

# **Rice pan-genome resources in Gramene**



<sup>1</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; <sup>2</sup>CEDE-BESE, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia; <sup>3</sup>IRRI, Los Baños, Philippines; <sup>4</sup>Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ, USA; <sup>5</sup>USDA ARS NAA Robert W. Holley Center for Agriculture and Health, Ithaca,

Abstract: Rice is one of the world's most important crops and, due to its small genome size (~400Mb), is a model system for plant genetics. The Gramene database was established 20 years ago to provide a comparative mapping resource for grasses, and the first rice genome was its cornerstone. Since then, Gramene has been involved in the sequencing, assembly and annotation of other rice varieties and crop genomes to expand the resource, in collaboration with community genome projects and Ensembl. Based on the Gramene/Ensembl infrastructure, we have established focused, pan-genome sites for specific crop research communities (rice, maize, sorghum, and grapevine). Gramene Oryza (https://oryza.gramene.org) release 5 (Oct. 2022) features 28 rice genomes from other species to support comparative genomic analysis and model the evolution of gene functions. The hosted rice genomes include the MAGIC16 collection, domesticated African rice varieties, short life cycle KitaakeX, and heirloom US Carolina Gold Rice. The pan-genome site is based on 38K protein-coding gene family trees, which enable phylogenetic inference of orthologs and paralogs, provide insights into allelic genes within and across species, and catalog lineage specific expansion and contraction. In addition to individual genome browsers, we provide an integrated search interface with various ways to interrogate and visualize gene tree data. We are integrating functional annotations, phenotypes, environmental data, and genetic variation into the site and look forward to working with the community to coordinate efforts on data stewardship, biocuration and training. Gramene Oryza is funded by USDA-ARS-8062-21000-041-00D.

# Gramene Oryza Release #6 (Jan. 2023) - Updates & highlights

Functional curation for > 4000 genes from RAP-DB and NCBI's geneRIFs

## **Gramene Oryza PanGenome (oryza.gramene.org)**

 28 domesticated & wild rice genomes V A CONTRACTOR V

- New Papers tab in Search results
- 3K SNP data for Nipponbare, MH63, IR64 and Azucena accessions
- 30 ortholog-based synteny maps between Nipponbare and other rice accessions, as well as Leersia, sorghum and maize
- •Whole-Genome Alignments between Nipponbare, sorghum and maize
- •Tissue specific RNAseq tracks for Nipponbare
- •Gramene Search link added to Browser Gene pages
- •Genome nomenclature updated based on recommended standards from the NSF-DBI (#2029854): CIBR-BBSRC: PanOryza: Globally coordinated genomes, proteomes and pathways for rice



Show All 💙 entries		Show/hide columns			
Dryza sativa genes	Location		Oryza carolina homologues	Location	
0s01g0267900 (Os01g0267900)	<u>1:9195546-9198962</u>	$\rightarrow$	Oscar.UnG003840 (Oscar.UnG003840)	tig00000085:1-3389	
0s01g0268000 (Os01g0268000)	<u>1:9201354-9202372</u>	$\rightarrow$	Oscar.UnG198850 (Oscar.UnG198850)	tig00027385:4908775-4913517	
		$\rightarrow$	Oscar.UnG362940 (Oscar.UnG362940)	tig00027436:6843385-6845839	
0s01g0268100 (Os01g0268100)	<u>1:9204377-9211643</u>	$\rightarrow$	Oscar.UnG003850 (Oscar.UnG003850)	tig00000085:4343-16203	
0s01g0268300 (Os01g0268300)	<u>1:9225213-9237025</u>	$\rightarrow$	Oscar.UnG003870 (Oscar.UnG003870)	tig00000085:29565-41628	
0s01g0268600 (Os01g0268600)	<u>1:9245577-9253404</u>	$\rightarrow$	Oscar.UnG003890 (Oscar.UnG003890)	tig00000085:49449-57858	
0s01g0268700 (Os01g0268700)	<u>1:9253938-9255239</u>	$\rightarrow$	Oscar.UnG003900 (Oscar.UnG003900)	tig00000085:58322-60964	
0s01g0268733 (Os01g0268733)	<u>1:9254067-9255167</u>		No homologues		
s01g0268800 (Os01g0268800)	1:9259148-9264227	$\rightarrow$	Oscar.UnG003910 (Oscar.UnG003910)	tig0000085:63367-68734	

Sorgham bicolor									
Chr. 9	1.485.200	1,485.300	1.485.400	1,485,500	906 bp	1.485,700	1,485,800	1,485,900	Forward strand
Genes	SORBI 3009G	6015900 >	_,,.		-,			_,,	_, ,
equence					ARYC03002107.1				
Sequence S.bic-O.sat lastz-net									
hr. 9	1,485,200 ◀ Reverse strand	1,485,300	1,485,400	1,485,500	1,485,600 906 bp	1,485,700	1,485,800	1,485,900	1,486,000
Oryza sativa Japoni	ica								
Chr. 5 S.bic-O.sat lastz-net	738,200	738,300	738,400	738,500	906 bp 738,600	738,700	738,800	738,900	Forward strand
ienes	0c05c0113000 >								
Sequence									
Contigs					5:700001800000				
		<u> </u>							
hr. 5	738,200	738,300	738,400	738,500	738,600 906 bp	738,700	738,800	738,900	739,000
hr. <b>5</b> Zea mays B73 v5 (	738,200 Reverse strand <i>Zea mays B73 v5&lt;</i>	738,300 :/i>)	738,400	738,500	738,600 906 bp —	738,700	738,800	738,900	739,000
ihr. 5 Zea mays B73 v5 (	738,200 → Reverse strand — <i>Zea mays B73 v5&lt;</i>	738,300 :/i>)	738,400	738,500	738,600906 bp 906 bp	738,700	738,800	738,900	739,000 Forward strand
hr. 5 Zea mays B73 v5 ( Chr. 10	738,200 → Reverse strand	738,300 :/i>) 142,929,500	738,400	738,500	738,600 906 bp <u>906 bp</u> 142,929,800	738,700 142,929,900	738,800	738,900	739,000 Forward strand 142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 D.sat-nam11L lastz	738,200 → Reverse strand <i>Zea mays B73 v5&lt; 142,929,400</i>	738,300 ;/i>) 142,929,500	738,400	738,500	738,600 906 bp 906 bp 142,929,800	738,700	738,800 142,930,000	738,900	739,000 Forward strand 142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 .sat-nam11L lastz ene (CSHL) equence	738,200 Reverse strand <i>Zea mays B73 v5&lt; 142,929,400</i>	738,300 :/i>) 142,929,500	738,400	738,500	738,600 906 bp 906 bp 142,929,800	738,700	738,800	738,900	739,000
hr. 5 Zea mays B73 v5 ( Chr. 10 .sat-nam11L lastz ene (CSHL) equence ontigs equence	738,200 → Reverse strand	738,300 ;/i>) 142,929,500	738,400	142,929,700	738,600 906 bp 906 bp 142,929,800 142,929,800 scaf_7 >	738,700	738,800	738,900	739,000
hr. 5 Zea mays B73 v5 ( Chr. 10 .sat-nam11L lastz equence ontigs equence ene (CSHL)	738,200 Reverse strand <i>Zea mays B73 v5&lt; 142,929,400</i>	738,300 ;/i>) 142,929,500	738,400	738,500	738,600 906 bp 906 bp 142,929,800 scaf_7 >	738,700	738,800	738,900	739,000 Forward strand 1 142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 sat-nam11L lastz ene (CSHL) equence ontigs equence ene (CSHL) hr. 10	738,200 → Reverse strand <i>Zea mays B73 v5&lt; 142,929,400 142,929,400 → Reverse strand</i>	738,300 ;/i>) 142,929,500	738,400 142,929,600 142,929,600	738,500	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp	738,700 142,929,900	738,800	738,900	739,000 Forward strand 142,930,200 142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 .sat-nam11L lastz ene (CSHL) equence ene (CSHL) hr. 10 ene Legend	<ul> <li>738,200</li> <li>■ Reverse strand</li> <li><i>Zea mays B73 v5</i></li> <li>142,929,400</li> <li>142,929,400</li> <li>■ Reverse strand</li> <li>Protein Coding</li> </ul>	738,300 ;/i>) 142,929,500	738,400 142,929,600 142,929,600	738,500	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp	738,700 142,929,900	738,800 142,930,000 142,930,000	738,900	739,000  Forward strand 142,930,200  142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 Lsat-nam11L lastz ene (CSHL) equence equence ene (CSHL) hr. 10 rene Legend	<ul> <li>738,200</li> <li>■ Reverse strand</li> <li><i>Zea mays B73 v5</i></li> <li>142,929,400</li> <li>142,929,400</li> <li>■ Reverse strand</li> <li>■ Protein Coding</li> <li>■ protein coding</li> </ul>	738,300 ;/i>) 142,929,500	738,400	738,500	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp 906 bp	738,700 142,929,900 142,929,900 142,929,900	738,800 142,930,000 142,930,000	738,900	739,000
Zea mays B73 v5 ( Chr. 10 ).sat-nam11L lastz Gene (CSHL) iequence Contigs iequence iene (CSHL) ihr. 10 Gene Legend	<ul> <li>738,200</li> <li>■ Reverse strand</li> <li><i>Zea mays B73 v5</i></li> <li>142,929,400</li> <li>142,929,400</li> <li>■ Reverse strand</li> <li>■ Protein Coding</li> <li>■ protein coding</li> </ul>	738,300 ;/i>) 142,929,500	738,400	738,500 142,929,700 142,929,700 142,929,700 Other — Orthologue 1-	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp 906 bp	738,700 142,929,900 142,929,900	738,800 142,930,000 142,930,000	738,900 142,930,100 142,930,100	739,000  Forward strand  142,930,200  142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 D.sat-nam11L lastz equence ontigs equence eiene (CSHL) hr. 10 Eiene Legend	<ul> <li>Zea mays B73 v5</li> <li>Zea mays B73 v5</li> <li>142,929,400</li> <li>142,929,400</li> <li>Reverse strand</li> <li>Protein Coding</li> <li>protein coding</li> </ul>	738,300 ;/i>) 142,929,500 	738,400 142,929,600 142,929,600	738,500 142,929,700 142,929,700 142,929,700 Other Othologue Othologue 1-	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp 906 bp	738,700 142,929,900 142,929,900 142,929,900	738,800 142,930,000 142,930,000	738,900	739,000  Forward strand 142,930,200  142,930,200
hr. 5 Zea mays B73 v5 ( Chr. 10 Lsat-nam11L lastz equence ontigs equence ene (CSHL) hr. 10 ene Legend	<ul> <li>Zea mays B73 v5</li> <li>Zea mays B73 v5</li> <li>142,929,400</li> <li>142,929,400</li> <li>Reverse strand</li> <li>Protein Coding</li> <li>protein coding</li> </ul>	738,300 ;/i>) 142,929,500 142,929,500 142,929,500	738,400 142,929,600 142,929,600	738,500 142,929,700 142,929,700 142,929,700 Other Othologue Othologue 1-	738,600 906 bp 906 bp 142,929,800 scaf_7 > 142,929,800 906 bp 906 bp	738,700 142,929,900 142,929,900 142,929,900	738,800 142,930,000 142,930,000	738,900	739,000  Forward strand 1 142,930,200

- Updated gene sets with Interpro protein domains annotated
- 37,423 gene trees
- 30 pairwise synteny maps
- 3 pairwise whole-genome alignments
- Gramene Search (c.v. keyword-based)
- BLAST Search (sequence-based)



## **Gramene Search & Enhanced Tree Views**

Os. indica XI-Os. indica XI-Ds. indica XI-Ds. indica XI-Os. indica XI-: Os. indica XI-3 Os. indica XI-2E Os. indica XI-2B O

Os. indica 93

Orvza nivara

Os. japonica Os. indica cB Os. aus cA1 N

Os. japonica G

### Pangenomic search summary

	- Os. Japonica	322			
24 genes in 36 genomes 🌣	<ul> <li>Os. indica XI-3B2 Os125827</li> </ul>	602			
aro 🛱	Os. japonica Carolina	445			
	Os. indica XI-3A Os127564	512			
	Os. japonica GJ-subtrp Os13227	8470			
erPro Domain   NB-ARC	<ul> <li>Os. indica XI-adm var MH63</li> </ul>	518			
	Os. japonica GJ-trop1 Azucena	480			
	<ul> <li>Os. japonica GJ-trop2 Os128077</li> </ul>	461			
	<ul> <li>Os. indica cB Os117425</li> </ul>	459			
	Os. indica XI-3B1 Os127518	499			
	Os. indica XI-1B2 Os127742	506			
	Os. aus cA1 N22	474			
	Os. indica XI-1A var ZS97	480			
	Os. indica XI-2B Os125619	495			
	Os. indica XI-2A Os132424	473			
	Os. aus cA2 Os127652	484			
	Os. indica XI-1B1 IR64	490			
	Os. japonica var. KitaakeX	460			
	• Os. indica IR8	411			
	<ul> <li>Oryza nivara</li> </ul>	439			
	<ul> <li>Oryza meridionalis</li> </ul>	435			
	<ul> <li>Oryza glumaepatula</li> </ul>	436			
	<ul> <li>Oryza glaberrima</li> </ul>	468			
	<ul> <li>Oryza barthii</li> </ul>	462			
0	<ul> <li>Oryza rufipogon</li> </ul>	447			
	<ul> <li>Oryza punctata</li> </ul>	322			
	<ul> <li>Oryza brachyantha</li> </ul>	250			
- Y	<ul> <li>Leersia perrieri</li> </ul>	369			
	Zea mays B73 v4	127			
0	$\sim$ 7 – 9 Zea mays B73 v5	127			

# Alternate gene-tree view

#### (MH63 gene: Osmh63.11G007180)

dm var MH63							
B2 Os127742			4-1-1-4				
B1 IR64							
A var ZS97							
B1 Os127518							
A Os127564							•
32 Os125827							•
Os125619							
125619							•
-11							
Carolina							
J-trop1 Azucena		_					-
Ds117425	<b>→</b>						-
2	<b></b>						_
127652	<b></b>						—
-subtrp Os132278	<b></b>						-
	<b>→</b>	-	<u>Öröttöri</u>				_
rolina	<b></b>				_		—
na	<b>→</b>	_		- <b>T</b> > -			_
	<b></b>						_
а	<b>→</b>	_					_
	<b>→</b>	_					_
	<b>+</b>		i-i-i-i				_
I	<b>→</b>					<b></b>	_
ıla	<b>→</b>			-()			—
s, 1 paralog			$\overline{+}$		<del></del>	+ $+$ $+$ $+$ $+$ $+$ $+$ $+$ $+$ $+$	-
			<u></u>	- <b>[</b> ])- ' - '		i-i-i-i-	
		_		-1)			
	-1-1-1			- <b>T</b> )			•
				-1)			•
genes			+ + + +				-
-							
genes							
			<b></b>				
		• <b>••••</b> ••		· ()	-		-
							<u> </u>



Nipponbare chr1 corresponding to 6 Carolina Gold contigs. The synteny was built based on colinear orthologs with less than 5 intervening genes. On the bottom is a table listing Nipponbare – Carolina orthologs in this selected region.

The figure to the left shows an example of synteny between Nipponbare

IRGSP chr1 and US Carolina Gold contigs. There are 6 syntenic regions on

 The figure above shows Whole Genome Alignment region comparison view between Nipponbare, sorghum and maize. Orthologs connected with a line.



Song J.M. et al. (2021). Two gap-free reference genomes and a global view of the centromere architecture in rice. Mol. Plant, 14 (2021), pp. 1757-1767



The search for Interpro domain NB-ARC results in 14,224 genes from the 36 genomes hosted on Gramene Oryza panSite. The genomes are organized as a phylogenomic tree with color-coded chromosomes concatenated. These hits are marked as hashes along each genome bar (contigs are grey bars). For example, O sativa Japonica Nipponbare has 322 hits distributed across all 12 chromosomes.

The right figure shows neighborhood gene tree view of a target gene, this gene family displays a clade specific conservation at this neighborhood region.

#### **BLAST Search**



Us. Japonica Os. japonica Car 🔾 Oryza glaberrii Oryza barthii Oryza glaberrim Oryza nivara Os. indica IR8 Oryza rufipogor Oryza glumaepat Oryza: 24 gene Oryza meridionalis Oryza punctata Oryza brachyantha Leersia perrieri Andropogoneae: rosids: 2 genes S. moellendorffii: 3 C. reinhardtii — D. melanogaster Eukaryota: 2 genes Eukaryota: 42 genes, 1 paralog

Job 1: BLASTP against Leersia perrieri Lperr_V1.4 (Proteins) Done: 3 hits found [View results]
Job 2: BLASTP against Oryza barthii AGI_PacBio (Proteins) Done: 7 hits found [View results]
Job 3: BLASTP against Oryza brachyantha AGI_PacBIO (Proteins) Done: 6 hits found [View results]
Job 4: BLASTP against Oryza sativa japonica var. Carolina CGRFv1 (Proteins) Done: 7 hits found [View results]
Job 5: BLASTP against Oryza glaberrima AGI_PacBIO (Proteins) Done: 4 hits found [View results]
Job 6: BLASTP against Oryza glumaepatula AGI_PacBIO (Proteins) Done: 5 hits found [View results]
Job 7: BLASTP against Oryza sativa indica 93-11 ASM386521v1 (Proteins) Done: 7 hits found [View results]
Job 8: BLASTP against Oryza sativa indica var. IR8 Rice_IR8_v1.7 (Proteins) Done: 10 hits found [View results]
Job 9: BLASTP against Oryza meridionalis AGI_PacBIO (Proteins) Done: 4 hits found [View results]
Job 10: BLASTP against Oryza nivara AGI_PacBio (Proteins) Done: 5 hits found [View results]
Job 11: BLASTP against Oryza punctata AGI_PacBIO (Proteins) Done: 4 hits found [View results]
Job 12: BLASTP against Oryza rufipogon OR_W1943 (Proteins) Done: 4 hits found [View results]
🗱 Job 13: BLASTP against Oryza sativa (Geng/Japonica-temp var. Nipponbare) IRGSP-1.0 (Proteins) Done: 4 hits found [View results]
Job 14: BLASTP against Oryza sativa (circum-Basmati var. ARC 10497) Os117425RS1 (Proteins) Done: 5 hits found [View results]
Job 15: BLASTP against Oryza sativa (Xian/Indica-2B var. Larha Mugad) Os125619RS1 (Proteins) Done: 3 hits found [View results]
Job 16: BLASTP against Oryza sativa (Xian/Indica-3B2 var. Liu Xu) Os125827RS1 (Proteins) Done: 4 hits found [View results]
Job 17: BLASTP against Oryza sativa (Xian/Indica-3B1 var. Khao Yai Guang) Os127518RS1 (Proteins) Done: 4 hits found [View results]
Job 18: BLASTP against Oryza sativa (Xian/Indica-3A var. Lima) Os127564RS1 (Proteins) Done: 4 hits found [View results]
Job 19: BLASTP against Oryza sativa (Geng/Japonica-trop1 var. Azucena) AzucenaRS1 (Proteins) Done: 5 hits found [View results]
Job 20: BLASTP against Oryza sativa (Xian/Indica-1B1 var. IR64) OsIR64RS1 (Proteins) Done: 4 hits found [View results]
Job 21: BLASTP against Oryza sativa ssp. japonica var. KitaakeX Osativa_Kitaake_v2.0 (Proteins) Done: 2 hits found [View results]
Job 22: BLASTP against Oryza sativa (Xian/Indica-adm var. Minghui 63) MH63RS2 (Proteins) Done: 6 hits found [View results]
Job 23: BLASTP against Oryza sativa (Xian/Indica-1A var. Zhenshan 97) ZS97RS3 (Proteins) Done: 6 hits found [View results]

- Zhou Y. et al. (2020). A platinum standard pan-genome resource that represents the population structure of Asian rice. Sci. Data, 7 (2020), p. 113 2. Stein J.C. et al. (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza Nature Genetics. 50:285.
- 3. Vilella A.J. et al. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Research 19 (2):327-35 4. Haas B.J. et al (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20 (18): 3643-46 Jones P. et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9): 1236-40

#### The BLAST Search can run sequence search against up to 25 genomes at the same time, the right figure shows the results of the 25 blast jobs.



Go

Oryza sativa indica 93-1

Oryza barthii

AGI\_PacBio

Orvza sativa japonica var. Carolina

✓ for