*Database*, 2025, baaf021
DOI: https://doi.org/10.1093/database/baaf021
**Original article**

# GrameneOryza: a comprehensive resource for *Oryza* genomes, genetic variation, and functional data

**Sharon Wei[1], Kapeel Chougule[1], Andrew Olson[1], Zhenyuan Lu[1], Marcela K. Tello-Ruiz[1], Vivek Kumar[1], Sunita Kumari[1], Lifang Zhang[1], Audra Olson[1], Catherine Kim[1], Nick Gladman[1,2], Doreen Ware[1,2,*]**

[1]Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, United States
[2]USDA ARS NEA, Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, 538 Tower Road, Ithaca, NY 14853-2901, United States

[*]Corresponding author. Cold Spring Harbor Laboratory, Williams #5, 1 Bungtown Road, Cold Spring Harbor, NY 11724, United States.
E-mail: Doreen.ware@usda.gov

Citation details: Wei, S., Chougule, K., Olson, A. *et al.* GrameneOryza: a comprehensive resource for *Oryza* genomes, genetic variation, and functional data. *Database* (2025) Vol. 2025: article ID baaf021; DOI: https://doi.org/10.1093/database/baaf021

## Abstract

Rice is a vital staple crop, sustaining over half of the global population, and is a key model for genetic research. To support the growing need for comprehensive and accessible rice genomic data, GrameneOryza (https://oryza.gramene.org) was developed as an online resource adhering to FAIR (Findable, Accessible, Interoperable, and Reusable) principles of data management. It distinguishes itself through its comprehensive multispecies focus, encompassing a wide variety of *Oryza* genomes and related species, and its integration with FAIR principles to ensure data accessibility and usability. It offers a community curated selection of high-quality *Oryza* genomes, genetic variation, gene function, and trait data. The latest release, version 8, includes 28 *Oryza* genomes, covering wild rice and domesticated cultivars. These genomes, along with *Leersia perrieri* and seven additional outgroup species, form the basis for 38 K protein-coding gene family trees, essential for identifying orthologs, paralogs, and developing pan-gene sets. GrameneOryza's genetic variation data features 66 million single-nucleotide variants (SNVs) anchored to the Os-Nipponbare-Reference-IRGSP-1.0 genome, derived from various studies, including the Rice Genome 3 K (RG3K) project. The RG3K sequence reads were also mapped to seven additional platinum-quality Asian rice genomes, resulting in 19 million SNVs for each genome, significantly expanding the coverage of genetic variation beyond the Nipponbare reference. Of the 66 million SNVs on IRGSP-1.0, 27 million acquired standardized reference SNP cluster identifiers (rsIDs) from the European Variation Archive release v5. Additionally, 1200 distinct phenotypes provide a comprehensive overview of quantitative trait loci (QTL) features. The newly introduced *Oryza* CLIMtools portal offers insights into environmental impacts on genome adaptation. The platform's integrated search interface, along with a BLAST server and curation tools, facilitates user access to genomic, phylogenetic, gene function, and QTL data, supporting broad research applications.

**Database URL**: https://oryza.gramene.org

## Introduction

Rice (*Oryza* spp.) is an essential staple crop worldwide, providing crucial nutrients such as carbohydrates, protein, and vitamins that supports more than half of the global population [1, 2]. It also serves as an important model organism for genetic research and breeding due to its relatively small genome size of 430 Mb [3], which is significantly smaller than that of maize and wheat. This small genome size makes rice a cost-effective option for genome sequencing, assembly, and analysis. The genome Os-Nipponbare-Reference-IRGSP-1.0 [3] of *Oryza sativa Japonica*, the first crop plant to undergo genome sequencing, has become the reference genome assembly for the *Oryza* genus and its varieties [3, 4]. The genetic diversity within domesticated rice and its wild relatives offers valuable tools for addressing the impacts of climate change and promoting sustainable agriculture [5, 6].

Plant genomes exhibit greater fluidity compared to vertebrate genomes, resulting in substantial genomic disparities even among varietal lines within the same species. For example, a study comparing seven wild soybean germplasms (*Glycine soja* Sieb. & *Zuccit*) found that approximately 80% of the genomes were shared, with an average genomic difference of about 20% among the accessions [7]. A similar phenomenon has been observed in maize, where a study analyzing the genomes of 26 diverse inbred lines revealed that more than half of the genes present in one genome may be absent in another, leading to extensive structural variations [8]. These differences include small single-nucleotide

Received 11 October 2024; Revised 29 January 2025

polymorphisms (SNPs), insertions and deletions (Indels), and large structural variations such as present–absent variation (PAV), copy number variation (CNV), inversions, transpositions, and differences in transposable elements (TEs) [9, 10]. These findings highlight the limitations of a single reference genome in capturing all functional elements and diversity within a species. Following the publication of the Nipponbare genome, numerous *Oryza* genomes have been sequenced and analyzed for various objectives using different sequencing technologies and assembly methodologies [5, 11–15].

Despite the abundance of rice genomes, the lack of a centralized platform has hindered access and comparison. To address this gap, we developed the GrameneOryza resource, a comprehensive comparative *Oryza* pan-genome database designed to adhere to the FAIR principles [16]. Utilizing the Ensembl platform [17–19] and Gramene infrastructure [20–22], this online resource incorporates unique features and functions from both Ensembl and Gramene software. Initially launched to host the rice genome using state-of-the-art Ensembl software, Gramene has evolved into a comprehensive plant genome browser [4, 21]. Its goal was to provide stewardship for the rice genome using an open-source web-based platform and facilitate knowledge transfer from well-studied model organisms like *Arabidopsis* [23] to rice. Building on the success of this initiative, Gramene has progressively expanded to integrate knowledge from various other databases. This includes curated pathways from Plant Reactome [20, 24], gene expression data from the EBI Atlas [25, 26], and genetic variation and gene function information from the Rice Annotation Project Database (RAP-DB) [4]. Additionally, it incorporates data from the 3000 Rice Genomes Project [27] and QTL information from Rice Qtaro [28] and Gramene's historic QTL database [29, 30]. The platform also integrates community tools like CLIMtools [31], which provide genotype-by-environment association data, and gene expression data across different tissues and developmental stages through EBI Expression Atlas widget. Gramene also developed a comprehensive and powerful gene search interface, henceforth referred to as the Gramene Search Interface (GSI). The GSI enables efficient navigation and exploration of genomic data, integrating multiple resources to provide detailed insights into gene functions, variations, and relationships. In collaboration with EnsemblGenomes, Gramene Plants adopted the rule of hosting only the International Nucleotide Sequence Database Collaboration (INSDC) assemblies, which improves genome reliability but delays the release of some emerging rice genomes yet to be deposited. To adapt to this requirement and better serve the rice community, Gramene launched in 2015 a dedicated subsite, GrameneOryza (https://oryza.gramene.org/), to host rice genomes and related data.

The current GrameneOryza pan-genome resource includes representatives of domesticated Asian and African rice, their progenitors, and wild rice relatives. This resource hosts a set of high-quality reference *Oryza* genomes presented in a phylogenetic framework, with uniform annotation of gene structure, repetitive features, and protein-coding gene phylogenetic trees built using the Ensembl protein compara pipeline. The gene tree analysis includes eight outgroup genomes spanning the plant kingdom and eukaryotic representatives (*Chlamydomonas reinhardtii, Selaginella moellendorffii, Sorghum BTx623, Zea mays B73, L. perrieri, Arabidopsis thaliana, Vitis vinifera, Drosophila melanogaster*) and provides insights into the evolutionary history within rice, plants, and eukaryotes. Special genomes of historical and cultural significance in the USA, such as cultivated Carolina Gold rice [14] and KitaakeX [13], a short-life cycle rice cultivar carrying the *Xa21* immune receptor gene, are also included. The latest release, Release 8, includes 27 million reference SNP cluster identifiers (rsIDs) assigned by European Variation Archive (EVA) to the variation database of the reference genome *O. sativa japonica* Nipponbare IRGSP-1.0, making it the only pan-*Oryza* genome resource with this critical information. Release 8 GrameneOryza also engages users in community curation by providing an interface for user suggestions on gene structure improvements and gene function annotations At the sametime, it serves as the only gateway for the *Oryza* CLIMtools, a set of interactive, web-based databases of the environment × genome associations and correlations between the local environment and a large pool of curated *Oryza* genotypes. Recent data ingestions include genetic variation and population data, for example: USDA minicore [32] and RAP-DB World Rice Collection (WRC) [33] and Japanese Rice Collection (JRC) [34] collection. A new function in the GSI is the Germplasm tab, which lists the potential mutant lines for the gene and directs the user to the germplasm center webpage for the germplasm when possible. These features distinguish GrameneOryza from other pan-*Oryza* resources, making it a unique and valuable resource for the rice community.

## Materials and methods

### GrameneOryza platform

The core functionality of the GrameneOryza resource is depicted in Fig. 1, which provides a systems-level view of the platform's architecture. The process begins with the input data (Fig. 1a), encompassing genome assemblies, gene structures, variation data, and quantitative trait loci (QTLs). These data are then incorporated into various databases (Fig. 1b), including the core database and specialized databases for search functionalities. The analysis workflow (Fig. 1c) involves in-house pipelines for generating structural and functional annotations, complemented by Ensembl workflows for constructing gene trees and predicting variant effects. Web services (Fig. 1d) facilitate programmatic access to the databases, supporting both user interfaces and custom scripts. The user interfaces (panel e) offer landing pages that enable users to interact with the data through powerful search tools and integrated views. The result pages consolidate diverse data types into a single panel organized under different tabs, enhancing the accessibility and usability of the GrameneOryza resource. This integrated approach ensures comprehensive and user-friendly access to the extensive genomic, phylogenetic, gene function, and QTL data hosted by GrameneOryza.

### Database design

GrameneOryza was crafted to serve as a repository for *Oryza* species and wild relatives, capitalizing on the Gramene search engine, dynamic interfaces, Ensembl workflows, and genome browser. The current iteration, Release 8 (Fig. 1), operates on the Ensembl v108 platform at its foundational layer, encompassing databases, API, analysis pipeline, and web code. As depicted in Fig. 2, the Gramene outer layer extracts data from Ensembl MySQL [19] databases and combines it with information from third-party resources. The fusion generates

collections of genes, synonyms, genetrees, pathways, domains, genetic variations, germplasms, and ontologies stored in document database, MongoDB (https://www.mongodb.com), facilitating integrated search and views. The gene collections are subsequently indexed by Solr (https://solr.apache.org), to streamline tailored searches with type-ahead suggestions. GrameneOryza also instituted a site-specific BLAST service [35], enabling sequence-based queries against all the genomes, transcriptomes, and proteomes. The Ensembl genome browser and GSI seamlessly cross-link at multiple levels, ensuring a fluid connection between these two components.

## Data resources

GrameneOryza knowledgebase hosts a comprehensive repository of 28 whole genome assemblies of *Oryza* accessions, representing both domesticated and wild rice varieties. This diverse collection features significant cultivars such as the US heirloom Carolina Gold [14] and the short-life cycle cultivar KitaakeX [13]. Additionally, the database incorporates *L. perrieri* [5], the nearest outgroup to the genus *Oryza*, providing valuable phylogenetic insights. Extending its comparative framework, GrameneOryza also includes seven additional outgroup species from across the plant kingdom and eukaryotic representatives, namely *C. reinhardtii* (green algae) [36], *S. moellendorffii* (clubmoss) [37], *Sorghum bicolor BTx623* (sorghum) [38], *Zea mays B73* (maize) [39], *A. thaliana* (model plant) [40], *V. vinifera* (grapevine) [41], and *D. melanogaster* (fruit fly) [42]. These genomes aid in constructing comprehensive phylogenetic gene trees [18], facilitating the identification of orthologs and paralogs, and providing a broader evolutionary context for the *Oryza* genomes.

GrameneOryza exemplifies a robust integration of diverse data types, including genomic, phylogenetic, gene function, genetic variation, germplasm, and QTL, facilitated by the GSI. This platform hosts single-nucleotide variants (SNVs) data from the RG3K project [27] aligned to reference genome Os-Nipponbare-Reference-IRGSP-1.0, and seven agronomically important platinum-quality Asian rice genomes: MH63, IR64, Aus, Azucena, ARC, ZS97, LiuXu, identifying approximately 19 million SNVs per genome [11]. It also hosts SNVs from USDA minicore and RAP-DB WRC/JRC collection. Furthermore, 27 million SNVs on IRGSP-1.0 are updated with rsIDs from EVA Release v5 (https://www.ebi.ac.uk/eva). All variants in the database are annotated with predicted consequences in Sequence Ontology term names [43], such as "start_lost" and "stop_gained," contextualized to nearby genes using the Ensembl Variant Effect Predictor [44]. The platform also hosts germplasm names of the RG3K, USDA minicore, and RAP-DB WRC/JRC populations alongside genotypes for each SNP and 1141 distinct phenotypes derived from 40 000 QTL regions on *O. sativa japonica* Nipponbare [29, 30], incorporating insights from published research papers and cross-species flanking marker sequence-based QTL mapping. The platform integrates over 5707 curated gene function annotations imported from RAP-DB [4] and NCBI geneRIF [4, 45], enhancing the depth of gene function data. GrameneOryza supports extensive phylogenetic analyses using protein gene trees and whole-genome DNA comparisons, featuring more than 38 000 protein-based gene trees with outgroups tracing back to the root of Eukarya using the compara workflow [18]. This framework supports paralog and ortholog assignments, aiding in the characterization of candidate split genes for gene model improvements. Synteny assignments between *O. sativa japonica* Nipponbare and all other *Oryza* genomes are based on collinear orthologs inferred from gene trees, while whole-genome alignments between *O. sativa japonica* Nipponbare, *Zea mays* B73, and *S. bicolor* BTx623 provide insights into conserved functional elements and synteny assignments.

**Table 1.** Data resources in GrameneOryza

| Data type | Count |
|---|---|
| Genomes | 28 *Oryza* genomes; 9 outgroup genomes from 8 species |
| Gene annotations | 1.2 million genes from 36 genomes |
| Gene trees | 38 000 protein coding gene trees |
| Genetic variation (SNP, rsID, QTL) | 66 million SNVs, 27 million rsIDs, 40 000 QTL regions |
| Gene function | 5700 genes with functions from RAP-DB and GeneRIF |
| Expression (base line, differential, single cell) | 15 baseline, 95 differential, 3 single cell studies |
| Pathways | 339 curated pathways |
| Synteny maps | 27 *Oryza* synteny maps |
| Whole genome alignments | 2 sets: IRGSPv1.0 vs BTx623 (Sorghum) and B73 (Maize) |
| Genotype X Environment (CLIMtools) | 413 geoenvironmental variables on 941 landraces |

The platform's integrated search and visualization tools allow keyword-based searches that return results from genome, gene, gene tree, pathway, variation, and literature databases. Community curation tools are provided for Gramene gene structure based on gene tree views, multiple protein alignments [46]. Additionally, the platform enables users to curate gene functions, allowing researchers to contribute valuable annotations that improve the understanding of gene roles across various species. This comprehensive integration enables intuitive access to multifaceted data, empowering researchers to explore and correlate genetic variations, gene functions, and environmental influences within a unified framework. Table 1 summarizes the data resources in GrameneOryza.

# Results

## User interface

The GrameneOryza home page follows a similar design as the Gramene and other Gramene pan-site home pages. The platform offers textual query search (Fig. 3a) where users can primarily search with specific gene identifiers (e.g. Os01g0100100), gene symbols (e.g *GS3*), specific biological pathways or ontological terms, such as those from Gene Ontology (GO) [47] or Plant Ontology (PO) [48] or phenotypic traits [49] or related terms, such as "drought tolerance," "yield," or "disease resistance" etc. The search index is designed to return genes [20]. Quick links in Fig. 3b include "Genome Browser," which directs users to the Ensembl browser with access to annotation, variation, and comparative gene tree views for the 28 rice genomes; "News," which provides detailed database release notes; and "Guides," which offers a step-by-step guide on using GrameneOryza with practical use cases. Additionally, there is a link for providing "Feedback." Access to additional various tools and

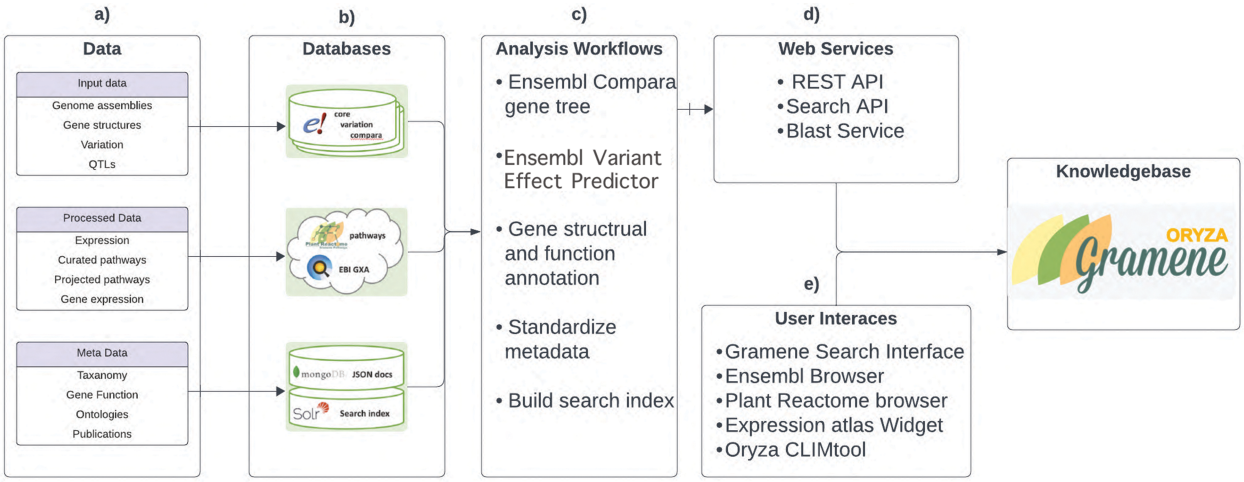


**Figure 1.** A systems level view of the core functionality of the GrameneOryza site. (a) Input data: omics data to load to the core database, this includes processed data as well as metadata, (b) Databases: build core database as well as database for search, (c) Analysis workflow: in-house workflows to generate structural and functional annotations as well as Ensembl-related workflows to build gene trees and provide variant effect prediction for genes, (d) Web services: to provide programmatic access to databases for the user interfaces or custom scripts, and (e) User interfaces: landing pages for users to interact with data for access to GrameneOryza's powerful search tools and integrated views. The result pages integrate different types of data into one panel under different tabs.

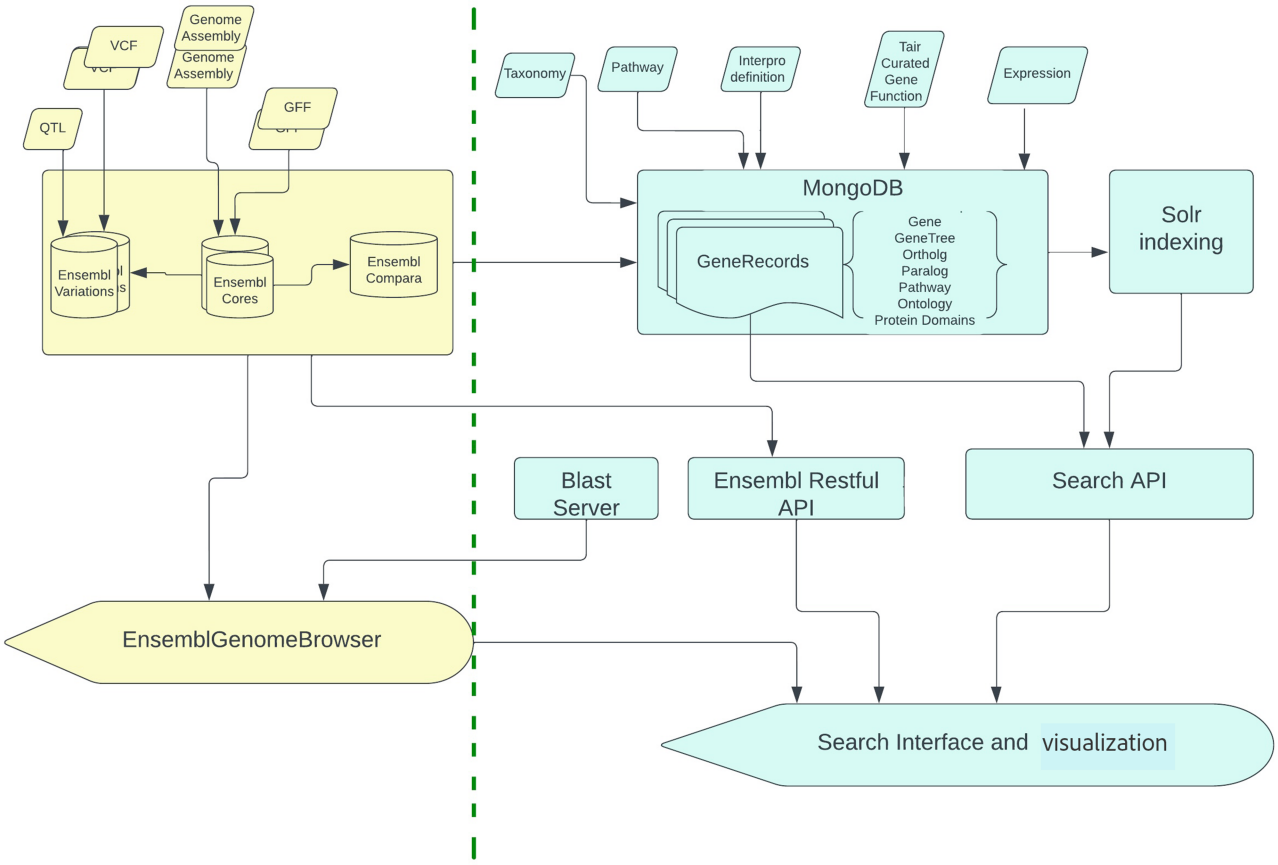


**Figure 2.** GrameneOrzya site infrastructure and its components. The Ensembl data cores for genomes, genetic variation, and comparative analyses (protein and DNA) were installed on a dedicated 250 GB Linux CentOS image with 16 GB memory and 2 CPUs. The core layer, represented in yellow, and the outer layer, depicted in turquoise, are separated by dotted green lines. The construction of the core layer databases is a prerequisite before the outer layer process, which aggregates and indexes them.

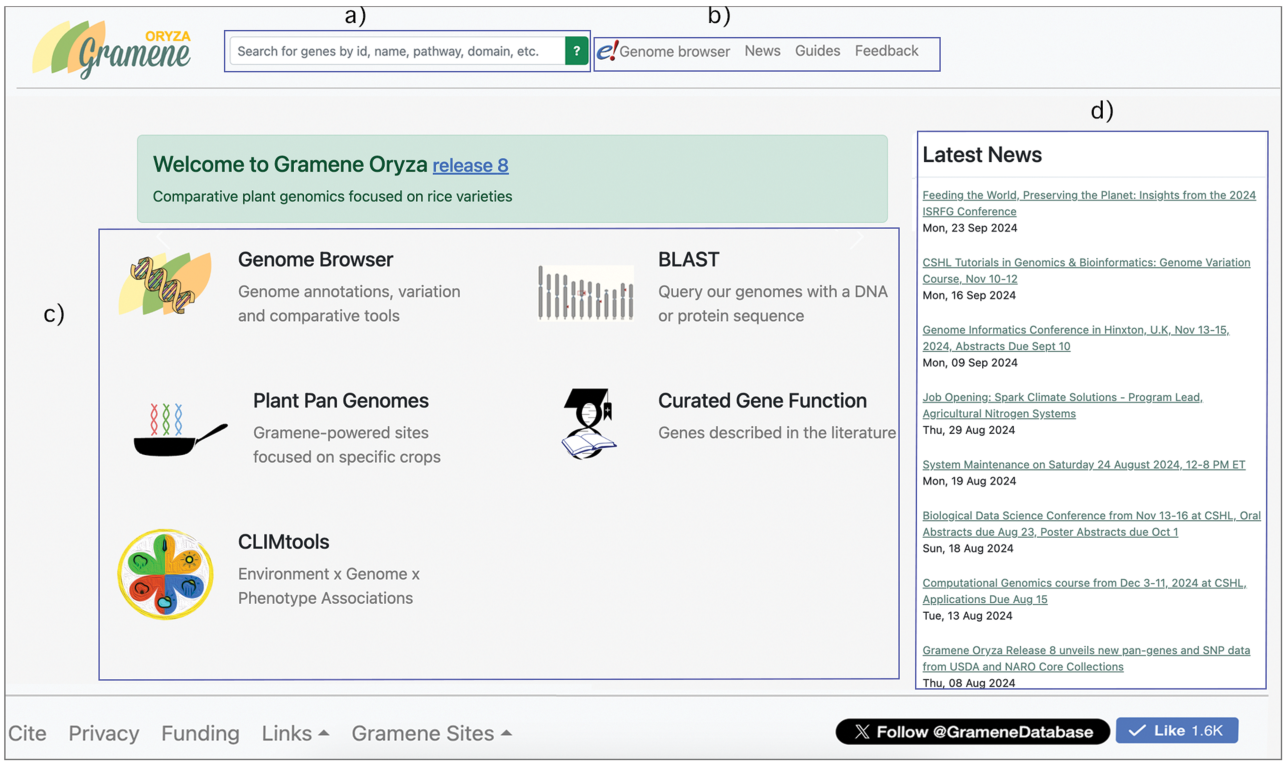


**Figure 3.** GrameneOryza user interface: (a) Search: textual query search for genes. (b) Quick links to Genome Browser, News, Release Notes, Guides, and Feedback form. (c) Tool and services panel with access to curated and published rice genes and other Gramene pansites as well as tools like BLAST, CLIMtools for analysis. (d) Latest news from the rice research community.

services can be found in the main section (Fig. 3c): "BLAST" allows user to query DNA or protein sequences against indexed database for 28 rice genomes along with their outgroups species; "Plant Pan Genomes" provides access to other Gramene powered crop species pan-sites including Sorghum, Maize, and Grapevine along with main GramenePlants site; and the "Curated Gene Function" tab offers quick access to rice genes published in literature and the inclusion of *Oryza* "CLIMtools" provides access to environment–genome associations and correlations for a large pool of curated *Oryza* genotypes. Finally, Fig. 3d provides the latest news coming from rice research. In addition, under each gene hit in the search results we can further study the gene with classified functional tabs and visualizations as described in Supplementary Table S1.

## Case studies

### Detecting CNVs for genes of interest

CNVs in rice (*O. sativa*) are a major source of genetic diversity, involving DNA segments that vary in copy number among individuals. First recognized in plant genomes in the early 2000s [50], CNVs have been linked to important traits like disease resistance [51], stress tolerance [52], and yield improvement [53]. Early work by Wang *et al.* [54] mapped CNVs across the rice genome, laying the foundation for understanding their functional significance. As sequencing technologies have advanced, researchers have been able to better identify and characterize these variations. Recent studies, such as Qin *et al.* [55], have employed whole-genome resequencing to discover novel CNVs and investigate their effects on gene function and phenotypic variation in rice. A well-known example of rice CNVs involves genes related to the hormone strigolactone (SL) biosynthesis, such as *MAX1* orthologs, which are part of the cytochrome P450 protein family. SL triggers the germination of the root parasitic plant *Striga hermonthica*, inhibiting shoot branching, and reduces tillering in the host plant [56, 57]. Studies have shown that variations in SL dosage due to CNVs in *SLB* genes are associated with different tillering phenotypes among rice varieties [56]. To explore this gene in the GrameneOryza portal, we type "SLB" into the search box at the top of the home page; the typeahead feature immediately suggests potential hits from different categories (Fig. 4a). There are 10 matching terms in the gene section, 3 in the Interpro Family. Selecting "OsSLB1" from "Genes," we landed on a page with gene ID *Os01g0700900*. Clicking the "Homology" tab brings up an embedded overview of the protein alignments of this gene family of 187 members (Fig. 4b). Select "Neighborhood conservation" from the "Display Mode" drop down list, we come to the genetree neighborhood view of this family (Fig. 4c). The conserved region on chromosome 1 (blue line) demonstrates good synteny across the 28 *Oryza* + *Leersia* genomes. The yellow internal nodes indicate potential split genes. We exclude them from our analysis by collapsing them. Highlighting the *SLB* gene family helps us to view the local CNV across the *Oryza* clade. The range is one to three copies in this region of ~200–300 Kb. In more detail, 5 genomes with one copy (Ketan, *O. rufipogon*, IR8, *O. punctata*, *O. brachyantha*), 16 with two copies (Azucena, ChaoMeo, *Leersia*, *O. glumipatula*, *O. barthii*, *O. glaberrima*, LMugael, Natel Bora, Carolina, YaiGuang, Lima, LiuXu, N22, ARC, Global Sail,

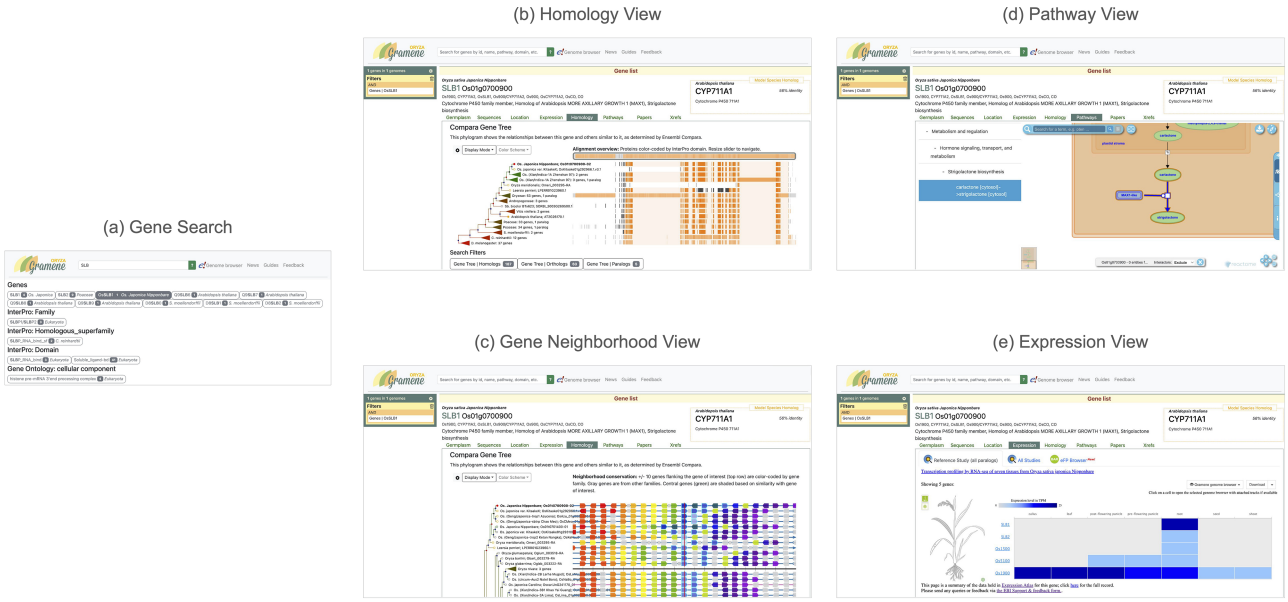


**Figure 4.** Use Case for finding copy number variations (CNVs) using GrameneOryza interface: (a) Gene Search: Keyword search produces many hits in different categories. One of the hits in the category "gene" is the rice version of SLB *OsSLB1*. (b) Homology View: the neighborhood view shows CNV across the landscape of the *Oryza* clade. (c) Pathway View: the pathway data from Reactome supports this gene's regulatory role in strigolactone biogenesis. (d) Expression View: The Atlas expression data shows *OsSLB1* is highly expressed in root and sometimes in leaf sheath. Three paralogs have similar expression patterns, but two have quite different expression patterns.

IR64), 3 genomes with 3 copies (Nipponbare, KitaakeX, *O. meridionalis*). Consistent with the expectation, the "Pathway" tab shows strigolactone biosynthesis (Fig. 4d) and the "Papers" tab lists curated publications. The "Expression" tab displays the *OsSLB1* expression pattern across tissues and studies as well as that of its paralogs, which indicates neofunctionalization among the paralogs (Fig. 4e).

### Using genetic variation for gene to determine protein truncation variants

Genetic variation in rice (*O. sativa*) has been extensively studied due to its critical role in crop improvement and adaptation. Researchers have focused on characterizing different types of variations, such as single SNPs, Indels, and CNVs, all of which contribute to the genetic diversity observed in rice populations [54]. This diversity is essential for understanding traits like disease resistance, drought tolerance, and yield, which are vital for rice breeding programs aimed at developing improved varieties [56]. Advancements in sequencing technologies have allowed for better identification and characterization of these variations. For instance, resequencing efforts involving both cultivated and wild rice accessions have provided valuable markers for the discovery of agronomically important genes, aiding breeders in selecting favorable traits for rice improvement [53]. Several large-scale projects and databases have been established to catalog rice genetic variations. For example, the 3000 Rice Genomes Project (RG3K), which IRRI spearheaded in collaboration with the Chinese Academy of Agricultural Sciences and BGI-Shenzhen, has an SNP-Seek database of SNPs derived from 3 K rice genomes [58]. The China National Rice Research Institute, in collaboration with other organizations, contributes to rice genomic databases and has developed the Rice Information System to access rice genomic data [59]. Japan has developed its own database as well, with the Rice Annotation Project Database, hosted by the National Institute of Agrobiological Sciences, as a major contributor to rice genomics [4]. The EVA is also one such database that provides open-access genomic variation data for non-human species, including rice [60]. These variation datasets are all linked to GrameneOryza, which focuses specifically on rice and provides curated information on genes, gene models, protein domains, and genetic variation. It integrates data from resources like EVA and other published studies to offer users a comprehensive platform for exploring rice genetics and diversity, including the functional annotations of variants and their associations with phenotypic traits. Each genetic variant is typically associated with a unique reference SNP cluster ID, or rsID. This rsID acts as a standardized identifier for SNV, allowing researchers to easily track, compare, and cite specific variations across different studies and databases [61]. These identifiers are critical for linking genetic variants to known phenotypes, making it easier to explore how specific SNPs or Indels contribute to important agricultural traits in rice.

Research in rice has increasingly focused on leveraging genetic variation to identify and characterize protein truncation variants (PTVs), which are mutations that lead to premature stop codons, frameshifts, or splice site disruptions, resulting in shortened, non-functional proteins. PTVs are of particular interest in rice breeding and functional genomics because they can reveal gene loss-of-function effects that are linked to key agronomic traits. By analyzing genome-wide variation data, researchers have identified PTVs associated with traits like disease resistance, stress adaptation, and yield improvement [62]. High-throughput sequencing technologies have enabled the detection of these variants across diverse rice populations. Studies have utilized bioinformatics pipelines to filter and validate PTVs, integrating this data with expression profiles and phenotypic information to determine the impact of PTVs on rice development and productivity. This research

offers valuable insights into gene function, contributing to the selection of beneficial mutations for rice breeding programs aimed at enhancing crop performance [54]. One prominent example is the *GS3* gene (Os03g0407400), which has been extensively studied for its role in controlling grain size. The GS3 protein is known to negatively regulate grain length in rice, with loss-of-function mutations leading to increased grain size, a trait that is often desirable in breeding programs. PTVs in *GS3*, such as premature stop codons, result in truncated and non-functional proteins, which disrupt the gene's normal function and lead to longer grains [63]. Research has shown that natural allelic variations in *GS3* are associated with variations in grain length among different rice varieties. For example, studies by Takano-Kai *et al.* [64] identified specific mutations in *GS3* that are responsible for the long-grain phenotype in certain indica rice varieties. Further genome-wide association studies (GWAS) have confirmed that these truncation variants are major contributors to grain size diversity in rice populations, making *GS3* a key target in rice breeding [65]. Additionally, targeted gene-editing approaches like CRISPR-Cas9 have been used to introduce PTVs in *GS3*, resulting in enhanced grain length and yield. These findings underscore the importance of understanding PTVs like those in GS3 for crop improvement efforts, as they directly impact traits that are critical for both productivity and market value [66].

To explore the *GS3* gene (*Os03g0407400*) in the GrameneOryza portal, enter "GS3" into the search box at the top of the homepage. Clicking the "Germplasm" tab brings up a table of germplasms containing predicted loss-of-function alleles for the *GS3* gene (Fig. 5a). The table also includes information like synonyms, studies, Variant Effect Predictor consequences, allele status, and links to view all loss-of-function genes for each germplasm. Clicking the "Search" button for any germplasm takes you to the Ensembl genome browser, where additional variants within the gene are highlighted (Fig. 5b—Variant image).

The *GS3* gene shows stop gained variants in 77 germplasms across 3 independent studies (as shown in Fig. 5a). You can further explore variants within this gene on the Variant image page (Fig. 5b) in the Ensembl genome browser. Notably, the stop gained variants appear on three exons (Os03t0407400-01-E1, Os03t0407400-01-E2, and Os03t0407400-01-E5) of the *GS3* gene. In the Variant table (Fig. 5c), you can filter variants based on their consequence type. Using the "Consequence Type" filter, you can narrow down the variants to display PTVs. Applying this filter reveals nine variant types resulting in PTVs. Additionally, by clicking the transcript ID Os03t0407400-01, you can view a variant comparison image (Fig. 5d), which graphically displays the transcript and variation data in a genomic context across different breeds or strains. The haplotype display shows allele positions where the reference matches in green and mismatches in purple. For heterozygous positions, a striped pattern of green and purple indicates the presence of both alleles.

## Discussion

The GrameneOryza resource offers a comprehensive platform for *Oryza* genomic data, addressing key challenges in rice research and providing tools that are valuable for both basic and applied research. One of the significant features of GrameneOryza is the incorporation of gene family trees, which play a crucial role in identifying orthologs and paralogs across the *Oryza* clade and related species. Gene trees provide a framework for understanding evolutionary relationships between genes, helping researchers identify gene families that are conserved across species or that have undergone lineage-specific expansions. This information is vital for developing *Oryza* pan-gene sets, which capture the full range of genetic diversity within the genus, including PAVs and CNVs. As rice breeding programs increasingly focus on tapping into this genetic diversity, pan-gene sets enable researchers to explore the full spectrum of functional diversity, leading to more targeted approaches in crop improvement. The inclusion of standardized rsIDs from the EVA supports Fairification of the data, further enhancing the utility of the genetic variation data accessible within GrameneOryza. rsIDs provide a globally recognized identifier for genetic variants, facilitating cross-referencing with other datasets and enabling reproducibility of agricultural research. The standardized rsID system ensures that researchers can easily compare genetic variation across different studies, accelerating efforts to link specific variants to phenotypic traits, such as disease resistance, stress tolerance, or yield improvement. The availability of 27 million rsIDs, anchored on the Nipponbare reference genome, provides a comprehensive view of rice genetic variation, which is particularly valuable for GWAS and other high-throughput function analyses.

One of the major strengths of GrameneOryza is its integrated search interface, which allows users to seamlessly navigate across various types of data, including genomic, phylogenetic, gene function, genetic variation, population, and QTL data. The platform's intuitive BLAST server provides researchers with a powerful tool for sequence-based searches, enabling them to quickly query their sequences against a curated database of *Oryza* genomes and related species. This functionality is particularly useful for identifying potential gene variants or homologs in newly sequenced rice varieties. The integration of diverse data types into a single, user-friendly platform significantly reduces the barriers to accessing and interpreting complex genomic data.

Another strength of GrameneOryza is the infrastructure for community curation. The phylogenetic tree-based gene structure interface provided by Homology tab enables the user to quality check the structures of the gene members in the context of homologous protein-alignments and flag the discordant ones (Fig. 6a). This unique function is especially useful in detecting potential split-gene instances, missing sequences, or annotation errors such as gene fusions or incomplete gene models and providing opportunities to identify the low-confident gene models for further inspection and improvement. Another place with community curation capability is the Papers tab, where "Submit a gene function here" feature (Fig. 6b) prompts users and researchers to submit their functional curation of the gene and meta data by filling out a Google form (Fig. 6c). As the volume of genomic data continues to grow, community curation will be critical in maintaining the accuracy and usability and scalability of public databases like GrameneOryza.

The need for consistent pan-gene sets and uniform annotations across the *Oryza* genus cannot be overstated. Given the high degree of genome fluidity observed in rice and related

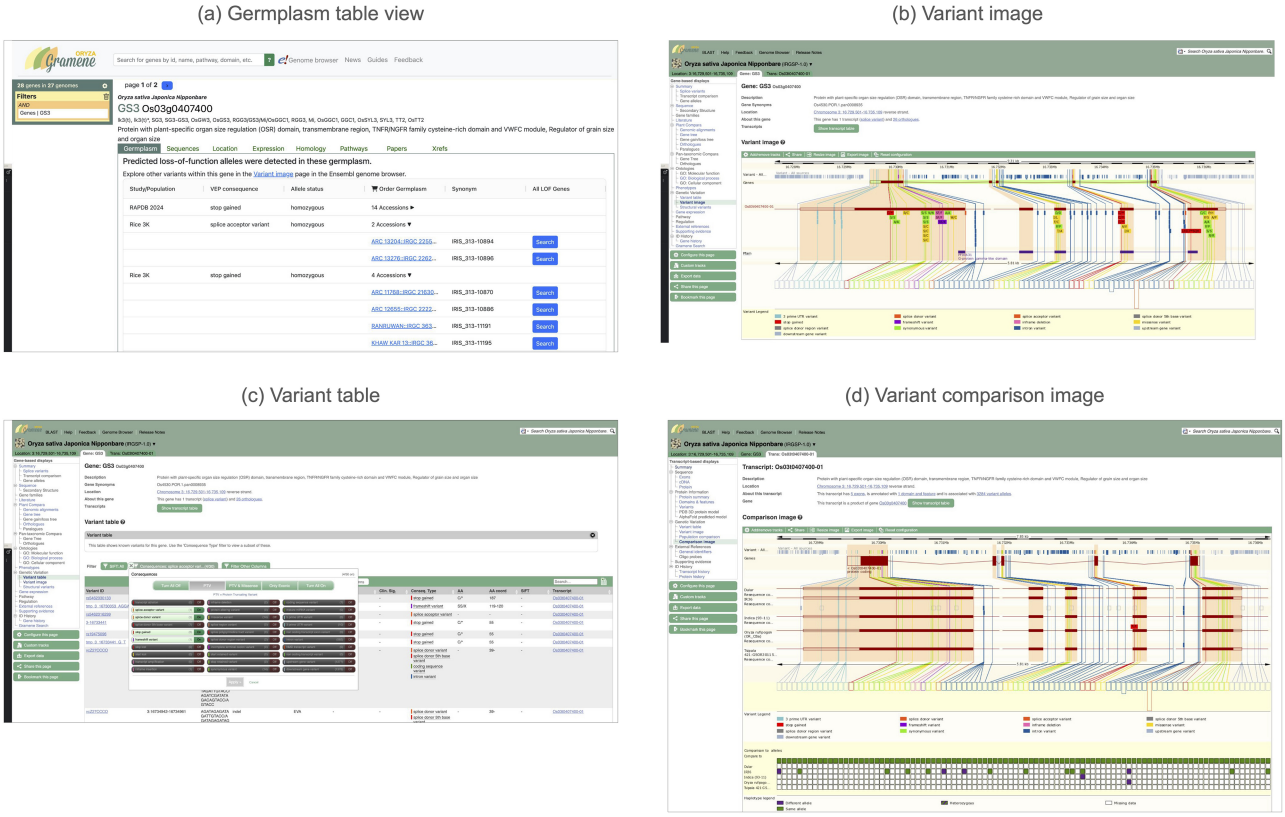


**Figure 5.** Identification of PTVs using the GrameneOryza interface. (a) Germplasm: displays a list of germplasms containing predicted loss-of-function alleles for the *GS3* gene. Clicking the hyperlinked text "Variant image" will bring you to the Ensembl genome browser's gene page Variant Image panel. Clicking the "Search" button for any germplasm links to additional genes with predicted loss-of-function mutations in the selected germplasm. (b) Variant Image: visualizes short sequence variants across different *GS3* transcripts. The variants are represented as vertical lines, color-coded based on their relative positions within the transcript. (c) Variant Table: displays variant information by consequence type, defaulting to all consequences for the *GS3* gene. The table allows filtering for PTVs, focusing specifically on variations that result in protein truncation. (d) Variant Comparison Image: clicking on a specific *GS3* transcript shows variation data in a broader genomic context, enabling comparison of variation across multiple germplasms or strains.

species, single reference genomes often fail to capture the full extent of functional diversity. Pan-gene sets address this limitation by incorporating genetic variations, such as PAVs, across multiple genomes, providing a more complete representation of the genetic landscape. Consistent annotations across these genomes ensure that the data are comparable and interoperable, enabling researchers to make accurate inferences about gene function and evolutionary relationships. As rice research moves toward a more systems-level understanding of gene function, having access to well-annotated pan-genomes will be essential for advancing both basic biology and applied breeding programs.

GrameneOryza is built on the principles of FAIR data management to ensure the highest standards of access and interoperability. GrameneOryza was developed within an ecosystem of community databases (Supplementary Table 2) that champion the principles of FAIR.

Of the 28 Oryza genomes, 27 have been deposited in the INSDC with GCA accessions (Supplementary Table 3). This ensures the provenance of genome assemblies with metadata are documented, organized, and traceable. The sole exception is the Carolina Gold genome [14] from USDA, which is not in INSDC; however, the genome sequences are fully released through Cyverse data stores, and can be downloaded from https://de.cyverse.org/dl/d/2C3CF540-2962-4BC2-8131-6CE8AA4FA4FE/oryza_carolina-top level-20180831.fa.gz, which is provided on the species info page of the site. The gene annotations were performed with a suite of published open source software such as Maker-P [67], PASA [68], etc, which guarantees reproducibility. The pan-genomes were imported from PanOryza project (https://panoryza.org/), which named the pan-genes based on the nomenclature guidelines set by the rice community. Protein features were annotated using InterproScan [69], an open-source, documented and well-maintained open-source tool from EBI. Interpro2Go [70], is an important step of this pipeline, assigns protein function with GO terms. These GO terms are community standards and that ensures the support of interoperability with other resources and tools.

The genetic variation data included in GrameneOrzya were obtained from both published studies and collaborators. To support FAIR compliance, these data were updated with rsIDs from EVA release 5 [60, 70]. When multiple identifiers were available for germplasm samples, priority was given to standard identifiers from germplasm centers such as GRIN-Global [71], RG-3 K project [27], National Agriculture and Food Research Organization (NARO, https://www.naro.go.jp/english/). Cross-links back to their respective germplasm center pages were provided, offering access to other sample names as synonyms.

GrameneOrzya also utilizes controlled vocabulary from established ontologies. Examples include GO for

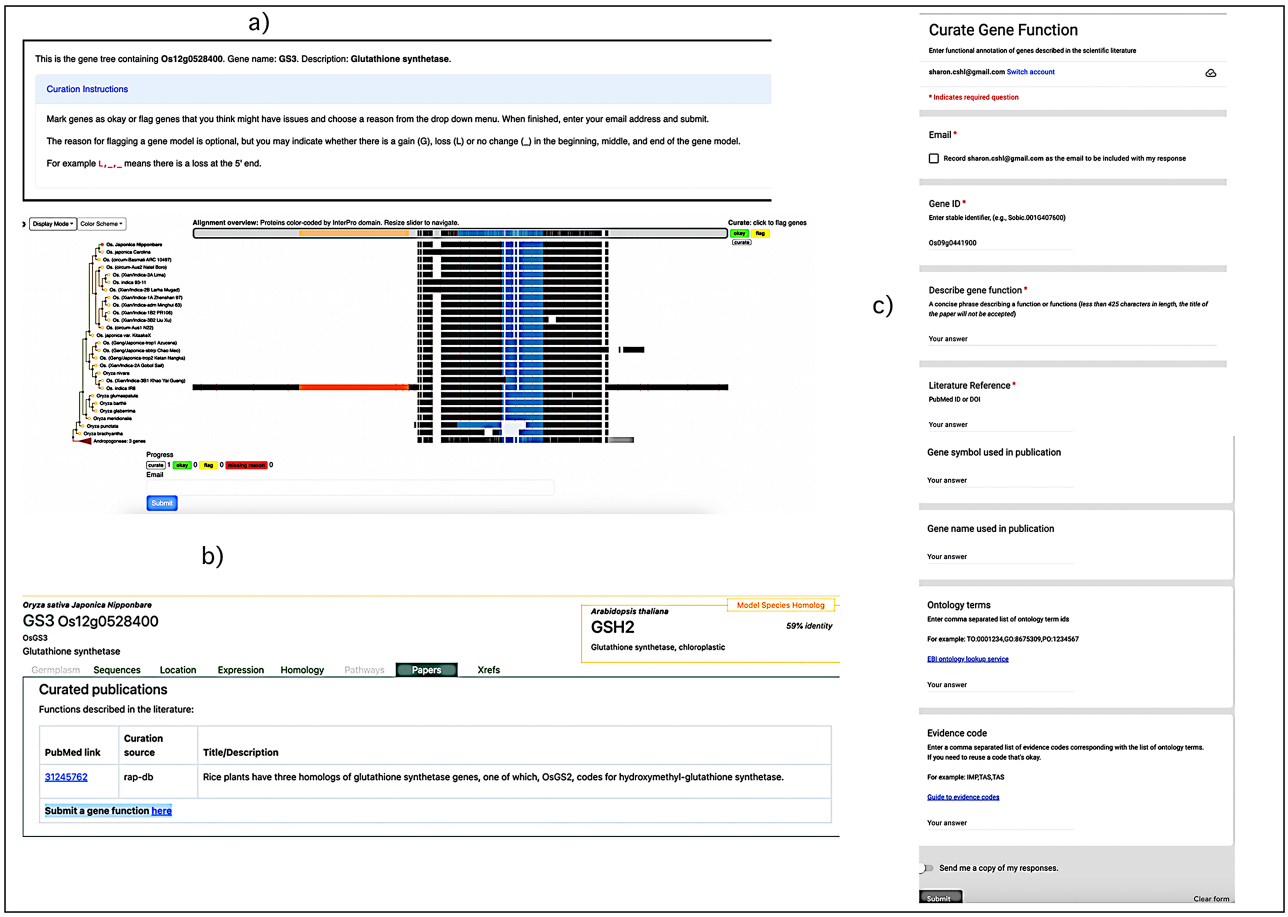


**Figure 6.** Community curations in GrameneOryza interface. (a) Homology-based gene structure inspection interface allows users to flag problematic models for further inspection and improvement. (b) Paper tab displays the functions curated by RAP-DB and GeneRIF, but also prompts users to enter their curation of the gene function. (c) The Google form that takes the curation of the gene function and its associated metadata provided by the user.

characterizing a gene's function and cellular localization, and TO, the Trait Ontology [49] for phenotypes associated with QTLs. To enhance accessibility, GrameneOryza developed powerful integrated search interface. Users can easily explore and export the underlying sequences and features. For bulk data needs, large datasets and batch downloads are supported through the dedicated ftp site (https://ftp.gramene.org/oryza/release-current/).

In summary, GrameneOryza offers an integrated, comprehensive, and user-friendly platform for rice genomic data. Its features, including gene trees, standardized rsIDs, enhanced data access tools, and community-driven curation, provide invaluable resources for the rice research community, supporting efforts to explore genetic diversity, improve crop varieties, and address challenges posed by climate change.

## Conclusion

GrameneOryza has made significant contributions to the rice research community by providing a comprehensive resource that integrates high-quality genomic, genetic variation, and functional data across multiple *Oryza* species. Its key features, such as gene family trees, standardized rsIDs, and an intuitive search interface, enable researchers to efficiently access and analyze diverse data types. The community-driven curation interface further enhances the accuracy of annotations, making the platform a valuable tool for exploring genetic diversity, evolutionary relationships, and functional genomics. Looking ahead, future enhancements could include expanding the repository with more emerging *Oryza* genomes and improving the integration of environmental-genomic data through tools like CLIMtools. Additionally, the development of more robust pan-gene sets and enhanced comparative genomic features would further support efforts to address challenges in rice breeding, climate resilience, and sustainable agriculture. As GrameneOryza continues to evolve, it will remain a critical resource for advancing both basic research and applied agricultural innovations.

## Acknowledgements

We would like to acknowledge our collaborators for their valuable contributions to the project, including King Abdullah University of Science and Technology (KAUST), the International Rice Research Institute (IRRI), the NSF Minicore Project, the Rice Annotation Project Database (RAP-DB), the European Variation Archive (EVA), the EBI Expression Atlas, the Germplasm Resources Information Network (GRIN), Pennsylvania State University (Oryza CLIMtools), and Oregon State University (Plant Reactome). Their expertise, data contributions, and ongoing support have been instrumental

in advancing this work. We thank Kate Dooling Cold Spring Harbor Laboratory (CSHL) and Christopher Olson (CSHL) for helping with the community curation through phylogenetic tree interface to detect potential split/merge gene models, and Peter Van Buren for the support of the computational systems at Cold Spring Harbor Laboratory including virtual machines, database servers, storage, etc. We also extend our thanks to the broader research community stakeholders for their input and participation in training and user support activities, which continue to enhance the success and impact of GrameneOryza.

## Supplementary data

Supplementary data is available at *Database* online.

Conflict of interest: None declared.

## Funding

Core funding for this project was provided by the National Science Foundation (NSF IOS-1127112). Ongoing support is received from the Agricultural Research Service of the U.S. Department of Agriculture (USDA ARS 8062–21000-044-000D and 8062–21000-051-00D). The computational resources used for this project were supported by the Elzar High Performance Computing facility at Cold Spring Harbor Laboratory, funded by the National Institutes of Health (NIH S10 OD0286321-01).

## Data availability

All data presented in this manuscript are available through the *GrameneOryza* platform, which can be accessed at https://oryza.gramene.org. The resource includes genomic assemblies, gene annotations, genetic variation data, and functional annotations for *Oryza* species. Additionally, SNP variation data with rsIDs from the EVA are available for download through the platform. Users can also access the curated pan-gene sets, gene family trees, and QTL data. For programmatic access, GrameneOryza provides web services, including APIs and a dedicated BLAST server. For further information or specific data requests, please refer to the GrameneOryza documentation or contact the platform support team: https://oryza.gramene.org/feedback

## References

1. Seck PA, Diagne A, Mohanty S *et al*. Crops that feed the world 7: rice. *Food Security* 2012;**4**:7–24.
2. Mohidem NA, Hashim N, Shamsudin R *et al*. Rice for food security: revisiting its production, diversity, rice milling process and nutrient content. *Collect FAO Agric* 2022;**12**:741.
3. Kawahara Y, de la Bastide M, Hamilton JP *et al*. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 2013;**6**:4.
4. Sakai H, Lee SS, Tanaka T *et al*. Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol* 2013;**54**:e6.
5. Stein JC, Yu Y, Copetti D *et al*. Publisher correction: genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* 2018;**50**:1618.
6. Wing RA, Ammiraju JSS, Luo M *et al*. The Oryza Map Alignment Project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol Biol* 2005;**59**:53–62.
7. Li Y-H, Zhou G, Ma J *et al*. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat Biotechnol* 2014;**32**:1045–52.
8. Hufford MB, Seetharam AS, Woodhouse MR *et al*. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 2021;**373**:655–62.
9. Golicz AA, Batley J, Edwards D. Towards plant pangenomics. *Plant Biotechnol J* 2016;**14**:1099–105.
10. Saxena RK, Edwards D, Varshney RK. Structural variations in plant genomes. *Brief Funct Genomics* 2014;**13**:296–307.
11. Zhou Y, Chebotarov D, Kudrna D *et al*. A platinum standard pan-genome resource that represents the population structure of Asian rice. *Sci Data* 2020;**7**:113.
12. Zhang Q *et al*. N6-Methyladenine DNA methylation in Japonica and Indica rice genomes and its association with gene expression, plant development, and stress responses. *Mol Plant* 2018;**11**:1492–508.
13. Jain R, Jenkins J, Shu S *et al*. Genome sequence of the model rice variety KitaakeX. *BMC Genomics* 2019;**20**:905.
14. Vaughn JN, Korani W, Stein JC *et al*. Gene disruption by structural mutations drives selection in US rice breeding over the last century. *PLoS Genet* 2021;**17**:e1009389.
15. Song JM *et al*. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol Plant* 2021;**14**:1757–67.
16. Wilkinson MD, Dumontier M, Aalbersberg IJ *et al*. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 2016;**3**:1–9.
17. Ruffier M, Kähäri A, Komorowska M *et al*. Ensembl core software resources: storage and programmatic access for DNA sequence and genome annotation. *Database* 2017;**2017**: bax020.
18. Vilella AJ, Severin J, Ureta-Vidal A *et al*. EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* 2009;**19**:327–35.
19. Hubbard T, Barker D, Birney E *et al*. The Ensembl genome database project. *Nucleic Acids Res* 2002;**30**:38–41.
20. Tello-Ruiz MK, Naithani S, Gupta P *et al*. Gramene 2021: harnessing the power of comparative genomics and pathways for plant research. *Nucleic Acids Res* 2021;**49**:D1452–D1463.
21. Ware D. Gramene. *Methods Mol Biol* 2007;**406**:315–29.
22. Tello-Ruiz MK, Naithani S, Stein JC *et al*. Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Res* 2018;**46**:D1181–D1189.
23. Reiser L, Bakker E, Subramaniam S *et al*. The Arabidopsis Information Resource in 2024. *Genetics* 2024;**227**:iyae027.
24. Gupta P, Naithani S, Preece J *et al*. Plant reactome and PubChem: the plant pathway and (bio)chemical entity knowledgebases. *Plant Bioinformatics* 2022;**2443**:511–525. 10.1007/978-1-0716-2067-0_27
25. Kapushesky M, Emam I, Holloway E *et al*. Gene expression atlas at the European Bioinformatics Institute. *Nucleic Acids Res* 2009;**38**:D690–D698.
26. Papatheodorou I, Moreno P, Manning J *et al*. Expression atlas update: from tissues to single cells. *Nucleic Acids Res* 2020;**48**:D77–D83.
27. The and 000 rice genomes project. The 3,000 rice genomes project. *Gigascience* 2014;**3**:7.
28. Nagamura Y, Antonio BA, Sato Y *et al*. Rice TOGO Browser: a platform to retrieve integrated information on rice functional and applied genomics. *Plant Cell Physiol* 2011;**52**:230–37.
29. Yonemaru J-I, Yamamoto T, Fukuoka S *et al*. Q-TARO: QTL Annotation Rice Online Database. *Rice* 2010;**3**:194–203.
30. Ni J, Pujar A, Youens-Clark K *et al*. Gramene QTL database: development, content and applications. *Database* 2009;**2009**: bap005.
31. Ferrero-Serrano Á, Chakravorty D, Kirven KJ. Oryza CLIMtools: a genome–environment association resource reveals adaptive roles

for heterotrimeric G proteins in the regulation of rice agronomic traits. *Plant Commun* 2024;**5**:100813.
32. Kumar A, Gupta C, Thomas J *et al*. Genetic dissection of grain yield component traits under high nighttime temperature stress in a rice diversity panel. *Front Plant Sci* 2021;**12**:712167.
33. Tanaka N, Shenton M, Kawahara Y *et al*. Whole-genome sequencing of the NARO World Rice Core Collection (WRC) as the basis for diversity and association studies. *Plant Cell Physiol* 2020;**61**:922–32.
34. Tanaka N, Shenton M, Kawahara Y *et al*. investigation of the genetic diversity of a rice core collection of Japanese landraces using whole-genome sequencing. *Plant Cell Physiol* 2020;**61**:2087–96.
35. Altschul SF, Gish W, Miller W *et al*. Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.
36. Merchant SS, Prochnik SE, Vallon O *et al*. The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 2007;**318**:245–50.
37. Banks JA, Nishiyama T, Hasebe M *et al*. The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 2011;**332**:960–63.
38. Paterson AH, Bowers JE, Bruggmann R *et al*. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 2009;**457**:551–56.
39. Schnable PS, Ware D, Fulton RS *et al*. The B73 maize genome: complexity, diversity, and dynamics. *Science* 2009;**326**:1112–15.
40. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000;**408**:796–815.
41. Jaillon O, Aury J-M, Noel B *et al*. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007;**449**:463–67.
42. Adams MD, Celniker SE, Holt RA *et al*. The genome sequence of *Drosophila melanogaster*. *Science* 2000;**287**:2185–95.
43. Eilbeck K, Lewis SE, Mungall CJ *et al*. The sequence ontology: a tool for the unification of genome annotations. *Genome Biol* 2005;**6**:R44.
44. McLaren W, Gil L, Hunt SE *et al*. The ensembl variant effect predictor. *Genome Biol* 2016;**17**:1–14.
45. Mitchell JA, Aronson AR, Mork JG *et al*. Gene indexing: characterization and analysis of NLM's GeneRIFs. *AMIA Annu Symp Proc* 2003;**2003**:460–64.
46. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80.
47. Gene Ontology Consortium, Aleksander SA, Balhoff J *et al*. The Gene Ontology knowledgebase in 2023. *Genetics* 2023;**224**.
48. Walls RL, Cooper L, Elser J *et al*. The Plant Ontology facilitates comparisons of plant development stages across species. *Front Plant Sci* 2019;**10**:631.
49. Cooper L, Elser J, Laporte M-A *et al*. Planteome 2024 update: reference ontologies and knowledgebase for plant biology. *Nucleic Acids Res* 2024;**52**:D1548–D1555.
50. Schranz ME, Song B-H, Windsor AJ *et al*. Comparative genomics in the Brassicaceae: a family-wide perspective. *Curr Opin Plant Biol* 2007;**10**:168–75.
51. Zhang F, Gu W, Hurles ME *et al*. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet* 2009;**10**:451–81.
52. Cook DE, Lee TG, Guo X *et al*. Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 2012;**338**:1206–9.
53. Xu X, Liu X, Ge S *et al*. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat Biotechnol* 2011;**30**:105–11.
54. Wang W, Mauleon R, Hu Z *et al*. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 2018;**557**:43–49.
55. Qin P, Lu H, Du H *et al*. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 2021;**184**:3542–3558.e16.
56. Jiang L, Liu X, Xiong G *et al*. DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* 2013;**504**:401–5.
57. Xie X, Yoneyama K, Yoneyama K. The Strigolactone story. *Annu Rev Phytopathol* 2010;**48**:93–117.
58. Alexandrov N, Tai S, Wang W *et al*. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res* 2015;**43**:D1023–7.
59. Zhao W, Wang J, He X *et al*. BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res* 2004;**32**:D377–D382.
60. Cezard T, Cunningham F, Hunt SE *et al*. The European Variation Archive: a FAIR resource of genomic variation for all species. *Nucleic Acids Res* 2022;**50**:D1216–D1220.
61. Sherry ST, Ward M-H, Kholodov M *et al*. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;**29**:308–11.
62. Zhao F, Wang Y, Zheng J *et al*. A genome-wide survey of copy number variations reveals an asymmetric evolution of duplicated genes in rice. *BMC Biol* 2020;**18**:1–16.
63. Fan C, Xing Y, Mao H *et al*. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 2006;**112**:1164–71.
64. Takano-Kai N, Jiang H, Kubo T *et al*. Evolutionary history of GS3, a gene conferring grain length in rice. *Genetics* 2009;**182**:1323–34.
65. Mao H, Sun S, Yao J *et al*. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc Natl Acad Sci* 2010;**107**:19579–84.
66. Yuyu C, Aike Z, Pao X. Effects of GS3 and GL3.1 for grain size editing by CRISPR/Cas9 in rice. *Rice Sci* 2020;**27**:405–13.
67. Campbell MS, Law M, Holt C *et al*. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol* 2014;**164**:513–24.
68. Haas BJ, Salzberg SL, Zhu W *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to assemble spliced alignments. *Genome Biol* 2008;**9**:R7.
69. Jones P, Binns D, Chang H-Y *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40.
70. Blum M, Chang H-Y, Chuguransky S *et al*. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 2021;**49**:D344–D354.
71. National Research Council, Board on Agriculture and Committee on Managing Global Genetic Resources: Agricultural Imperatives (1990) The U.S. National Plant Germplasm System. *The U.S. National Plant Germplasm System*; National Academies Press, 1990.