Rice Annotation Project Database (RAP-DB) Update 2025 How We Maintain Up-to-Date Functional Information on Rice Genes

Yoshihiro Kawahara¹, Tomoko Hirozane-Kishikawa¹, Ryo Hirata², Yuki Tamagaki³, Yumiko Teramoto², Xiaohui Wang¹, Norio Tabei⁴, Masahiko Kumagai¹, Hiroaki Sakai¹, Takeshi Itoh⁵



The Rice Annotation Project Database (RAP-DB) has been providing functional annotation for rice genes for over 20 years. Over the past year, we have manually curated and updated information on the exon-intron structure and function of approximately 700 loci by thoroughly reviewing relevant literatures. To date, we have curated rice gene information for approximately 6,000 loci. Notably, in response to user feedback, we recently revised the gene annotations for over 1,000 nucleotide-binding leucine-rich repeat containing receptor genes reported in Gottin et al. (2021). Furthermore, to improve the efficiency of literature curation, we have integrated natural language processing technology into our workflow. A newly developed system utilizes AI models to identify publications containing rice gene-related information, greatly accelerating the curation process.

The following data and features will be available in near future. 1) Orthologous Gene Information Across Model **Crops**: we have inferred orthologous gene candidates for model crops such as wheat, maize and soybean based on amino acid sequence similarity and synteny. This information will be accessible via each gene page, enabling the use of agronomically important gene information across species. 2) User Feedback System: A feedback system will be launched to allow users to contribute additional literature references, functional annotation, or proposed corrections to gene structures. This will facilitate **community-driven updates** and ensure RAP-DB remains a comprehensive resource of rice genes. We hope that these updates will make RAP-DB a more user-friendly and reliable resource for rice genomics research. Please follow our X account (@rapdbjp), where we post information about rice-related research, updates on RAP-DB, etc.

Description

waxy, Waxy

Os06g0133000

Other variants

Gene ontology (GO)

Starch synthase, catalytic domain (IPR013534)

Position

Locus

Wx

Comparison of two major gene annotation for the 'Nipponbare' reference genome

	RAP-DB	RGAP
Organization	NARO, Japan	UGA, USA
Release	2024-07-12	Release 7
Last upadate	Jul 12, 2024	Feb 7, 2012
Frequency of updates	twice a year	no updates
# of loci	45,878	55,986
# of transcripts		
- Protein-coding	50,778	66,338
- Non-coding	2,332	0
BUSCO (%, Protein)		



JRL	https://rapdb.dna.affrc.go.jp/	https://rice.uga.edu/
- Missing	0.4	3.4
- Fragmented	0.5	5.0
- Complete (Single, Dup.)	99.1 (85.3, 13.8)	91.6 (70.7 <i>,</i> 20.9)

Functional annotation of rice genes in RAP-DB



upon the completion of the *Oryza sativa* ssp. *japonica* cv. Nipponbare genome sequencing by the **International Rice** quencing Project with the aim of providing the scientific community with an accurate and timely annotation of the rice genome sequence. One of the major objectives of this project is to facilitate a comprehensive analysis of the genome structure and function of rice on the basis of the annotation



12/Jul/2024 NEW • We have updated CGSNL annotation and manual curation data (see update 2024-<u>07-12.txt</u>). QTL annotation in Q-TARO (QTL Annotation Rice Online database) is now available in the JBrowse ("Other rice annotations" -> "Q-TARO") 1/Jan/2024 We have updated CGSNL annotation and manual curation data (see update 2024-<u>01-11.txt</u>). <u>TASUKE+ (685 varieties)</u> has been updated to the latest version (TASUKE+ ver. 20231214) 7/Sep/2023 · We have updated CGSNL annotation and manual curation data (see update 2023-

<u>09-07.txt</u>). "Known functional alleles & mutations" data and RNA-Seq data have been 🕑 More 💥 @rapdbjp

Genes can be searched by ID, symbol, etc., and results can be downloaded.

Select search fields
A Search and Retrieval





Genome-wide polymorphisms (SNPs and InDels) can be browsed on TASUKE+.



[PubMed] Isshiki M et al. Plant J, 15(1):133-8 (1998) [PubMed] Hirano HY et al. Mol Biol Evol 15(2):022 83 (100

Cai XL et al. Plant J, 4(4):459-65 (1998) [PubMed] Zhao Q et al. Nat Gen 50(2):278-284 (2018) amanaka S et al. The ppl Genet, 108(7):12

Yong-zhong Li et al. Pl Science, 108(2):181-1

Liu L et al. Plant Mol Bio 71(6):609-26 (2009)

Bao JS et al. Theor Ap

Isshiki M et al. Plant Physiol, 125(3):1388-

Isshiki M et al. Plant

Isshiki M et al. Plant

Bao JS et al. Theor App Senet, 113(7):1171-83

Ikuo Ando et al. Breeding Science, 60(3):187-194

🔽 Literat

Expres

V Diver 🗸 Oryza

KEGG

🗸 Seque

Note

ikami I et al. Theor Ap enet, 116(7):979-89

Gene Description Gene Symbol Gene Name	Keyword Se	earch A	dvanced Search	ID Search	Reset						
Gene Ontology (GO)											
Trait Ontology (TO) Functional domain (InterPro)	Search Resu	Ilts Locus 	○ Transcript								
PubMed ID	<< < > >	> 1 to 1 of 1	hits 10 🗸 per page								
□ Transcript Evidence	RAP ID					Gene Sym	bol		Gene Name		
Display options A	Locus ID	Transcript ID 🍦	Description		Position	RAP-DB	CGSNL	Oryzabase	RAP-DB	CGSNL	Oryzabase
Description Position Position Gene Symbol CGSNL Oryzabase Gene Name RAP-DB CGSNL Oryzabase Oryzabase Ontology Gene Ontology (GO) Plant Ontology (TO) Trait Ontology (TO) Functional domain (InterPro) MSU ID Reference Evidence	□ Os06g0133000 C	Ds06t0133000-01 Ds06t0133000-02	Granule-bound starch syntha amylose in endosperm (OsOG Granule-bound starch syntha biosynthesis (OsO6t0133000-	ise, Synthesis of ist0133000-01) ise 1, Starch -02)	chr06:17656221770 653 (+ strand)	WX, WX1, WX (WX(am)), WX-B, GBSS-I, GBSS-I, GBSSI, OSGBSSI, GBSSI, GBSSI	WX1	wx (Wx(am)), Wx, WX-B, GBSS-1, GBSS-1, OsGBSS1, OsGBSS1, OsGBSS1, GBSS1, GSS, OsWx	waxy, Waxy, GLUTINOUS ENDOSPERM, glutinous endosperm, WAXY, "Granule-bound starch synthase 1, chloroplastic/amyloplastic", Granule-bound starch synthase, I, UDP-glycogen synthase, "Granule-bound starch synthase, chloroplast precursor", glycogen [starch] synthase, Granule- bound glycogen synthase, UDPG-glycogen glucosyltransferase, glycogen [starch] synthetase, Granule-bound glycogen [starch] synthase, UDPG-glycogen synthetase, granule bound starch synthase 1, granule-bound starch synthase 1, granule- bound starch synthase 1	GLUTINOUS ENDOSPERM	glutinous endosperm, waxy, Waxy, WAXY, "Granule- bound starch synthase 1, chloroplastic/amyloplastic", Granule-bound starch synthase I, UDP-glycogen synthase, "Granule-bound starch synthase, chloroplast precursor", glycogen [starch] synthase, Granule- bound glycogen synthase, UDPG-glycogen transglucosylase, uridine diphosphoglucose-glycogen glucosyltransferase, glycogen [starch] synthetase, Granule-bound glycogen [starch] synthase, UDPG-glycogen synthestase, granule bound starch synthase I, granule-bound starch synthase I, granule-bound starch synthase I granule-bound starch synthase I
Transcript evidence Protein evidence	<< < >	> 1 to 1 of 1	hits 10 🗸 per page						bound starch synthase I		
Save options Annotation data Tab-delimited file Sequence data Genomic sequence Genomic sequence											
Selection All results v Save											

Known functional alleles for approximately 600 loci extracted from the literature.

How we update gene annotation

Manual Curation through Literature Review by experts

Based on the information reported in the literature, we have updated gene structures, gene functions, etc.



New features are coming soon!

TASUKE+ with over 1,300 rice accessions

We have obtained and analyzed whole genome resequence data of over 1,300 rice accessions from various research groups, making genome-wide polymorphism data accessible on TASUKE+. Additionally, we plan to release other TASUKE+ systems, which allows users to view polymorphism data for NARO rice collection and major cultivars commonly used in Japan.

opaque endosperm; effect on of GBSSI to starch granules

high amylose content

low amylose content

low amylose content

intermediate and high amylose conten effect on viscosity characteristic

Community feedback system for curation

We have implemented a feedback system for users to provide information about literature, functions, and structures of specific genes. If you have any literature or information on gene symbols that are not described in RAP-DB, please give us feedback!!



*fast*Text Natural Language Processing library developed by Google and Meta.

0.20 0.10 0.00 Precisior

Curated gene information from

the research community

Curated gene annotation from research groups can also be a very important source for the updates.

We have been conducting manual curation by referring to several papers so far. The curated gene annotation data

is greatly welcome!!

have developed a system that uses 'AI' to 0.70 0.60 evaluate papers in PubMed and identify 0.50 0.40 those that describe functional information 0.30 on rice genes, which can be 'Useful' for curation. F1 score Recall

Beyond gene ontology (GO): using biocuration approach to improve the gene	the plant journal				
nomenclature and functional annotation	The Plant Journal (2021) 108 , 492–508				
of rice S-domain kinase subfamily	A new comprehensive annotation of leucine-rich				
Sushma Naithani, Daemon Dikeman, Priyanka Garg, Noor Al-Bader and Pankaj Jaiswal	repeat-containing receptors in rice a o				

Céline Gottin^{1,2} 🝈, Anne Dievart^{1,2} 🝈, Marilyne Summo^{1,2} 🍈, Gaëtan Droc^{1,2} 💿, Christophe Périn^{1,2} Vincent Ranwez¹ () and Nathalie Chantret^{1,4} () ¹UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 I e S-domain subfamily of receptor-like kinases (SDRLKs) in plants is poorl CIRAD, UMR AGAP Institut, F-34398 Montpellier, Franc acterized. Most members of this subfamily are currently assigned gene funct

Oryza sativa (rice) plays an essential food security role for more than half of the world's population. Obtain ing crops with high levels of disease resistance is a major challenge for breeders, especially today, given the urgent need for agriculture to be more sustainable. Plant resistance genes are mainly encoded by three

S 🗗 B

Leucine-rich repeat-

containing receptors (LRRs) tural' annotations, we propose that the community adopts this approach, which we call 'comprehensiv annotation. The resulting data are crucial for further studies on the natural variability and evolution CR genes in order to promote their use in breeding future resilient varie

NRAe

Ortholog information

Information on agriculturally useful rice genes can be utilized for the molecular breeding of other crops and plants. We will release ortholog information for the following plant species.



ure sion (TENOR)	bound glycogen synthase, UDPG-gl glucosyltransferase, glycogen [star UDPG-glycogen synthetase, granul granule-bound starch synthase I	RAPDB entry update request
base	Position	
nce	chr06:17664701770653(+ strand	OsID or Transcript ID * Os06t0133000-02
	Locus Os06g0133000	PMID or DOI*
	Other variants	Gene symbol, Gene name
	Gene ontology (GO) Molecular Function: glycogen (sta	Message
	glycosyltrans Functional domain (InterPro	
	Glycosyl transferase, family 1 (IPR) Bacterial/plant glycogen synthase (Starch synthase, catalytic domain (Confirm >

Scan to get: the PDF version of this poster -



Gramene

S-domain subfamily

(SDRLKs)

Additional Information a Declarations can be fou page 24

DOI 10.7717/peerj.11052

of receptor-like kinase and 98 gence and 98

ased on the S-locus Receptor Kinase from Brassica that acts as the female

determinant of self-incompatibility (SI). However, Brassica like SI mechanisms of

ot exist in most plants. Thus, automated Gene Ontology (GO) pipelines are no sufficient for functional annotation of SDRLK subfamily members and lead to s association with the GO biological process of SI. Here, we show that manual bio-curation can help to correct and improve the gene annotations and

ociation with relevant biological processes. Using publicly available genomic a

ativa) SDRLK subfamily, the structure of individual genes and protei

nducted a detailed analysis of the expansion of the rid

-like protein s and RLKs

es, and

gap that exists due to

helping to improve th

in the principles of findabi

Acknowledgment:

This research was partly supported by Cabinet Office, Government of Japan, Moonshot Research and Development Program for Agriculture, Forestry and Fisheries (funding agency: Bio-oriented Technology Research Advancement Institution) grant no. JPJ009237.