Rice Annotation Project Database (RAP-DB) Update 2024

Yoshihiro Kawahara[1], Tomoko Hirozane Kishikawa[1], Xiaohui Wang[1], Ryo Hirata[2], Masahiko Kumagai[1] and Takeshi Itoh[3]

[1]Research Center for Advanced Analysis, NARO, Japan, [2]IMSBIO Co., Ltd., Japan, [3]International College, National Taiwan University, Taiwan

RAP-DB: Rice Annotation Project Database (https://rapdb.dna.affrc.go.jp) has been providing rice genome and gene annotation for nearly two decades since the completion of rice genome sequencing in 2004. Initially, we identified gene models primarily using full-length cDNAs, and assigned functions based on homology to known protein sequences. Recently, we have **improved gene models by incorporating large-scale transcriptome data such as RNA-Seq and IsoSeq**. In addition, we **keep our gene annotation up-to-date through manual curation**, carefully reviewing the latest literatures. For the past five years, we have been **updating our gene annotation twice a year**, providing the latest information on rice genes. RAP-DB has recently begun providing information on gene expression and genome diversity. **Gene expression profiles under 739 cultivation conditions** were available for each gene. Furthermore, **genome-wide variation data for 685 rice varieties** are provided through the multiple genome browser 'TASUKE+.' We have compiled a list of crucial genes related to agronomic traits and provided information on known alleles from literature. This resource is invaluable for efficient breeding using genomic and genetic information. The X/Twitter account (@rapdbjp) also provides the information about latest rice gene publications, updating RAP-DB, etc. Please follow us!

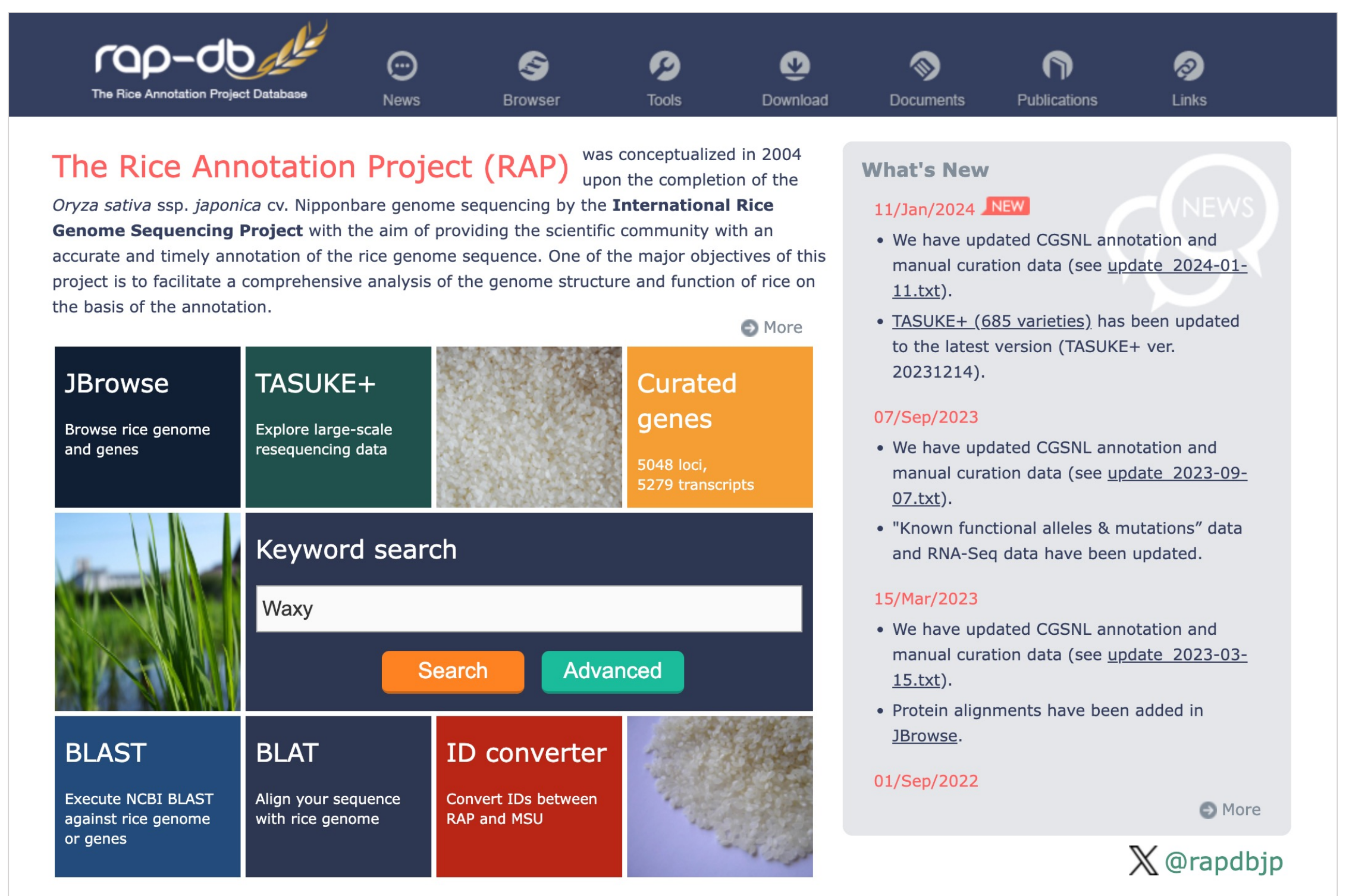
Multiple gene annotation for one reference genome (IRGSP-1.0, Os-Nipponbare-Reference-IRGSP-1.0)

The gene annotations in RAP-DB continue to be updated 'twice a year'.

	RAP-DB	RGAP	RefSeq
Organization	NARO	UGA	NCBI
Release	20240111	Release 7	Release 102
Last upadate	Jan 11, 2024	Feb 7, 2012	Aug 7, 2018
Frequency of updates	twice a year	no updates	unspecified
# of loci	45,896	55,986	35,223
# of transcripts			
Protein-coding	53,075	66,338	53,457
Non-coding	2,354	0	10,987
BUSCO (%, Protein)			
Complete (Single, Dup.)	99.4 (85.6, 13.8)	92.0 (71.1, 20.9)	99.6 (69.1, 30.5)
Fragmented	0.4	4.8	0.1
Missing	0.2	3.2	0.3

Links to the databases:
- **RAP-DB** https://rapdb.dna.affrc.go.jp
- **RGAP** https://rice.uga.edu/
- **RefSeq** https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Oryza_sativa_Japonica_Group/102/

Top page (https://rapdb.dna.affrc.go.jp)



The Rice Annotation Project (RAP) was conceptualized in 2004 upon the completion of the *Oryza sativa* ssp. *japonica* cv. Nipponbare genome sequencing by the **International Rice Genome Sequencing 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.

More

JBrowse Browse rice genome and genes

TASUKE+ Explore large-scale resequencing data

Curated genes 5048 loci, 5279 transcripts

Keyword search Waxy — Search / Advanced

BLAST Execute NCBI BLAST against rice genome or genes

BLAT Align your sequence with rice genome

ID converter Convert IDs between RAP and MSU

What's New

11/Jan/2024 NEW
- We have updated CGSNL annotation and manual curation data (see update_2024-01-11.txt).
- TASUKE+ (685 varieties) has been updated to the latest version (TASUKE+ ver. 20231214).

07/Sep/2023
- We have updated CGSNL annotation and manual curation data (see update_2023-09-07.txt).
- "Known functional alleles & mutations" data and RNA-Seq data have been updated.

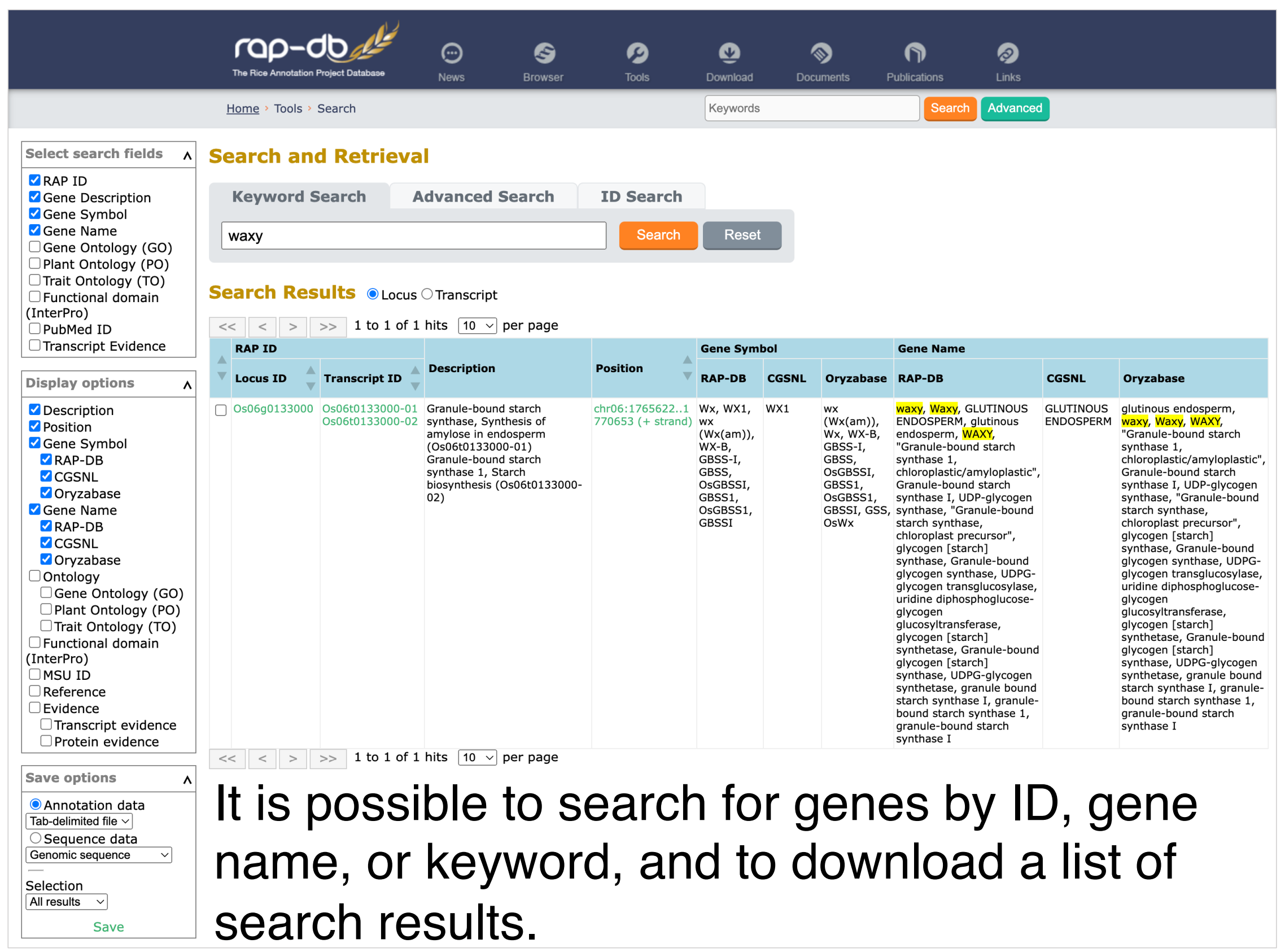
15/Mar/2023
- We have updated CGSNL annotation and manual curation data (see update_2023-03-15.txt).
- Protein alignments have been added in JBrowse.

01/Sep/2022

More

@rapdbjp

Search by ID or keywords (eg. Waxy)



Search and Retrieval

Keyword search: waxy

Select search fields
☑ RAP ID
☑ Gene Description
☑ Gene Symbol
☑ Gene Name
☐ Gene Ontology (GO)
☐ Plant Ontology (PO)
☐ Trait Ontology (TO)
☐ Functional domain (InterPro)
☐ PubMed ID
☐ Transcript Evidence

Display options
☑ Description
☑ Position
☑ Gene Symbol
  ☑ RAP-DB
  ☑ CGSNL
  ☑ Oryzabase
☑ Gene Name
  ☑ RAP-DB
  ☑ CGSNL
  ☑ Oryzabase
☐ Ontology
  ☐ Gene Ontology (GO)
  ☐ Plant Ontology (PO)
  ☐ Trait Ontology (TO)
☐ Functional domain (InterPro)
☐ MSU ID
☐ Reference
☐ Evidence
  ☐ Transcript evidence
  ☐ Protein evidence

Save options
● Annotation data — Tab-delimited file
○ Sequence data — Genomic sequence
Selection — All results

Search Results ● Locus ○ Transcript

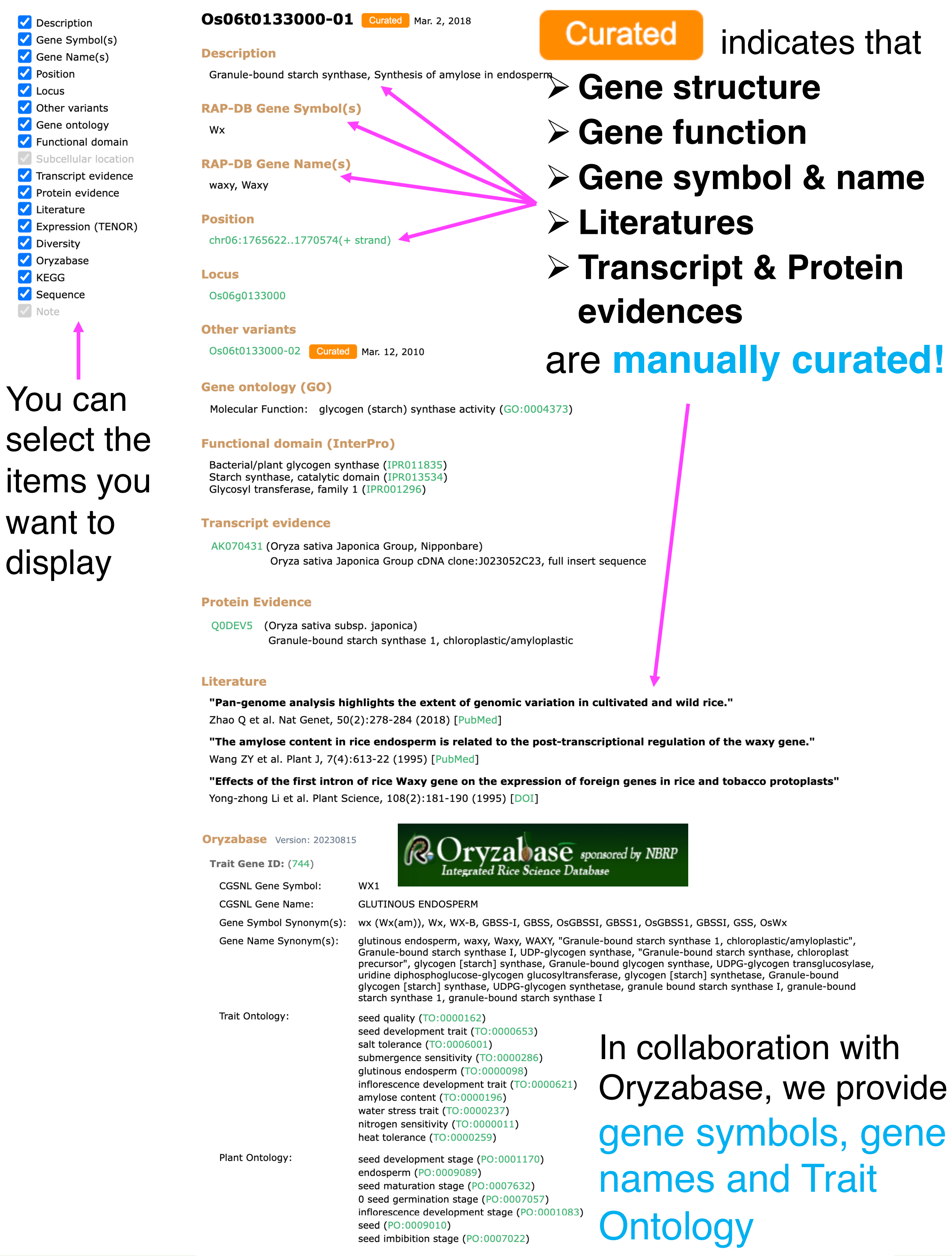
1 to 1 of 1 hits 10 per page

RAP ID		Description	Position	Gene Symbol			Gene Name		
Locus ID	Transcript ID			RAP-DB	CGSNL	Oryzabase	RAP-DB	CGSNL	Oryzabase
Os06g0133000	Os06t0133000-01 Os06t0133000-02	Granule-bound starch synthase, Synthesis of amylose in endosperm (Os06t0133000-01) Granule-bound starch synthase 1, Starch biosynthesis (Os06t0133000-02)	chr06:1765622..1770653 (+ strand)	Wx, WX1, wx (Wx(am)), WX-B, GBSS-I, GBSS, OsGBSSI, GBSS1, OsGBSS1, GBSSI	WX1	wx (Wx(am)), Wx, WX-B, GBSS-I, GBSS, OsGBSSI, GBSS1, OsGBSS1, GBSSI, 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 transglucosylase, uridine diphosphoglucose-glycogen glucosyltransferase, glycogen [starch] synthetase, Granule-bound glycogen [starch] synthase, UDPG-glycogen synthetase, granule bound starch synthase I, granule-bound starch synthase 1, granule-bound starch synthase I	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 synthetase, granule bound starch synthase I, granule-bound starch synthase 1, granule-bound starch synthase I

1 to 1 of 1 hits 10 per page

It is possible to search for genes by ID, gene name, or keyword, and to download a list of search results.

Annotation for individual loci/transcripts
1. Gene symbol/name, function, literatures, etc



☑ Description
☑ Gene Symbol(s)
☑ Gene Name(s)
☑ Position
☑ Locus
☑ Other variants
☑ Gene ontology
☑ Functional domain
☑ Subcellular location
☑ Transcript evidence
☑ Protein evidence
☑ Literature
☑ Expression (TENOR)
☑ Diversity
☑ Oryzabase
☑ KEGG
☑ Sequence
☑ Note

You can select the items you want to display

Os06t0133000-01 Curated Mar. 2, 2018

Description
Granule-bound starch synthase, Synthesis of amylose in endosperm

RAP-DB Gene Symbol(s)
Wx

RAP-DB Gene Name(s)
waxy, Waxy

Position
chr06:1765622..1770574(+ strand)

Locus
Os06g0133000

Other variants
Os06t0133000-02 Curated Mar. 12, 2010

Gene ontology (GO)
Molecular Function: glycogen (starch) synthase activity (GO:0004373)

Functional domain (InterPro)
Bacterial/plant glycogen synthase (IPR011835)
Starch synthase, catalytic domain (IPR013534)
Glycosyl transferase, family 1 (IPR001296)

Transcript evidence
AK070431 (Oryza sativa Japonica Group, Nipponbare)
Oryza sativa Japonica Group cDNA clone:J023052C23, full insert sequence

Protein Evidence
Q0DEV5 (Oryza sativa subsp. japonica)
Granule-bound starch synthase 1, chloroplastic/amyloplastic

Literature

"Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice."
Zhao Q et al. Nat Genet, 50(2):278-284 (2018) [PubMed]

"The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene."
Wang ZY et al. Plant J, 7(4):613-22 (1995) [PubMed]

"Effects of the first intron of rice Waxy gene on the expression of foreign genes in rice and tobacco protoplasts"
Yong-zhong Li et al. Plant Science, 108(2):181-190 (1995) [DOI]

Oryzabase Version: 20230815

Oryzabase sponsored by NBRP
Integrated Rice Science Database

Trait Gene ID: (744)

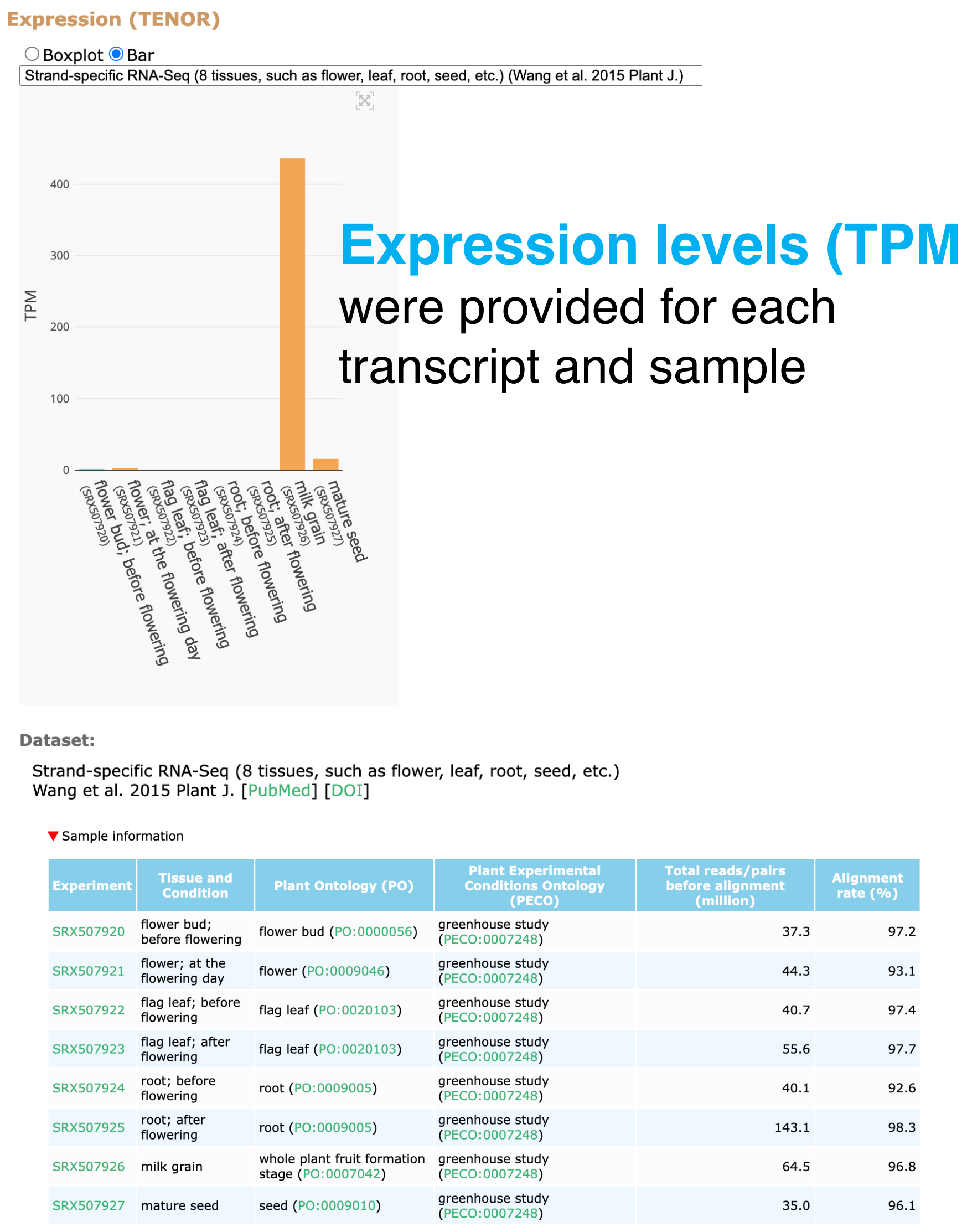
CGSNL Gene Symbol:	WX1
CGSNL Gene Name:	GLUTINOUS ENDOSPERM
Gene Symbol Synonym(s):	wx (Wx(am)), Wx, WX-B, GBSS-I, GBSS, OsGBSSI, GBSS1, OsGBSS1, GBSSI, GSS, OsWx
Gene Name Synonym(s):	glutinous endosperm, waxy, Waxy, WAXY, "Granule-bound starch synthase 1, chloroplastic/amyloplastic", Granule-bound 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 synthetase, granule bound starch synthase I, granule-bound starch synthase 1, granule-bound starch synthase I
Trait Ontology:	seed quality (TO:0000162) seed development trait (TO:0000653) salt tolerance (TO:0006001) submergence sensitivity (TO:0000286) glutinous endosperm (TO:0000098) inflorescence development trait (TO:0000621) amylose content (TO:0000196) water stress trait (TO:0000237) nitrogen sensitivity (TO:0000011) heat tolerance (TO:0000259)
Plant Ontology:	seed development stage (PO:0001170) endosperm (PO:0009089) seed maturation stage (PO:0007632) 0 seed germination stage (PO:0007057) inflorescence development stage (PO:0001083) seed (PO:0009010) seed imbibition stage (PO:0007022)

Curated indicates that
- Gene structure
- Gene function
- Gene symbol & name
- Literatures
- Transcript & Protein evidences

are **manually curated!**

In collaboration with Oryzabase, we provide **gene symbols, gene names and Trait Ontology**

Annotation for individual loci/transcripts
2. Gene expression profiles under 739 cultivation conditions



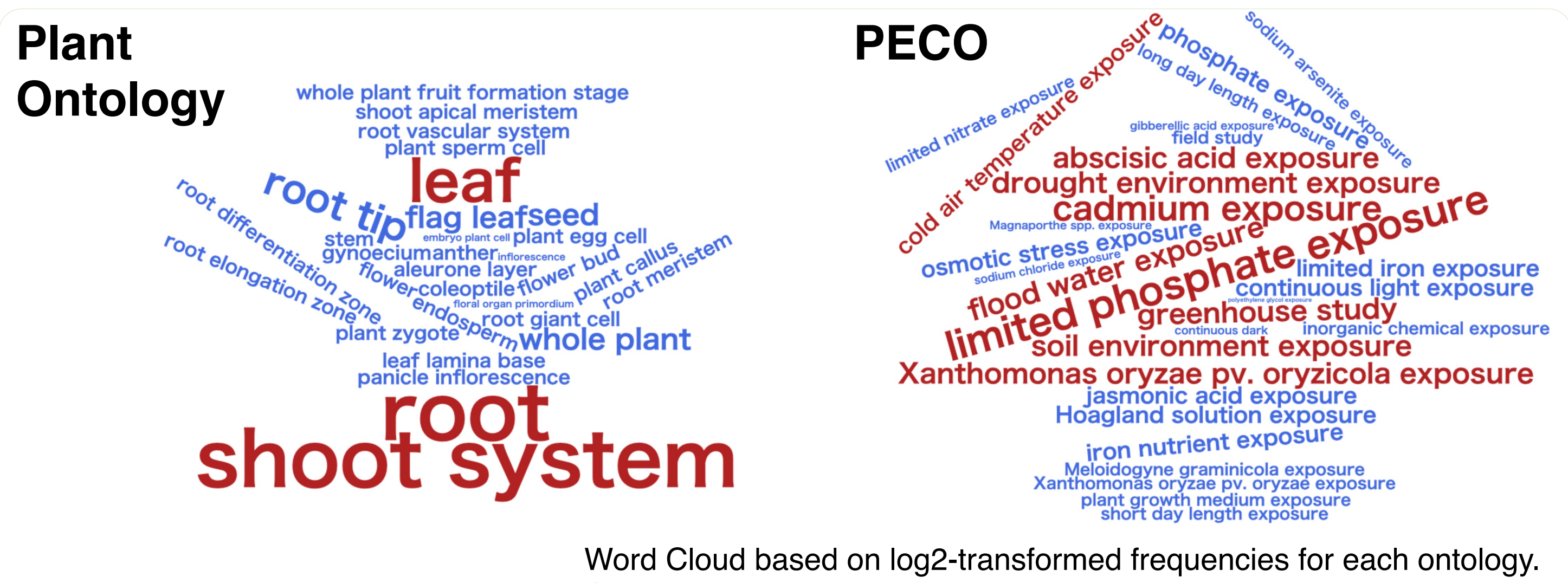
Expression levels (TPM) were provided for each transcript and sample

Dataset:
Strand-specific RNA-Seq (8 tissues, such as flower, leaf, root, seed, etc.)
Wang et al. 2015 Plant J. [PubMed] [DOI]

▼ Sample information

Experiment	Tissue and Condition	Plant Ontology (PO)	Plant Experimental Conditions Ontology (PECO)	Total reads/pairs before alignment (million)	Alignment rate (%)
SRX507920	flower bud; before flowering	flower bud (PO:0000056)	greenhouse study (PECO:0007248)	37.3	97.2
SRX507921	flower; at the flowering day	flower (PO:0009046)	greenhouse study (PECO:0007248)	44.3	93.1
SRX507922	flag leaf; before flowering	flag leaf (PO:0020103)	greenhouse study (PECO:0007248)	40.7	97.4
SRX507923	flag leaf; after flowering	flag leaf (PO:0020103)	greenhouse study (PECO:0007248)	55.6	97.7
SRX507924	root; before flowering	root (PO:0009005)	greenhouse study (PECO:0007248)	40.1	92.6
SRX507925	root; after flowering	root (PO:0009005)	greenhouse study (PECO:0007248)	143.1	98.3
SRX507926	milk grain	whole plant fruit formation stage (PO:0007042)	greenhouse study (PECO:0007248)	64.5	96.8
SRX507927	mature seed	seed (PO:0009010)	greenhouse study (PECO:0007248)	35.0	96.1

Plant Ontology & Plant Environmental Condition Ontology (PECO) assigned to **739 RNA-Seq samples**

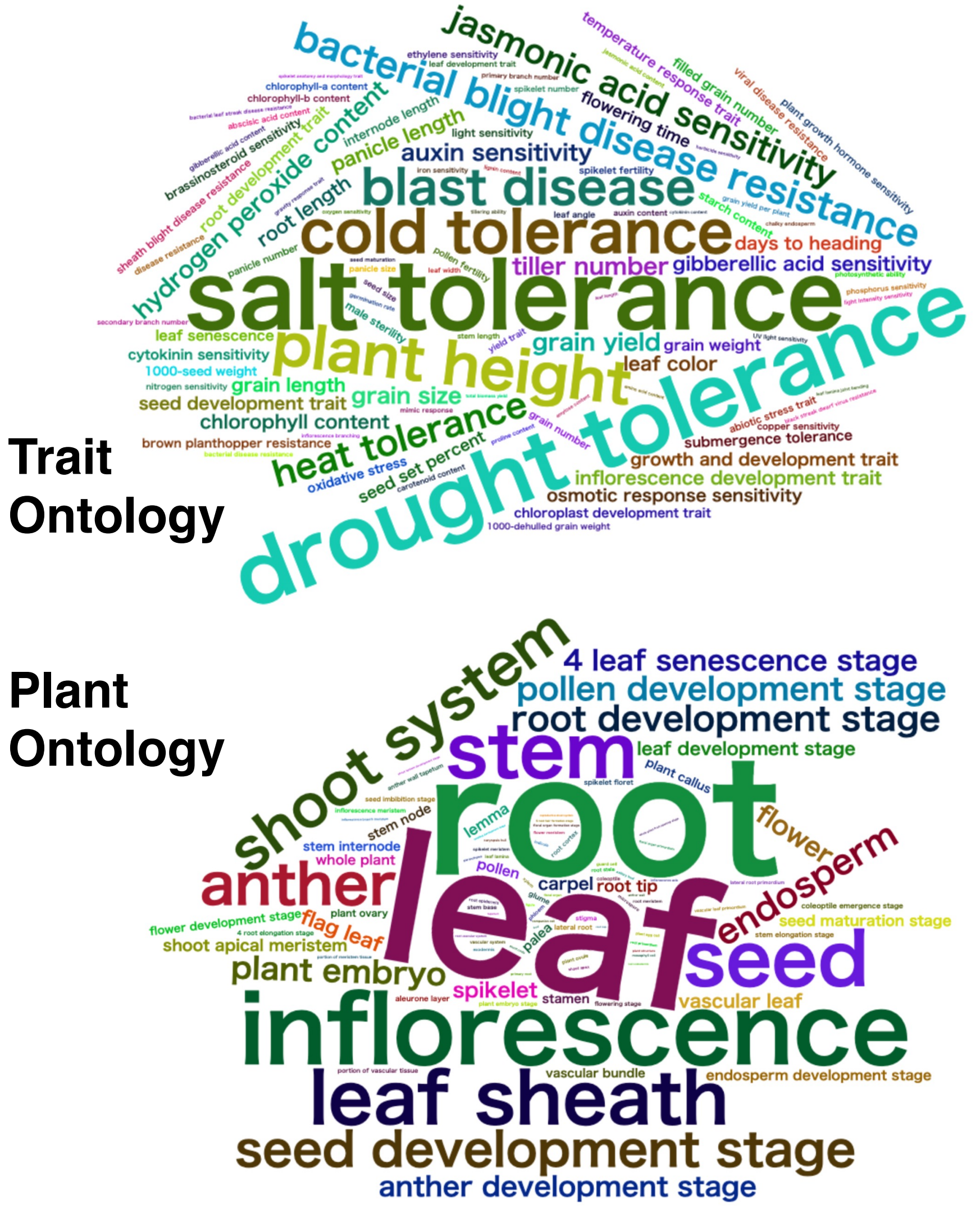


Word Cloud based on log2-transformed frequencies for each ontology. Ontologies that have a frequency of 5 or more log2 are shown in red.

Extraction of knowledge information from literature

Manual curation by experts
Based on the information reported in the literature, we are modifying the gene structures and adding functional and allele information. **>5,000 genes have been manually curated!!**

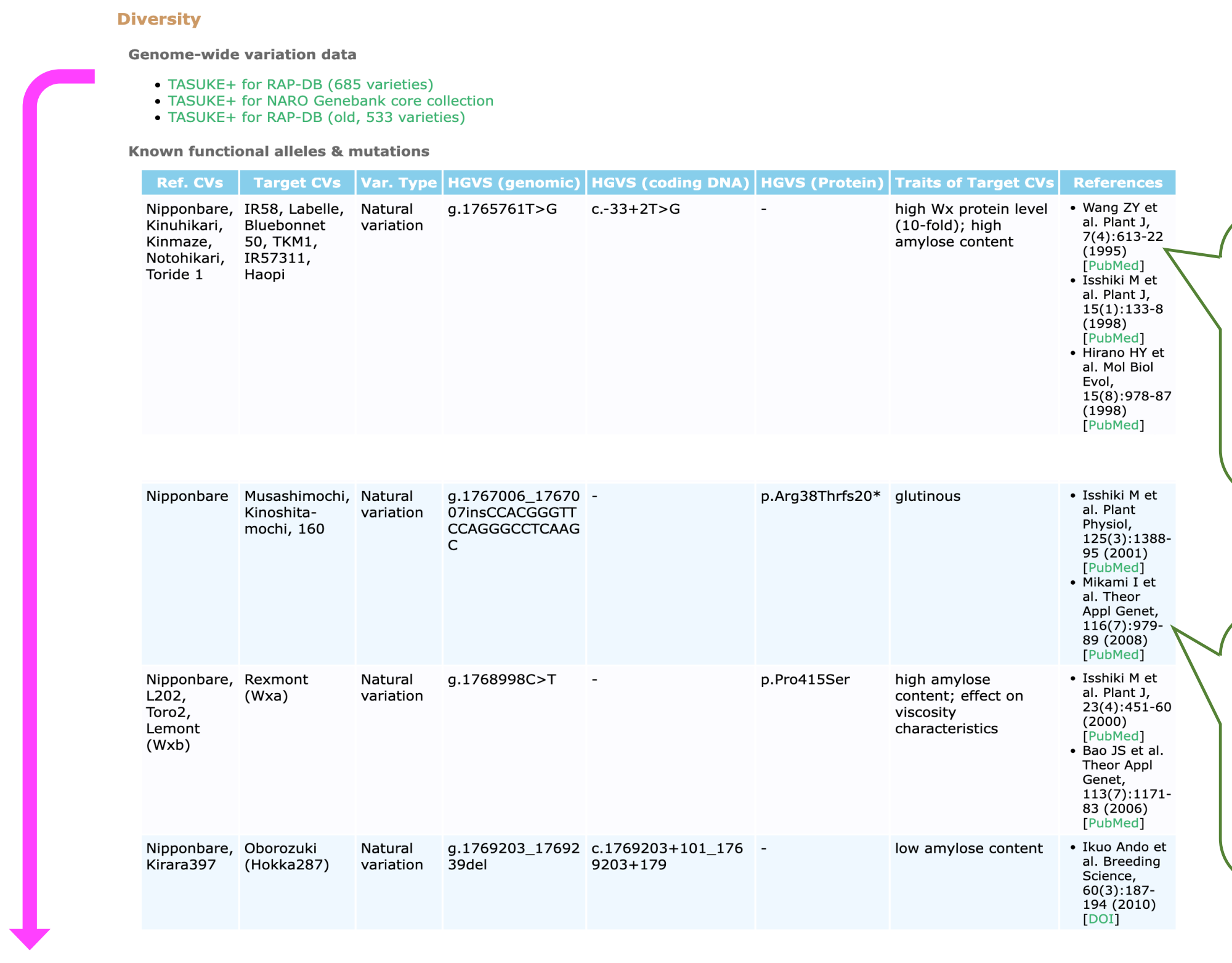
Frequently observed Trait Ontology and Plant Ontology terms for **manually curated genes**.



Annotation for individual loci/transcripts
3. Genome diversity and know functional alleles

Known allele information is extracted from literature.

Functional alleles in about **600 loci** are provided.



Diversity

Genome-wide variation data
- TASUKE+ for RAP-DB (685 varieties)
- TASUKE+ for NARO Genebank core collection
- TASUKE+ for RAP-DB (old, 533 varieties)

Known functional alleles & mutations

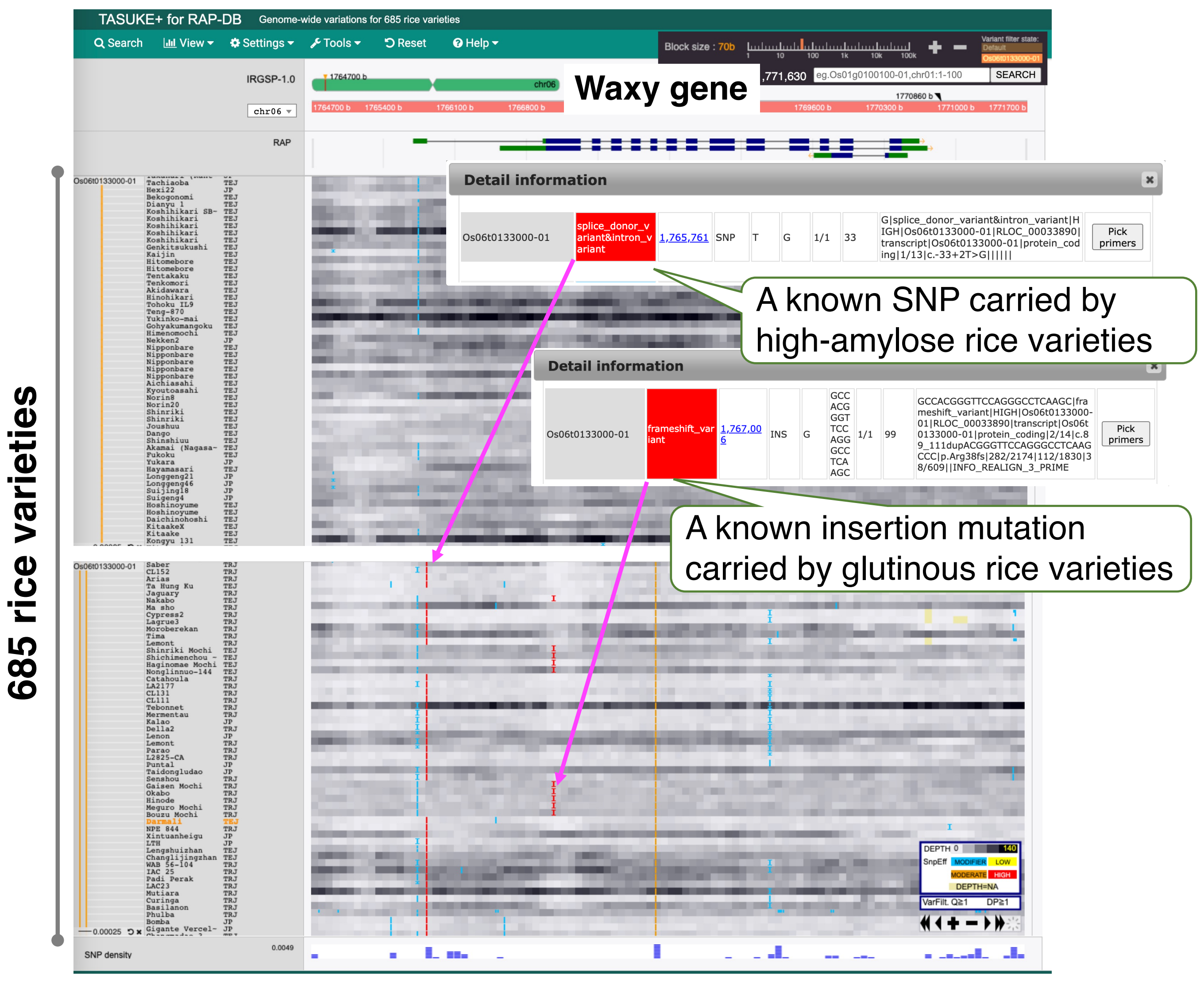
Ref. CVs	Target CVs	Var. Type	HGVS (genomic)	HGVS (coding DNA)	HGVS (Protein)	Traits of Target CVs	References
Nipponbare, Kinuhikari, Kinmaze, Notohikari, Toride 1	IR58, Labelle, Bluebonnet 50, TKM1, IR57311, Haopi	Natural variation	g.1765761T>G	c.-33+2T>G	-	high Wx protein level (10-fold); high amylose content	Wang ZY et al. Plant J, 7(4):613-22 (1995) [PubMed] Isshiki M et al. Plant J, 15(1):133-8 (1998) [PubMed] Hirano HY et al. Mol Biol Evol, 15(8):978-87 (1998) [PubMed]
Nipponbare	Musashimochi, Kinoshita-mochi, 160	Natural variation	g.1767006_1767007insCCACGGGTTCCAGGGCCTCAAGC	-	p.Arg38Thrfs20*	glutinous	Isshiki M et al. Plant Physiol, 125(3):1388-95 (2001) [PubMed] Mikami I et al. Theor Appl Genet, 116(7):979-89 (2008) [PubMed]
Nipponbare, L202, Toro2, Lemont (Wxb)	Rexmont (Wxa)	Natural variation	g.1768998C>T	-	p.Pro415Ser	high amylose content; effect on viscosity characteristics	Isshiki M et al. Plant J, 23(4):451-60 (2000) [PubMed] Bao JS et al. Theor Appl Genet, 113(7):1171-83 (2006) [PubMed]
Nipponbare, Kirara397	Oborozuki (Hokka287)	Natural variation	g.1769203_1769239del	c.1769203+101_1769203+179	-	low amylose content	Ikuo Ando et al. Breeding Science, 60(3):187-194 (2010) [DOI]

A SNP carried by high-amylose rice varieties reported in the literatures

A known insertion mutation carried by glutinous rice varieties

TASUKE+:
Genome-wide variations for 685 rice varieties

Genome-wide variant information in any genomic region can be viewed for **685 rice varieties** for which genome resequencing data are publicly available.



A known SNP carried by high-amylose rice varieties

A known insertion mutation carried by glutinous rice varieties

For more information about TASUKE+, go to poster **PO0675 TASUKE+: A Web-Based Platform for Visualizing Large-Scale Resequencing Data and GWAS**

Contact: rapdb@ml.affrc.go.jp
X/Twitter: @rapdbjp

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.