Molecular Plant, Volume *16*

Supplemental information

Rice Gene Index: A comprehensive pan-genome database for comparative and functional genomics of Asian rice

Zhichao Yu (于志超), Yongming Chen (陈永明), Yong Zhou (周勇), Yulu Zhang (张雨露), Mengyuan Li (李梦圆), Yidan Ouyang (欧阳亦聃), Dmytro Chebotarov, Ramil Mauleon, Hu Zhao (赵虎), Weibo Xie (谢为博), Kenneth L. McNally, Rod A. Wing, Weilong Guo (郭伟龙), and Jianwei Zhang (张建伟)

Supplemental Information

Supplemental Information 1 – Functional annotation of Ortholog Gene Index (OGI) genes across species

To gain insights into the evolutionary history of OGI genes, we aligned their protein sequences to the non-redundant protein (NR) database partitioned into 13 taxonomic levels (Wang et al., 2018). We found that new genes explosively emerged with the appearance of *Poaceae* (PS10) and *O. sativa* (PS13) respectively (Supplemental Figures 1B, C, see Methods) and the specific genes are younger and shorter (T-test, p = 2e-16) than core genes (Supplemental Figure 1D). By functional analysis with InterPro domains screened for the gene set, we found that ~88.35% of coding genes contained InterPro domains in the core gene set, a much higher proportion than that in the specific gene set (~35.63%), implying that accession-specific genes may be new genes or pseudogenes. Furthermore, core genes are enriched in essential functions for development, regulation of transcription, DNA-binding transcription factor activity, and so on (using Gene Ontology, GO) (Supplemental Figure 3A), whereas accession-specific genes are enriched in biotic responses like viral penetration into host nucleus (Supplemental Figure 3B).

Supplemental Information 2 – Extra useful functional applications

The “JBrowse” is deployed to present various data tracks like coding and non-coding annotations, transcriptome data, etc. (Figure 1I). “GOEnrichment” (Figure 1H) provides GO enrichment analysis, “GeneDescription” may retrieve gene functions by batches, “OGI” displays the topological structure of each Ortholog Gene Index, and “Download” supplies the genomes and annotations of 16 rice accessions.

Supplemental Information 3 – System construction

Rice Gene Index (RGI) is hosted on a Linux operation system and Nginx web server (https://www.nginx.com) (Figure 1B). All featured data (e.g., homologs, collinearity blocks, transcripts, AS events, and gene functions) of RGI were organized and stored in the MySQL database (http://www.mysql.com). The website was constructed by using Shiny (https://shiny.rstudio.com) and Django

(https://www.djangoproject.com) for the web framework, Bootstrap (https://getbootstrap.com) for the front-end design, and Echarts (https://echarts.apache.org) for data visualization.

Methods

Iso-Seq data processing

Iso-Seq data were sequenced from leaves and roots in 14 accessions, and panicles in 9 accessions. The leaves and roots data were processed to high quality (>98%) reads from PacBio subreads raw data with -ccs_max_length 15000 -ccs_min_length 50 -ccs_polish false -hq_cutoff 0.98 -Maximum_Fuzzy_Junction_Difference 5 -Minimum_Mapped_Concordance 95 -Minimum_Mapped_Coverage 99 -Minimum_Mapped_Length 50 -Require_and_Trim_Poly(A)_Tail true -Run_Clustering true, using SMRTLink 9.0.0.92188 from Pacific Biosciences. The panicles data were sequenced in a mixed pool (9 samples), and the data were processed to circular consensus sequence (CCS) from PacBio subreads raw data with -Maximum_CCS_Read_Length 50000 -Minimum_CCS_Read_Length 10 -Minimum_Number_of_Passes 3 -Minimum_Predicted_Accuracy 0.98, using SMRTLink 8.0.0.80529. Then, we split 9 sample data by barcode and generated high quality (>98%) reads, using the standard IsoSeqv3 (https://github.com/PacificBiosciences/IsoSeq) pipeline. All high-quality (>98%) reads were mapped to the respective genome using minimap2 (v2.17) (Li, 2018). TAMA (Kuo et al., 2020) with -x no_cap -i 0.95 was used to collapse redundant transcripts. Furthermore, TAMA was used to merge the transcripts between tissues and those from CDS prediction.

RNA-Seq data processing

RNA-Seq data were sequenced from multiple tissues (i.e., leaves, roots, and immature panicles) in 16 accessions. The RNA-Seq data was mapped to respective genomes using HISAT2 (v2.1.0) (Kim et al., 2019) with default parameters. The gene expression was calculated from the alignment results using StringTie (v1.3.4d) (Pertea et al., 2015).

Annotation reconstruction

Gene annotations were reconstructed by *de novo* annotation and transcripts from Iso-Seq data. Based on the *de novo* annotation, we identify the new genes and transcripts from Iso-Seq data by comparing them in each variety using SQANTI3 (Tardaguila et al., 2018). Based on the *de novo* annotation gene ID, we inserted the new genes and transcripts to *de novo* annotations by gene locations using python scripts.

Homologs identification

Homologs were identified using reconstructed annotations. Proteins translated by the longest transcript in gene were used to identify homologs. Then, GeneTribe (Chen et al., 2020) was used to perform the homology inference, which combined sequence similarity and collinearity block information, and generated homology relationships including “reciprocal best hits” (RBHs), “single-side best hits” (SBHs), one-to-many, and singletons.

Ortholog Gene Index establishment

First, we removed redundant of the homologous gene groups across 18 annotations to obtain 119,783 non-redundant homologous gene groups. Furthermore, using the above data, homologous gene groups were clustered with the connected graph algorithm to obtain 112,658 Ortholog Gene Indices. Furthermore, using the above data, homologous gene groups were clustered with the connected graph algorithm (McColl et al., 1986) to obtain 112,658 Ortholog Gene Indices. If two gene groups have overlapped genes, they will be clustered to a new gene group by removing the redundant homologous genes. And we iterated the step until there were no connected homologous relationships found between any groups. The indices were then named with the following rules: the first two digits are the chromosome number containing the most genes in each OGI, and the last six digits represent the order of genes on the genome (Using the order of genes in MH63 as a reference to determine the basic order of OGI, and for OGI not containing MH63 gene, query the position of MH63 gene in the OGI where its upstream

gene is located, make insertion, and finally rearrange the order at intervals of 10.). For each OGI, a score was calculated and assigned with the number of involved accessions in each OGI divided by the total number of all accessions, and will be updated as the data increases.

Gene ages identification

Gene ages were identified by previous methods (Wang *et al.*, 2018). We downloaded the NR database (https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/) from NCBI (13 October 2020) and classified the protein sequence to 13 taxonomic levels (PS1 [Cellular organisms]; PS2 [Eukaryota]; PS3 [Viridiplantae]; PS4 [Streptophyta, Streptophytina]; PS5 [Embryophyta]; PS6 [Tracheophyta, Euphyllophyta]; PS7 [Spermatophyta]; PS8 [Magnoliophyta, Mesangiospermae]; PS9 [Liliopsida, Petrosaviidae, Commelinids, Poales]; PS10 [Poaceae]; PS11 [BOP clade]; PS12 [Oryzoideae, Oryzeae, Oryza]; and PS13 [Oryza.sativa]) based on NCBI taxonomy. OGI core gene set and specific gene set were translated into proteins and were aligned to the 13 databases using BLASTP (2.11.0+) (Camacho et al., 2009) with E-value < 1e-5 and identity > 30%. The age of a gene was considered as the taxonomic level of the oldest aligned protein.

Gene function annotation

Proteins from annotation were mapped to the Swiss-Prot database with -evalue 1e-10 using BLAST (2.11.0+) (Camacho *et al.*, 2009). The relationships between genes and Gene Ontology (GO) terms were extracted from the BLAST result by python scripts. We annotated the pan gene sets by InterProScan (v5.55-88.0) (Jones et al., 2014) with -dp -f tsv, and screened by evalue < 1E-10.

Alternative splicing events identification

SUPPA2 (2.3) (Trincado et al., 2018) was used to identify AS events for all transcripts, transcripts from Iso-Seq data and transcripts from various tissues (i.e., leaves, roots, and panicles). The AS events were classified into various types such as intron retention, exon skipping, alternative donor site, alternative accepter site, and alternative position.

Phylogenetic analysis

We first performed the multiple sequence alignment in homolog groups (longest protein of each gene) by mafft (7.475) (Katoh and Standley, 2013) with –auto –treeout –maxiterate 1000 –thread 8 –quiet –inputorder. Second, the result were trimmed by trimAl (1.4.1) (Capella-Gutiérrez et al., 2009) with -automated1. Finally, we used IQ-TREE (1.6.12) (Nguyen et al., 2014) to generate the phylogenetic tree with default parameters and show the tree by ggtree (Yu et al., 2017). By default, the "TreePlot" tool extracts all pairwise homolog scores calculated by GeneTribe (Chen et al., 2020) in RGI, and then uses the MCL algorithm (van Dongen, 2000) to separate genes into different homologous groups. An option is provided to users to determine whether homologs in each group should be analyzed independently, or all genes together.

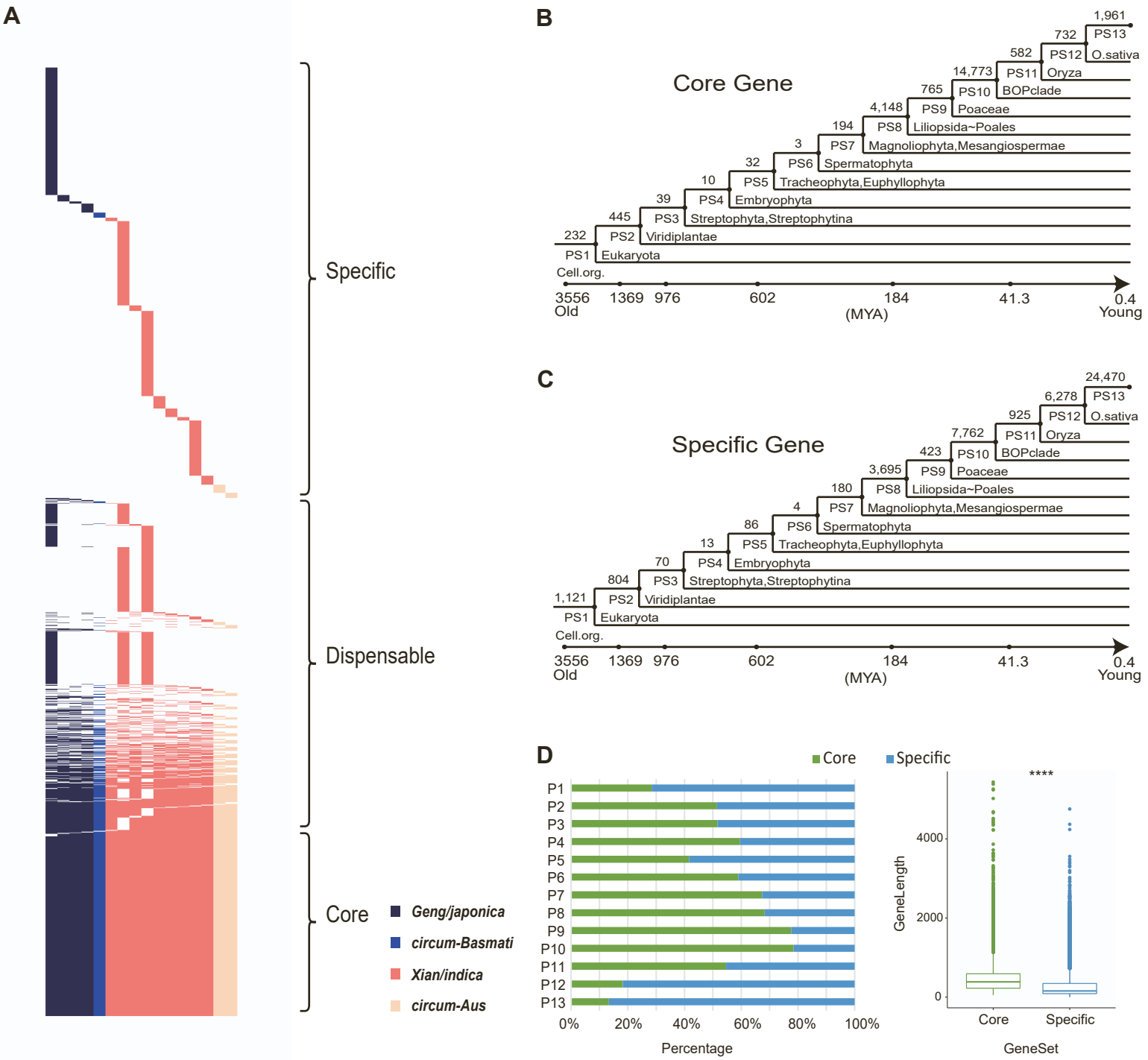
Other integrated Tools

The "BLAST" tool was based on Sequenceserver (Priyam et al., 2019), and the "JBrowse" was adapted from Buels, et al (2016).

Data availability

The datasets generated during and/or analyzed during the current study are available in https://riceome.hzau.edu.cn. PacBio Iso-Seq raw data are available from NCBI under BioProject PRJNA760839. The RNA-Seq raw data are available from NCBI under BioProject PRJNA659864 and PRJNA597070.

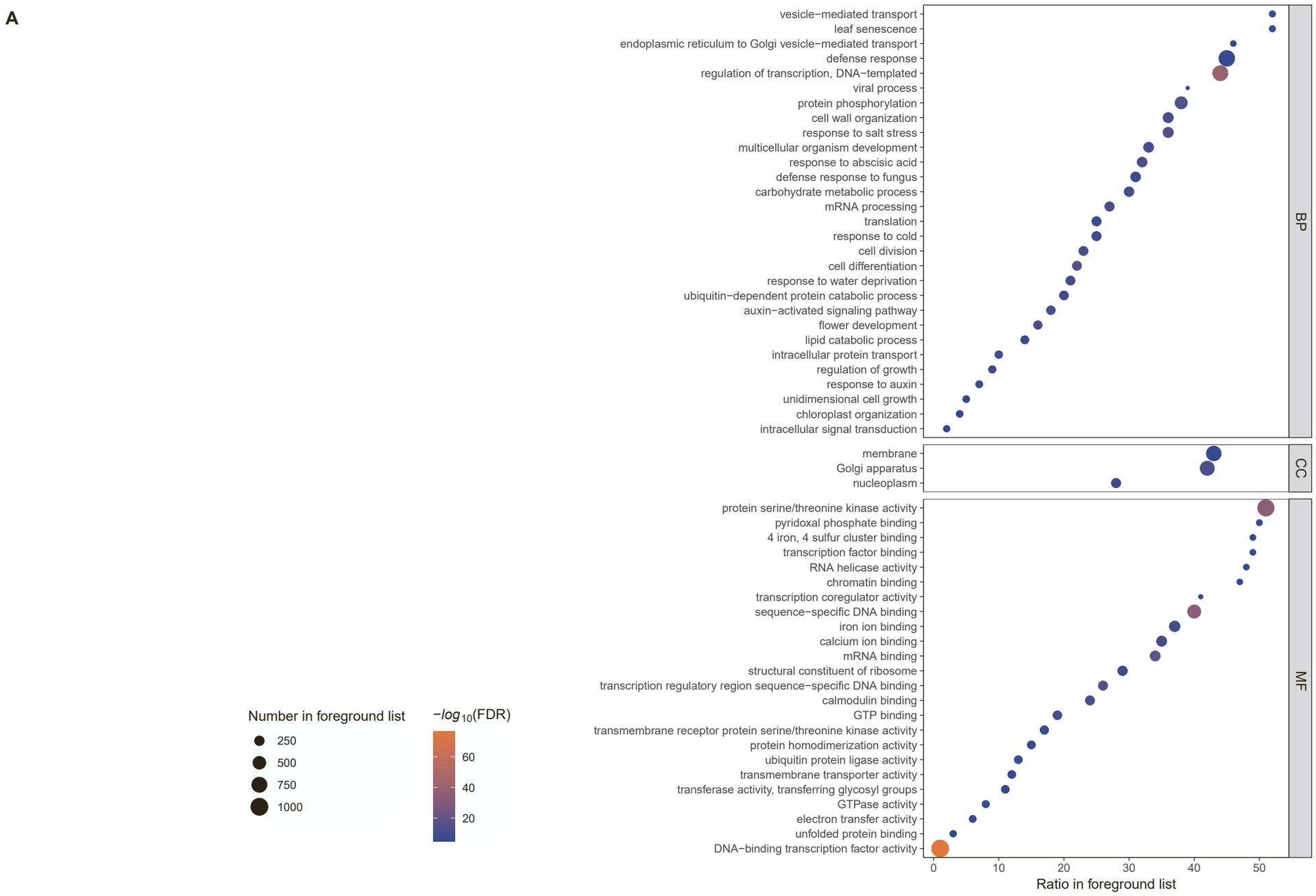
Supplemental Figures

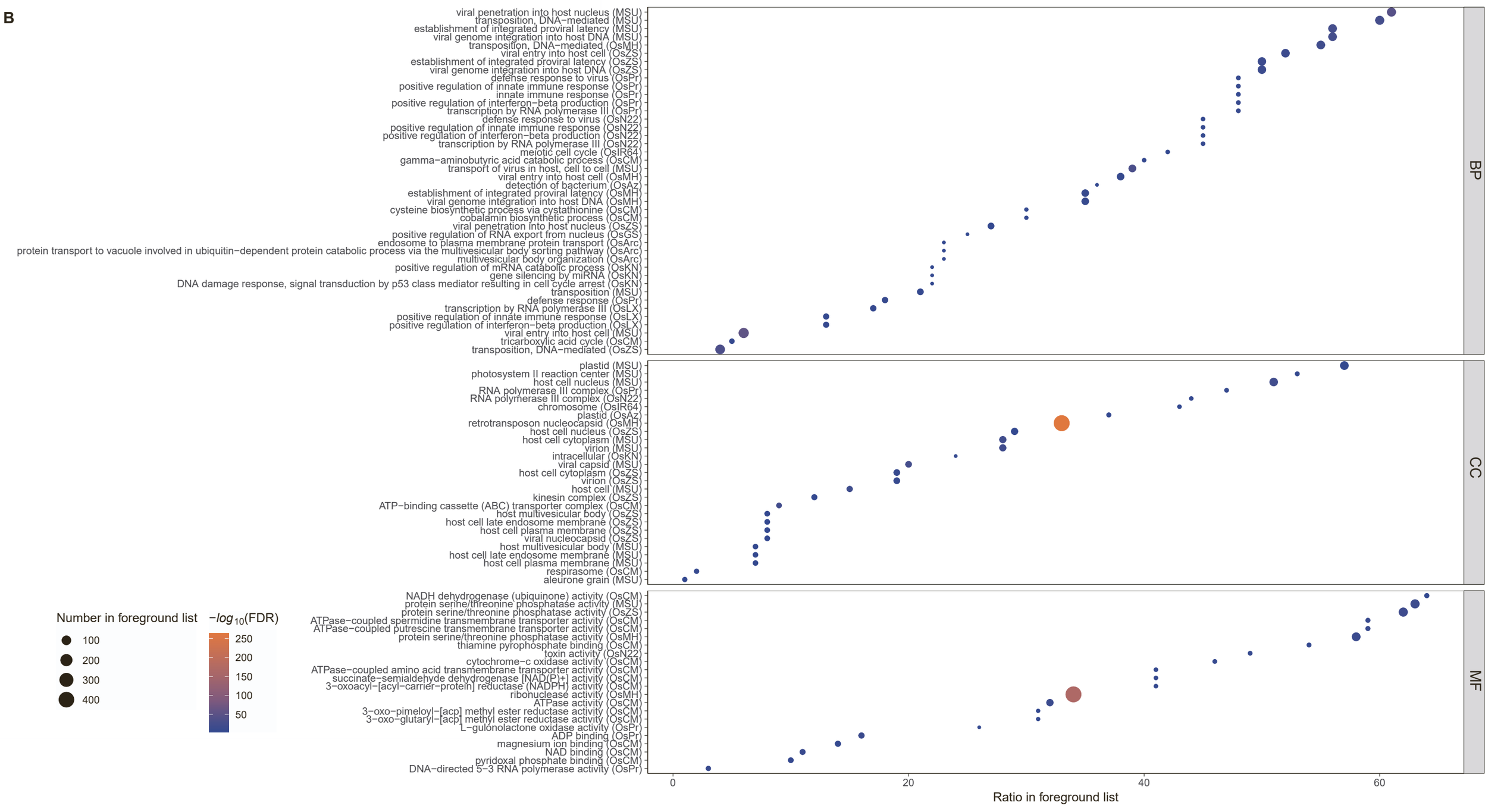


Supplemental Figure 1. Gene presence/absence variations in 16 Asian rice accessions. (A) The core genes are present in all accessions, the dispensable genes are present in <16 of accessions, and the accession-specific genes are only present in one accession. The presence of genes is colorful (4 subpopulations in Asian rice: *Geng/japonica*, *Xian/indica*, *circum-Aus*, and *circum-Basmati*), and the absence of genes is white. (B and C) The numbers of core genes (B) and accession-specific genes (C) that emerged at different evolutionary times, from PS1 (single-cell organisms) to PS13 (*O. sativa*). (D) The age distribution (left) and gene length (right) of the core and genes. (See Methods)

	Os GJ-temp: Nipponbare | IRGSP 1.0 | MSU	Os GJ-temp: Nipponbare | IRGSP 1.0 | RAPdb	Os GJ-temp: Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	Os GJ-subtrp: CHAO MEO | Os132278RS1 | Gramene(+IsoSeq)	Os GJ-trop1: Azucena | AzucenaRS1 | Gramene(+IsoSeq)	Os GJ-trop2: KETAN NANGKA | Os128077RS1 | Gramene(+IsoSeq)	Os cB: ARC 10497 | Os117425RS1 | Gramene(+IsoSeq)	Os XI-1B2: PR 106 | Os127742RS1 | Gramene(+IsoSeq)	Os XI-adm: Minghui 63 | MH63RS3 | HZAU	Os XI-1B1: IR 64 | OsIR64RS1 | Gramene(+IsoSeq)	Os XI-1A: Zhenshan 97 | ZS97RS3 | HZAU	Os XI-3A: LIMA | Os127564RS1 | Gramene(+IsoSeq)	Os XI-3B1: KHAO YAI GUANG | Os127518RS1 | Gramene(+IsoSeq)	Os XI-2A: GOBOL SAIL | Os132424RS1 | Gramene(+IsoSeq)	Os XI-3B2: LIU XU | Os125827RS1 | Gramene(+IsoSeq)	Os XI-2B: LARHA MUGAD | Os125619RS1 | Gramene(+IsoSeq)	Os cA1: N22 | OsN22RS2 | Gramene(+IsoSeq)	Os cA2: NATEL BORO | Os127652RS1 | Gramene(+IsoSeq)
Os GJ-temp: Nipponbare | IRGSP 1.0 | MSU	-	28365	28169	32747	32703	32678	32354	32121	47376	31789	49471	32849	32318	32118	38300	32139	32139	31858
Os GJ-temp: Nipponbare | IRGSP 1.0 | RAPdb	32576	-	35396	28760	28730	28820	28523	28665	29786	28326	30393	29211	28879	28639	33808	28632	28595	28462
Os GJ-temp: Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	32598	35493	-	28773	28724	28839	28572	28563	29691	28287	30336	29127	28792	28521	33747	28573	28552	28399
Os GJ-subtrp: CHAO MEO | Os132278RS1 | Gramene(+IsoSeq)	34911	28552	28396	-	35441	35287	34813	34106	31273	33680	31991	34811	34300	34015	40535	34072	34121	33727
Os GJ-trop1: Azucena | AzucenaRS1 | Gramene(+IsoSeq)	34915	28534	28390	35447	-	35421	34869	34220	31330	33747	32018	34921	34319	34087	40717	34143	34199	33875
Os GJ-trop2: KETAN NANGKA | Os128077RS1 | Gramene(+IsoSeq)	35565	28522	28410	35224	35359	-	35002	34225	32020	33950	32636	34980	34399	34076	40793	34253	34329	33828
Os cB: ARC 10497 | Os117425RS1 | Gramene(+IsoSeq)	34440	28374	28285	34813	34877	35075	-	34156	31195	33954	31863	34971	34335	34022	40739	34252	34400	33829
Os XI-1B2: PR 106 | Os127742RS1 | Gramene(+IsoSeq)	34296	28360	28160	33989	34111	34190	34035	-	31787	34672	32299	35342	35013	34757	41352	34606	34277	33850
Os XI-adm: Minghui 63 | MH63RS3 | HZAU	48304	27546	27342	31531	31592	31666	31459	31934	-	31581	54830	32475	31995	31815	37886	31671	31632	31383
Os XI-1B1: IR 64 | OsIR64RS1 | Gramene(+IsoSeq)	34077	28141	28000	33696	33758	34040	33966	34797	31673	-	32164	35247	34873	34557	41177	34465	34254	33658
Os XI-1A: Zhenshan 97 | ZS97RS3 | HZAU	48497	27855	27665	31751	31810	31901	31650	32044	52951	31712	-	32602	32136	31891	38134	31856	31854	31585
Os XI-3A: LIMA | Os127564RS1 | Gramene(+IsoSeq)	34186	28375	28186	34004	34103	34248	34127	34669	31610	34435	32106	-	35182	34816	41504	34657	34264	33871
Os XI-3B1: KHAO YAI GUANG | Os127518RS1 | Gramene(+IsoSeq)	34071	28399	28213	34007	33999	34168	34037	34800	31478	34557	32041	35718	-	34867	41504	34686	34329	33849
Os XI-2A: GOBOL SAIL | Os132424RS1 | Gramene(+IsoSeq)	33982	28385	28168	33887	33968	34006	33887	34724	31411	34394	31913	35495	35029	-	41318	34583	34281	33997
Os XI-3B2: LIU XU | Os125827RS1 | Gramene(+IsoSeq)	34362	28517	28346	34181	34290	34451	34357	34887	31709	34654	32239	35728	35217	34890	-	34767	34589	34013
Os XI-2B: LARHA MUGAD | Os125619RS1 | Gramene(+IsoSeq)	34159	28333	28166	33985	34026	34221	34149	34628	31377	34353	31955	35372	34890	34597	41120	-	34363	33988
Os cA1: N22 | OsN22RS2 | Gramene(+IsoSeq)	34442	28425	28290	34059	34172	34353	34362	34359	31536	34192	32188	35002	34575	34347	41010	34411	-	34592
Os cA2: NATEL BORO | Os127652RS1 | Gramene(+IsoSeq)	34053	28365	28190	33779	33978	33974	33934	34067	31239	33727	31788	34728	34230	34220	40557	34140	34713	-

Supplemental Figure 2. The 18 x 18 gene annotation matrix, which shows the number of 1-to-1 homologs, and demonstrates the homologous gene pairs. The color corresponds to the number of homologs. Red represents higher gene pairs number and blue represents lower gene pairs number.





Supplemental Figure 3. The bubble plot of the Gene Ontology enrichment analysis in (A) the core gene set and (B) the accession-specific gene set.

A

Basic Info

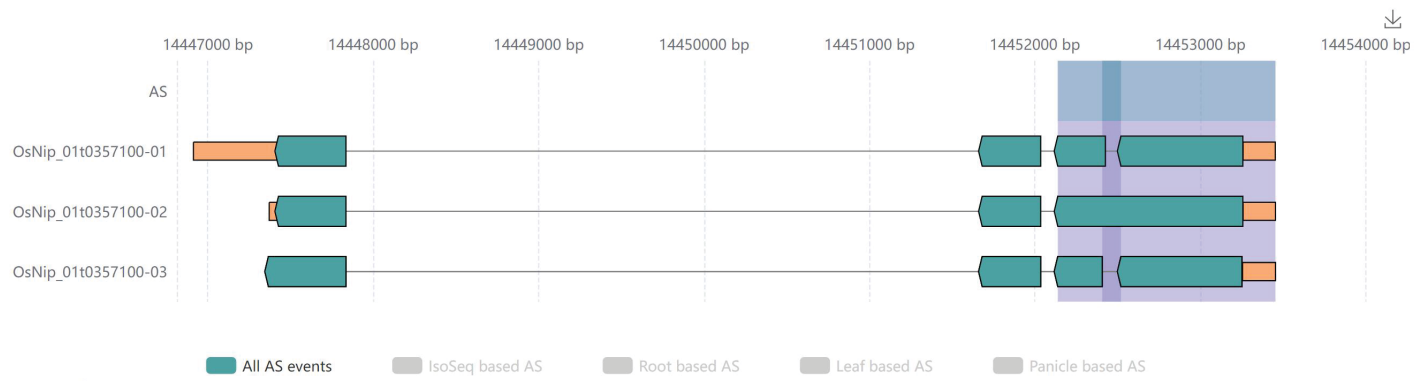
Search in RGI | Search

Gene ID	OGI # | Score	Accession | Assembly | Annotation	Location	Links
OsNip_01g0357100	OGI:01044620 | 16/16	Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	Chr01:14446913-14453454	Homologues
Gene Symbol	PSR1, PSR1, NiR, OsNiR, OsNIR1, NIR1, FD-NiR, OsNiR2			MicroCollinearity
DNA seq	GAACCTTATCTCCTTCTCTCTCGTCGCTTTCTGCGTCTCCCCGTCTCTCCTTCGCCAACAGCCGAGAAGAGGCAGAGAGAGCGCCGCCCCCCGTCCCTCTCTCTCCC TCTCGTCCTCGCCCCCATCCCTCTCGTCTTTCCCTTGCCGGCAGCAGAGGAGGCGGCAGCGACGGCTTCAGCTGCTCCCACGGGCCGGATCGGGCAGTGGCGG... More			RGI-JBrowse
CDS seq	ATGGCCTCCTCCGCCTCCCTGCAGCGCTTCCTCCCCCCGTACCCCCACGCGGCAGCATCCCGCTGCCGCCCTCCCGGCGTCCGCGCCCGCCCCGTGCAGTCGTCG ACGGTGTCCGCACCGTCCTCCTCGACTCCGGCGGCGGACGAGGCCGTGTCGGCGGAGCGGCTGGAGCCGCGGGTGGAGCAGCGGGAGGGCCGGTACTGGGT... More			
Protein seq	MASSASLQRFLPPYPHAAASRCRPPGVRARPVQSSTVSAPSSSTPAADEAVSAERLEPRVEQREGRYWVLKEKYRTGLNPQEKVKLGKEPMSLFMEGGIKELAKMPM EEIEADKLSKEDIDVRLKWLGLFHRRKHQYGRFMMRLKLPNGVTTSEQTRYLASVIEAYGKEGCADVTTRQNWQIRGVTLPDVPAILDGLNAVGLTSLQSGMDNVRN PVGNPLAGIDPDEIVDTRSYTNLLSSYITSNFQGNPTITNLPRKWNVCVIGSHDLYEHPHINDLAYMPAVKGGKFGFNLLVGGFISPKRWEEALPLDAWVPGDDIIPVC KAVLEAYRDLGTRGNRQKTRMMWLIDELVNHFFLHSSTPLTELRMSQCSYQLIAVLALQGMEAFRSEVEKRMPNGVLERAAPEDLIDKKWQRRDYLGVHPQKQEG MSYVGLHVPVGRVQAADMFELARLADEYGSGELRLTVEQNIVIPNVKNEKVEALLSEPLLQKFSPQPSLLLKGLVACTGNQFCGQAIIETKQRALLVTSQVEKLVSVPRA VRMHWTGCPNSCGQVQVADIGFMGCLTKDSAGKIVEAADIFVGGRVGSDSHLAGAYKKSVPCDELAPIVADILVERFGAVRREREEDEE			
Function	Ferredoxin--nitrite reductase, chloroplastic [UniProtKB/Swiss-Prot:Q42997]			

B

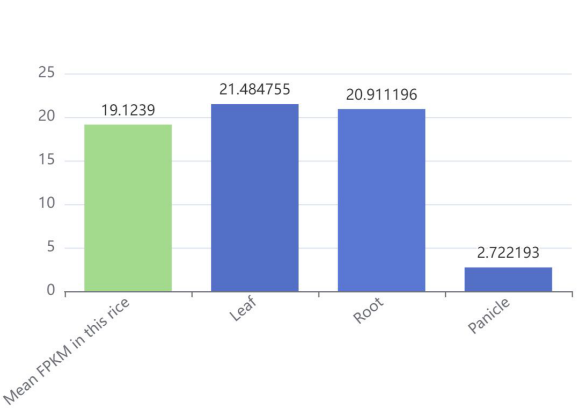
Transcripts

Transcripts browser | Transcripts table | Alternative Splicing events browser | Alternative Splicing events table



Expression

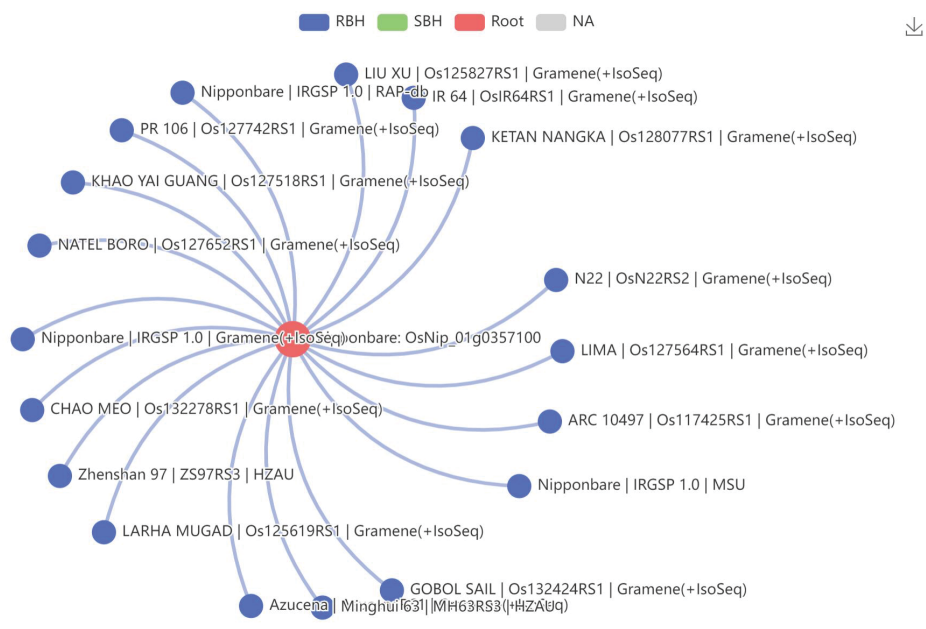
Browser | Table



C

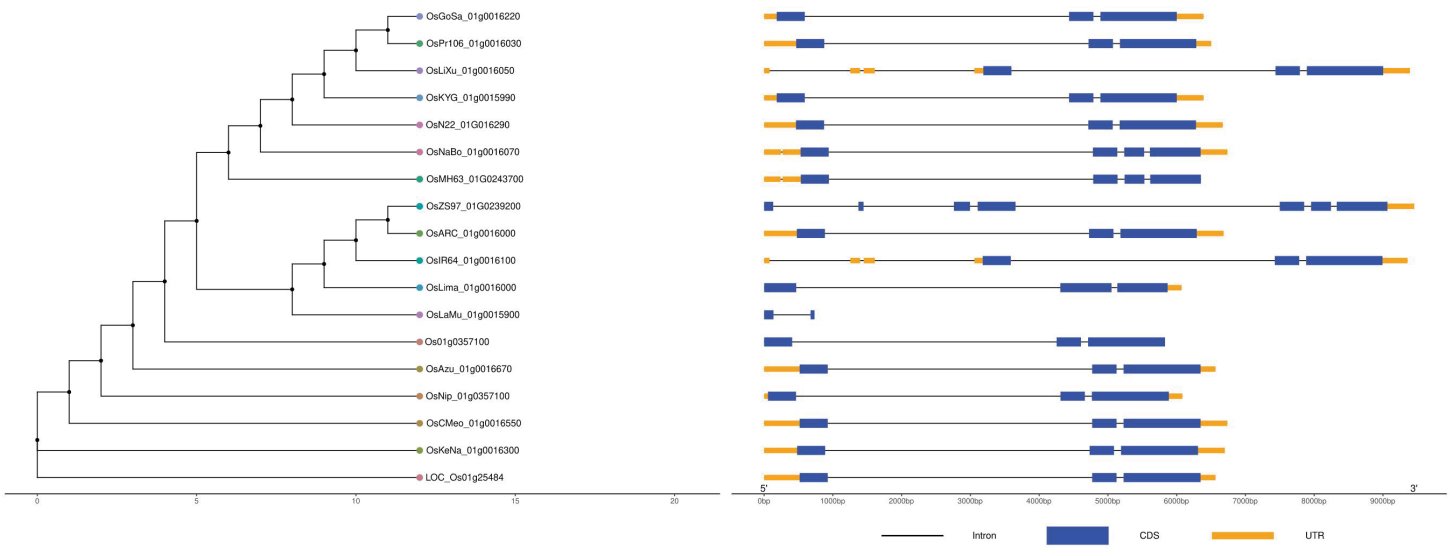
Homologues

Browser | Table

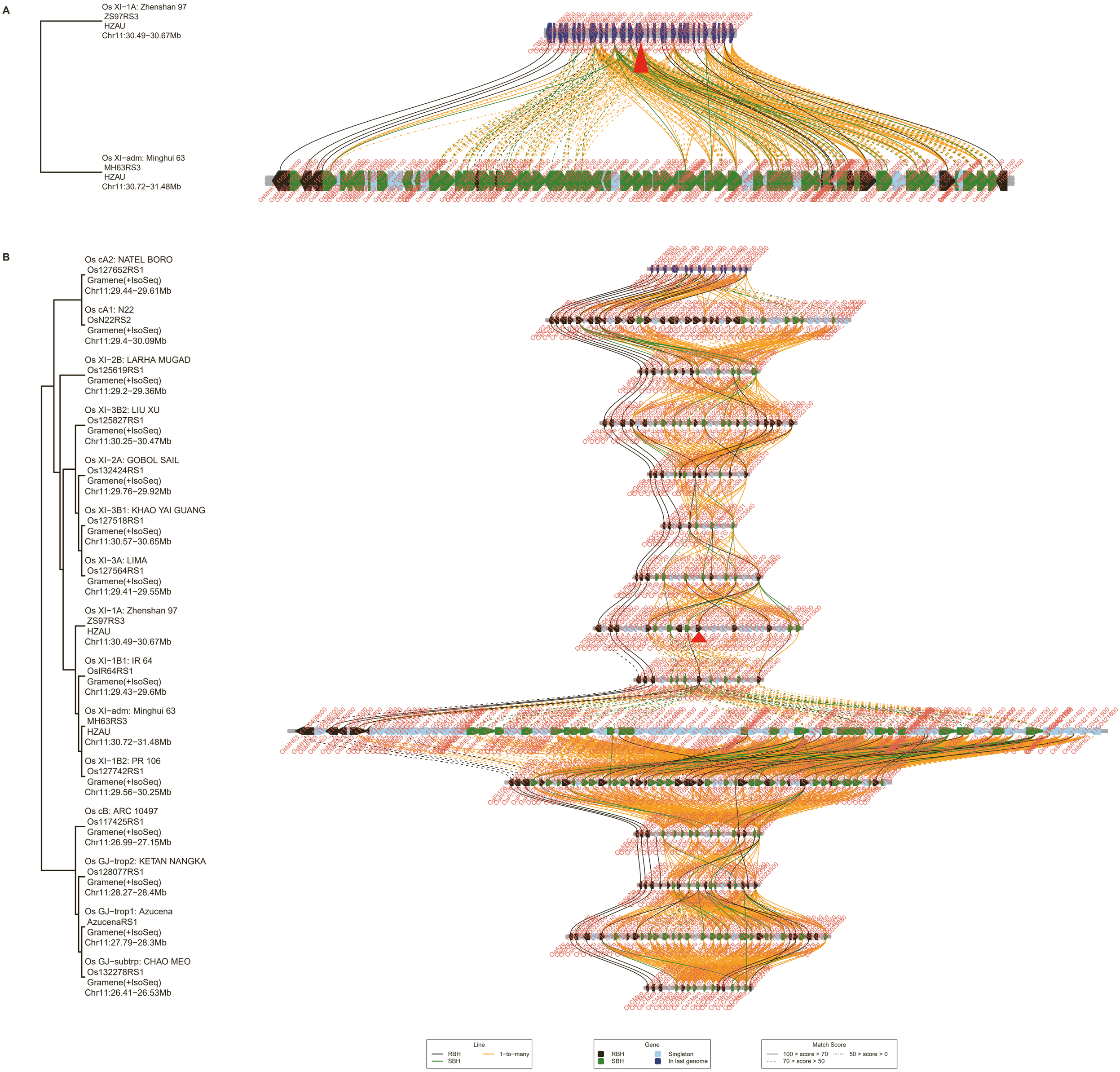


TreePlot

Custom TreePlot



Supplemental Figure 4. The "GeneCard" page. (A) Basic information. (B)Transcriptome information. (C) Homologues information.



Supplemental Figure 5. In previous studies, nucleotide-binding site leucine-rich repeat (NLR) genes were found to be highly duplicated in chromosome 11 of MH63 compared with other genomes. (A) displays the micro-collinearity of NLR genes in MH63 and ZS97 by searching NLR gene OsZS97_11G0430400 in the "MicroCollinearity" module. (B) displays the micro-collinearity of NLR genes in 16 varieties by searching NLR gene OsZS97_11G0430400 in the "MicroCollinearity" module. (A) and (B) show OsZS97_11G0430400's 15 flanking genes.

Supplemental Tables

Supplemental Table 1. Summary of RGI data composition.

Accession Name	Subpopulations	Assembly	Annotation[a.]	Locus Tag Prefix
Os GJ-temp: Nipponbare	Geng-japonica-temp	IRGSP1.0	Gramene (+IsoSeq) MSU RAP-db	OsNip_ LOC_Os Os
Os GJ-subtrp: CHAO MEO	Geng-japonica-trop1	AzucenaRS1	Gramene (+IsoSeq)	OsAzu_
Os GJ-trop1: Azucena	Geng-japonica-trop2	Os128077RS1	Gramene (+IsoSeq)	OsKeNa_
Os GJ-trop2: KETAN NANGKA	Geng-japonica-subtrp	Os132278RS1	Gramene (+IsoSeq)	OsCMeo_
Os cB: ARC 10497	circum-Basmati	Os117425RS1	Gramene (+IsoSeq)	OsARC_
Os XI-1B2: PR 106	Xian-indica-1B2	Os127742RS1	Gramene (+IsoSeq)	OsPr106_
Os XI-adm: MH63	Xian-indica-adm	MH63RS3	HZAU	OsMH63_
Os XI-1B1: IR 64	Xian-indica-1B1	OsIR64RS1	Gramene (+IsoSeq)	OsIR64_
Os XI-1A: ZS97	Xian-indica-1A	ZS97RS3	HZAU	OsZS97_
Os XI-3A: LIMA	Xian-indica-3A	Os127564RS1	Gramene (+IsoSeq)	OsLima_
Os XI-3B1: KHAO YAI GUANG	Xian-indica-3B1	Os127518RS1	Gramene (+IsoSeq)	OsKYG_
Os XI-2A: GOBOL SAIL	Xian-indica-2A	Os132424RS1	Gramene (+IsoSeq)	OsGoSa_
Os XI-3B2: LIU XU	Xian-indica-3B2	Os125827RS1	Gramene (+IsoSeq)	OsLiXu_
Os XI-2B: LARHA MUGAD	Xian-indica-2B	Os125619RS1	Gramene (+IsoSeq)	OsLaMu_
Os cA1: N22	circum-Aus1	OsN22RS2	Gramene (+IsoSeq)	OsN22_
Os cA2: NATEL BORO	circum-Aus2	Os127652RS1	Gramene (+IsoSeq)	OsNaBo_

[a.] this row is the annotation source, Gramene (+IsoSeq) means de novo annotation attached transcripts from Iso-Seq data

Supplemental Table 2. Iso-Seq and from multiple tissues (i.e., leaves, roots, and immature panicles) of 16 rice accessions.

Accessions	SMRT CELL[a.]			HQ FLNC Reads			HQ Transcripts			
	Leaf	Panicle	Root	Leaf	Panicle	Root	Leaf	Panicle	Root	Merged
Os GJ-temp: Nipponbare	1	1	1	19,101	26,760	34,261	14,751	20,883	27,415	40,644
Os GJ-subtrp: CHAO MEO	1	1	1	26,589	28,501	26,100	22,916	17,237	22,084	40,025
Os GJ-trop1: Azucena	1	1	1	29,209	33,939	32,440	24,908	25,060	23,557	45,578
Os GJ-trop2: KETAN NANGKA	1	1	2	17,148	53,122	28,529	9,905	25,429	24,123	37,942
Os cB: ARC 10497	1	-	1	25,932		25,265	19,225	-	19,584	29,880
Os XI-1B2: PR 106	1	1	1	31,088	44,374	26,093	19,008	21,528	21,720	38,785
Os XI-adm: MH63	-	-	-	-	-	-	-	-	-	-
Os XI-1B1: IR 64	1	1	1	19,168	31,556	31,937	14,951	24,625	21,376	39,818
Os XI-1A: ZS97	-	-	-	-	-	-	-	-	-	-
Os XI-3A: LIMA	1	-	1	32,067	-	26,603	17,771	-	14,487	23,989
Os XI-3B1: KHAO YAI GUANG	1	1	1	31,036	-	43,031	19,756	-	22,178	30,559
Os XI-2A: GOBOL SAIL	2	-	1	31,967	-	42,680	15,857	-	21,485	26,869
Os XI-3B2: LIU XU	1	1	1	23,628	31,519	23,625	14,308	16,786	19,427	32,223
Os XI-2B: LARHA MUGAD	1	-	1	29,618	-	38,265	15,838	-	19,437	25,832
Os cA1: N22	1	1	1	43,709	50,872	29,457	26,290	24,509	21,098	43,742
Os cA2: NATEL BORO	1	1	1	37,614	41,834	28,826	17,928	20,225	21,157	35,923

[a.] this row shows the number of PacBio SMRT cells.

Supplemental Table 3. RNA-Seq and from multiple tissues (i.e., leaves, roots, and immature panicles) of 16 rice accessions.

Accessions	Biological Replication[a.]			Alignment Rate		
	Leaf	Panicle	Root	Leaf	Panicle	Root
Os GJ-temp: Nipponbare	1	1	1	99.54%	99.29%	99.57%
Os GJ-subtrp: CHAO MEO	1	1	1	96.14%	93.87%	92.04%
Os GJ-trop1: Azucena	1	1	1	99.50%	99.41%	98.48%
Os GJ-trop2: KETAN NANGKA	1	1	1	71.11%	37.23%	17.79%
Os cB: ARC 10497	1	-	1	98.54%	-	94.72%
Os XI-1B2: PR 106	1	1	1	62.25%	29.27%	20.68%
Os XI-adm: MH63	2	2	2	95.12%	89.89%	85.26%
Os XI-1B1: IR 64	1	1	1	99.24%	99.22%	98.29%
Os XI-1A: ZS97	2	2	2	88.25%	88.36%	60.60%
Os XI-3A: LIMA	1	-	1	98.56%	-	96.37%
Os XI-3B1: KHAO YAI GUANG	1	1	1	94.99%	-	99.45%
Os XI-2A: GOBOL SAIL	1	-	1	96.66%	-	93.84%
Os XI-3B2: LIU XU	1	1	1	99.51%	99.29%	99.42%
Os XI-2B: LARHA MUGAD	1	-	1	86.39%	-	78.36%
Os cA1: N22	1	1	1	99.60%	99.73%	98.58%
Os cA2: NATEL BORO	1	1	1	96.37%	86.24%	86.43%

[a.] this row shows the number of biological replications.

Supplemental Table 4. Numbers of annotated genes in 16 rice accessions.

Accession|Assembly|Annotation	Annotated Genes	Genes from Iso-Seq Data
Os GJ-temp: Nipponbare | IRGSP 1.0 | MSU	55,801	-
Os GJ-temp: Nipponbare | IRGSP 1.0 | RAPdb	37,859	-
Os GJ-temp: Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	38,404	733
Os GJ-subtrp: CHAO MEO | Os132278RS1 | Gramene(+IsoSeq)	37,240	1,054
Os GJ-trop1: Azucena | AzucenaRS1 | Gramene(+IsoSeq)	36,882	1,079
Os GJ-trop2: KETAN NANGKA | Os128077RS1 | Gramene(+IsoSeq)	37,994	1,388
Os cB: ARC 10497 | Os117425RS1 | Gramene(+IsoSeq)	37,181	844
Os XI-1B2: PR 106 | Os127742RS1 | Gramene(+IsoSeq)	36,720	1,255
Os XI-adm: Minghui 63 | MH63RS3 | HZAU	59,903	-
Os XI-1B1: IR 64 | OsIR64RS1 | Gramene(+IsoSeq)	36,925	987
Os XI-1A: Zhenshan 97 | ZS97RS3 | HZAU	60,935	-
Os XI-3A: LIMA | Os127564RS1 | Gramene(+IsoSeq)	37,994	973
Os XI-3B1: KHAO YAI GUANG | Os127518RS1 | Gramene(+IsoSeq)	37,522	1,105
Os XI-2A: GOBOL SAIL | Os132424RS1 | Gramene(+IsoSeq)	36,467	1,810
Os XI-3B2: LIU XU | Os125827RS1 | Gramene(+IsoSeq)	44,942	1,183
Os XI-2B: LARHA MUGAD | Os125619RS1 | Gramene(+IsoSeq)	37,474	1,244
Os cA1: N22 | OsN22RS2 | Gramene(+IsoSeq)	37,598	1,280
Os cA2: NATEL BORO | Os127652RS1 | Gramene(+IsoSeq)	36,392	1,664

Supplemental Table 5. Summary of alternative splicing events in 16 rice accessions.

Accession|Assembly|Annotation	A3	A5	AF	AL	MX	RI	SE	Total
Os GJ-temp: Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	1,250	634	237	91	2	1,181	326	3,721
Os GJ-subtrp: CHAO MEO | Os132278RS1 | Gramene(+IsoSeq)	2,761	2,574	530	182	113	2,228	1,715	10,103
Os GJ-trop1: Azucena | AzucenaRS1 | Gramene(+IsoSeq)	2,735	2,583	575	187	100	2,257	1,719	10,156
Os GJ-trop2: KETAN NANGKA | Os128077RS1 | Gramene(+IsoSeq)	2,484	2,350	509	167	101	2,072	1,581	9,264
Os cB: ARC 10497 | Os117425RS1 | Gramene(+IsoSeq)	2,253	2,009	451	151	91	1,912	1,396	8,263
Os XI-1B2: PR 106 | Os127742RS1 | Gramene(+IsoSeq)	2,530	2,415	519	169	116	2,153	1,637	9,539
Os XI-adm: Minghui 63 | MH63RS3 | HZAU	-	-	-	-	-	-	-	-
Os XI-1B1: IR 64 | OsIR64RS1 | Gramene(+IsoSeq)	2,634	2,434	546	208	96	2,177	1,646	9,741
Os XI-1A: Zhenshan 97 | ZS97RS3 | HZAU	-	-	-	-	-	-	-	-
Os XI-3A: LIMA | Os127564RS1 | Gramene(+IsoSeq)	1,841	1,670	367	127	62	1,791	1,182	7,040
Os XI-3B1: KHAO YAI GUANG | Os127518RS1 | Gramene(+IsoSeq)	1,986	1,770	403	149	60	1,792	1,236	7,396
Os XI-2A: GOBOL SAIL | Os132424RS1 | Gramene(+IsoSeq)	1,563	1,355	351	115	45	1,450	940	5,819
Os XI-3B2: LIU XU | Os125827RS1 | Gramene(+IsoSeq)	2,537	2,424	493	168	97	1,981	1,683	9,383
Os XI-2B: LARHA MUGAD | Os125619RS1 | Gramene(+IsoSeq)	1,803	1,588	377	154	57	1,611	1,129	6,719
Os cA1: N22 | OsN22RS2 | Gramene(+IsoSeq)	2,742	2,548	514	937	108	2,464	1,705	11,018
Os cA2: NATEL BORO | Os127652RS1 | Gramene(+IsoSeq)	2,281	2,198	474	156	104	1,796	1,469	8,478

Supplemental Table 6. Homologues of gene LOC_Os11g29290 in 16 accessions. The result was produced by the "Homologues" module and checked manually. "RBH" means reciprocal best hits, "NA" means no homologous relationship.

Accession|Assembly|Annotation	Homologous	Type
Os GJ-temp: Nipponbare | IRGSP 1.0 | MSU	LOC_Os11g29290	RBH
Os GJ-temp: Nipponbare | IRGSP 1.0 | RAPdb	Os11g0483000	RBH
Os GJ-temp: Nipponbare | IRGSP 1.0 | Gramene(+IsoSeq)	OsNip_11g0483000	RBH
Os GJ-subtrp: CHAO MEO | Os132278RS1 | Gramene(+IsoSeq)	OsCMeo_11g0013950	RBH
Os GJ-trop1: Azucena | AzucenaRS1 | Gramene(+IsoSeq)	OsAzu_11g0014030	RBH
Os GJ-trop2: KETAN NANGKA | Os128077RS1 | Gramene(+IsoSeq)	OsKeNa_11g0013990	RBH
Os cB: ARC 10497 | Os117425RS1 | Gramene(+IsoSeq)	OsARC_11g0013770	RBH
Os XI-1B2: PR 106 | Os127742RS1 | Gramene(+IsoSeq)	OsPr106_11g0014050	RBH
Os XI-adm: Minghui 63 | MH63RS3 | HZAU	OsMH63_11G0261700	RBH
Os XI-1B1: IR 64 | OsIR64RS1 | Gramene(+IsoSeq)	NA	NA
Os XI-1A: Zhenshan 97 | ZS97RS3 | HZAU	OsZS97_11G0273800	RBH
Os XI-3A: LIMA | Os127564RS1 | Gramene(+IsoSeq)	OsLima_11g0014120	RBH
Os XI-3B1: KHAO YAI GUANG | Os127518RS1 | Gramene(+IsoSeq)	OsKYG_11g0014170	RBH
Os XI-2A: GOBOL SAIL | Os132424RS1 | Gramene(+IsoSeq)	OsGoSa_11g0013820	RBH
Os XI-3B2: LIU XU | Os125827RS1 | Gramene(+IsoSeq)	OsLiXu_11g0013660	RBH
Os XI-2B: LARHA MUGAD | Os125619RS1 | Gramene(+IsoSeq)	OsLaMu_11g0013960	RBH
Os cA1: N22 | OsN22RS2 | Gramene(+IsoSeq)	OsN22_11G013780	RBH
Os cA2: NATEL BORO | Os127652RS1 | Gramene(+IsoSeq)	OsNaBo_11g0014000	RBH

Supplemental References

Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biology **17**:66. 10.1186/s13059-016-0924-1.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics **10**:421. 10.1186/1471-2105-10-421.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics **25**:1972-1973. 10.1093/bioinformatics/btp348.

Chen, Y., Song, W., Xie, X., Wang, Z., Guan, P., Peng, H., Jiao, Y., Ni, Z., Sun, Q., and Guo, W. (2020). A Collinearity-Incorporating Homology Inference Strategy for Connecting Emerging Assemblies in the Triticeae Tribe as a Pilot Practice in the Plant Pangenomic Era. Molecular Plant **13**:1694-1708. 10.1016/j.molp.2020.09.019.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics **30**:1236-1240. 10.1093/bioinformatics/btu031.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution **30**:772-780. 10.1093/molbev/mst010.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology **37**:907-915. 10.1038/s41587-019-0201-4.

Kuo, R.I., Cheng, Y., Zhang, R., Brown, J.W.S., Smith, J., Archibald, A.L., and Burt, D.W. (2020). Illuminating the dark side of the human transcriptome with long read transcript sequencing. BMC Genomics **21**:751. 10.1186/s12864-020-07123-7.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**:3094-3100. 10.1093/bioinformatics/bty191.

McColl, W.F., Noshita, K. (1986). On the number of edges in the transitive closure of a graph. *Discrete Applied Mathematics* **15**:67-73. 10.1016/0166-218X(86)90020-X

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution **32**:268-274. 10.1093/molbev/msu300.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology **33**:290-295. 10.1038/nbt.3122.

Priyam, A., Woodcroft, B.J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L.J., Gibbins, M.A., et al. (2019). Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. Molecular Biology and Evolution **36**:2922-2924. 10.1093/molbev/msz185.

Tardaguila, M., de la Fuente, L., Marti, C., Pereira, C., Pardo-Palacios, F.J., Del Risco, H., Ferrell, M., Mellado, M., Macchietto, M., Verheggen, K., et al. (2018). SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res **28**:396-411. 10.1101/gr.222976.117.

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., and Eyras, E. (2018).

SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biology **19**:40. 10.1186/s13059-018-1417-1.

van Dongen, S. (2000). A cluster algorithm for graphs. Information Systems.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**:43-49. 10.1038/s41586-018-0063-9.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution **8**:28-36. 10.1111/2041-210X.12628.