960
	<i>O. eichingeri</i>	20,448
	<i>O. glumaepatula</i>	22,627
Cultivated rice	2428	27,601
	9311	27,599
	Basmati1	28,042
	CG14	26,671
	CN1	27,738
	D62	26,266
	FH838	26,011
	G8	27,266
	II32	27,586
	IR64	26,517
	J4155	27,328
	Kosh	29,061
	KY131	28,832
	Lemont	27,983
	LJ	28,325
	N22	27,267
	R498	29,623
	Y3551	27,285
	ZH11	27,988

Supplementary Table 10. The number of R gene identified within *Oryza* genus.

Species	R gene number
<i>O. glaberrima</i>	419
<i>O. meyeriana</i>	238
<i>O. sativa japonica</i>	473
<i>O. sativa indica</i>	511
<i>O. rufipogon</i>	260
<i>O. grandigumis</i> C _t	224
<i>O. grandigumis</i> D _t	183
<i>O. latifolia</i> C _t	314
<i>O. latifolia</i> D _t	206
<i>O. malampuzhaensis</i> B _t	302
<i>O. malampuzhaensis</i> C _t	277
<i>O. alta</i> C _t	260
<i>O. alta</i> D _t	216
<i>O. minuta</i> B _t	258
<i>O. minuta</i> C _t	336
<i>O. australiensis</i>	195
<i>O. officinalis</i>	356
<i>O. punctata</i> B _t	320
<i>O. punctata</i> C _t	349
<i>O. rhizomatis</i>	343
<i>O. branchyantha</i>	245
<i>O. eichingeri</i>	288
<i>O. glumaepatula</i>	354