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

Yoshihiro Kawahara[1], Tomoko Hirozane-Kishikawa[1], Ryo Hirata[2], Yuki Tamagaki[3], Yumiko Teramoto[2], Xiaohui Wang[1], Norio Tabei[4], Masahiko Kumagai[1], Hiroaki Sakai[1], Takeshi Itoh[5]

[1]Research Center for Advanced Analysis, NARO, Japan, [2]IMSBIO Co., Ltd., Japan, [3]Mizuho Research & Technologies, Ltd., [4]DYNACOM Co., Ltd., [5]International College, National Taiwan University, Taiwan

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.**

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

Functional annotation of rice genes in RAP-DB

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

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)
glycosyltransferase activity (GO:0016757)

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

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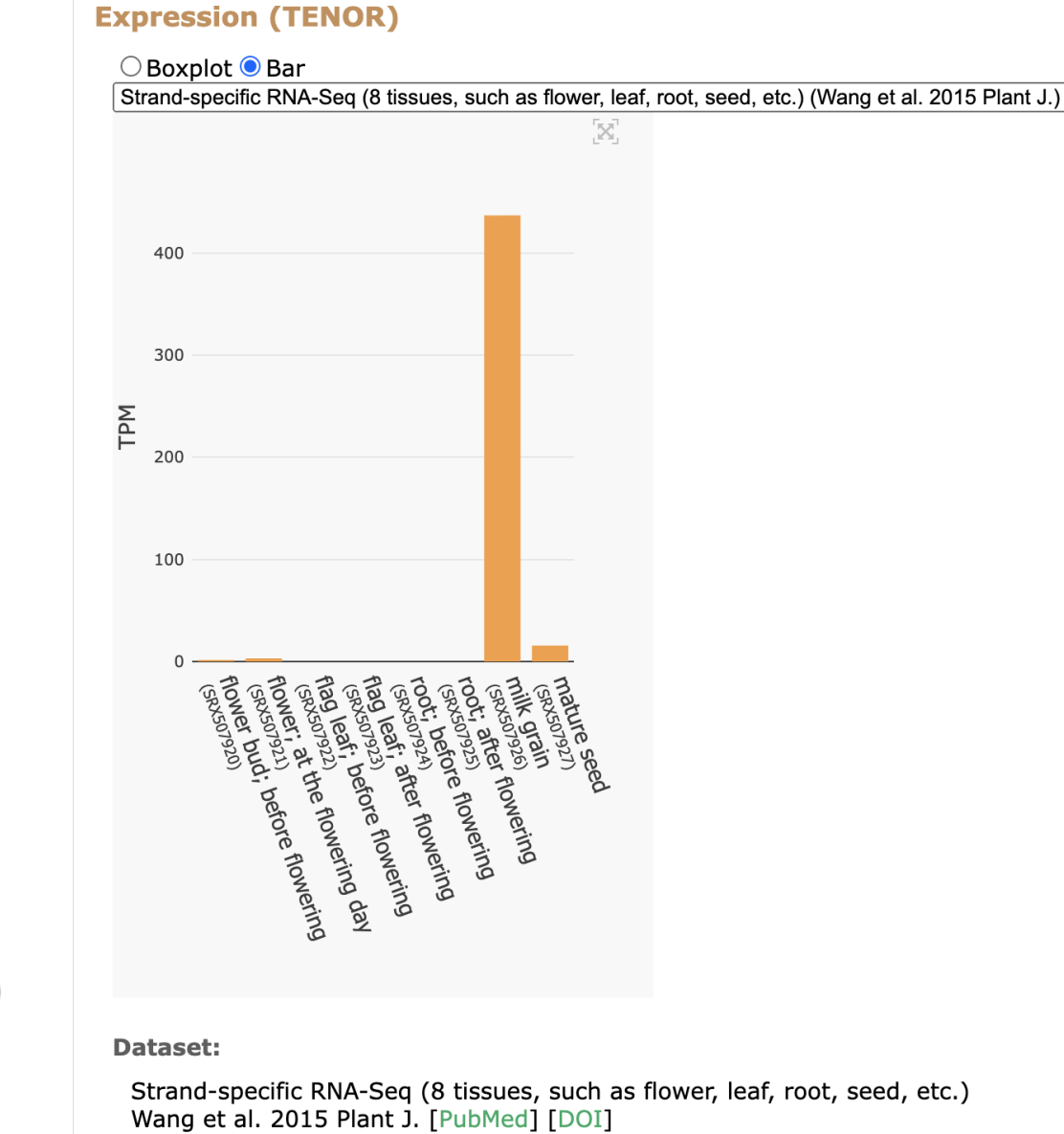Oryza sativa Japonica Group cDNA clone:J023052C23, full insert sequence

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."
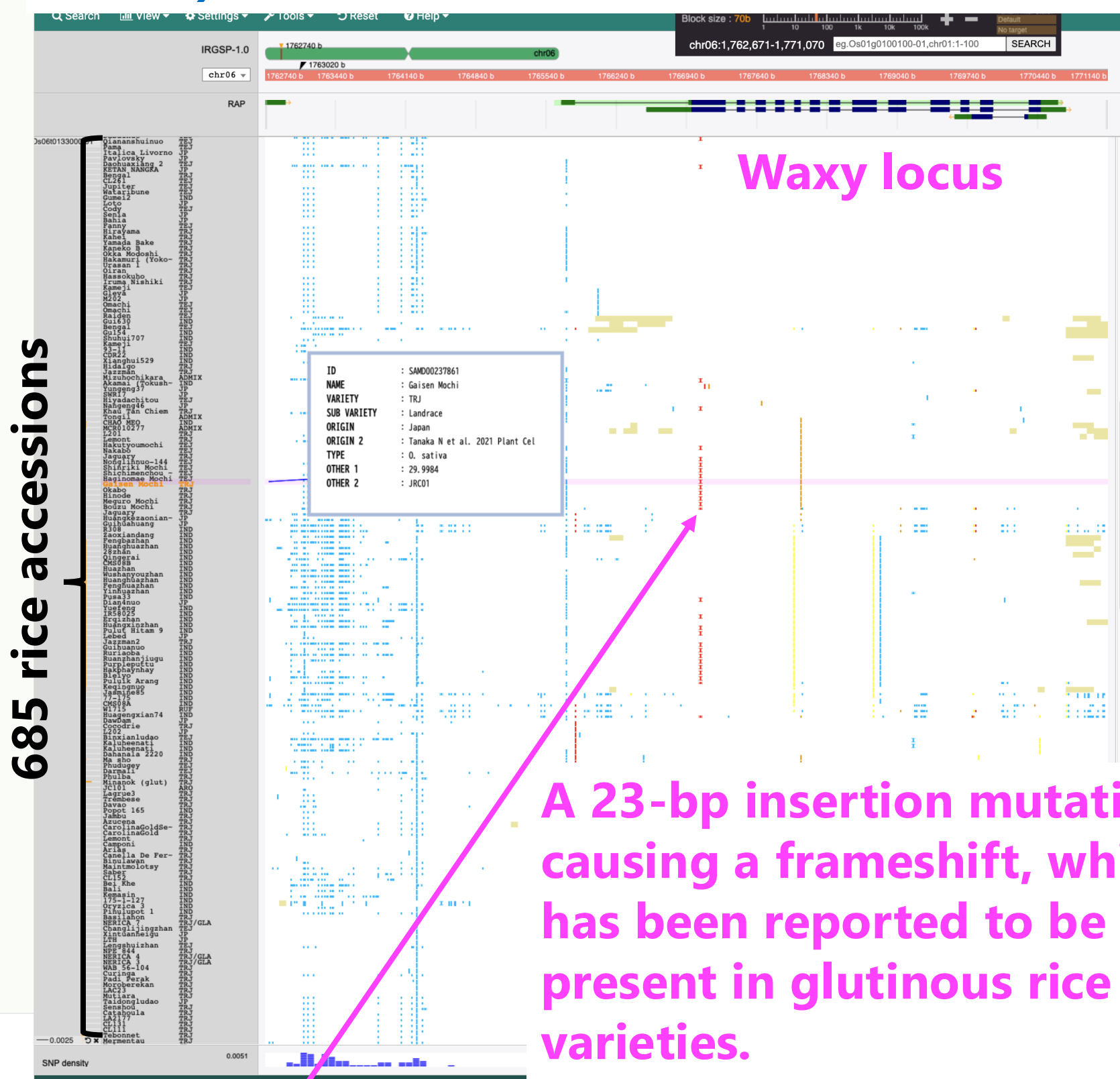Zhao Q et al. Nat Genet, 50(2):278-284 (2018) [PubMed]

"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]

Expression profile under more than 700 experimental conditions

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

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

Waxy locus

685 rice accessions

A 23-bp insertion mutation causing a frameshift, which has been reported to be present in glutinous rice varieties.

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.
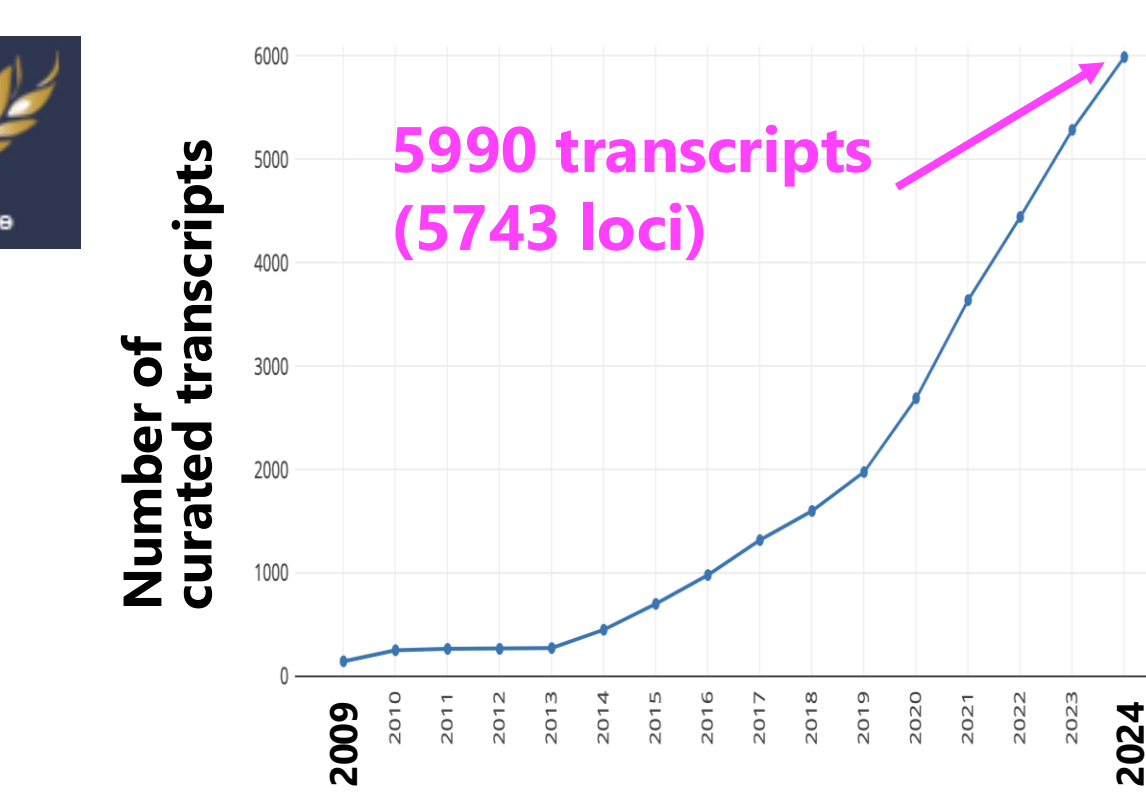

5990 transcripts (5743 loci)

Number of curated transcripts

AI model performance evaluation with newly published 216 papers

fastText / BERT

Natural Language Processing library developed by Google and Meta.

To improve the efficiency of curation, we have developed a system that uses 'AI' to evaluate papers in PubMed and identify those that describe functional information on rice genes, which can be 'Useful' for curation.

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!!

S-domain subfamily of receptor-like kinase (SDRLKs)

Leucine-rich repeat-containing receptors (LRRs)

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.

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!!

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.

Tree scale: 0.1

C. quinoa (High-protein superfood)
V. marina (Salt-tolerant wild azuki bean)
G. max (Soybean)
A. thaliana
Z. mays (Maize)
H. vulgare (Barley)
T. aestivum (Wheat)
O. sativa (Rice)