



HHS Public Access

Author manuscript

Curr Plant Biol. Author manuscript; available in PMC 2017 November 01.

Published in final edited form as:

Curr Plant Biol. 2016 November ; 7-8: 10–15. doi:10.1016/j.cpb.2016.12.005.

Gramene Database: Navigating Plant Comparative Genomics Resources

Parul Gupta^{1,\$}, Sushma Naithani^{1,\$}, Marcela Karey Tello-Ruiz^{2,\$}, Kapeel Chougule², Peter D'Eustachio³, Antonio Fabregat⁴, Yinping Jiao², Maria Keays⁴, Young Koung Lee², Sunita Kumari², Joseph Mulvaney², Andrew Olson², Justin Preece¹, Joshua Stein², Sharon Wei², Joel Weiser⁵, Laura Huerta⁴, Robert Petryszak⁴, Paul Kersey⁴, Lincoln D. Stein⁵, Doreen Ware^{2,6}, and Pankaj Jaiswal^{1,*}

¹Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, USA

²Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

³NYU School of Medicine, New York, NY, USA

⁴European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, UK

⁵Ontario Institute for Cancer Research, Toronto, ON, Canada

⁶USDA ARS NEA Plant, Soil & Nutrition Laboratory Research Unit, Ithaca, NY, USA

Abstract

Gramene (<http://www.gramene.org>) is an online, open source, curated resource for plant comparative genomics and pathway analysis designed to support researchers working in plant genomics, breeding, evolutionary biology, system biology, and metabolic engineering. It exploits phylogenetic relationships to enrich the annotation of genomic data and provides tools to perform powerful comparative analyses across a wide spectrum of plant species. It consists of an integrated portal for querying, visualizing and analyzing data for 44 plant reference genomes, genetic variation data sets for 12 species, expression data for 16 species, curated rice pathways and orthology-based pathway projections for 66 plant species including various crops. Here we briefly describe the functions and uses of the Gramene database.

1. DATABASE OVERVIEW

Gramene (<http://www.gramene.org>) supports researchers working with various crops, models, and other economically important plant species by providing online resources for comparative analyses of Genomes and Pathways [1, 2]. The Genomes portal, developed collaboratively with Ensembl Plants, hosts annotated genome assemblies for 44 plant species

*To whom correspondence should be addressed Address of the corresponding author: Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR, USA, Phone: +1-541-737-8471, Fax: +1-541-737-3573, jaiswalp@oregonstate.edu.

[§]These authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

including lower plants, gymnosperms, and flowering plants. Users can access detailed information regarding individual genes and proteins, genetic and physical maps, phylogenetic trees based on whole-genome alignments, protein-based Compara gene family trees, genetic and structural variants, and expression data (*e.g.*, ESTs and alternatively spliced transcript isoforms). Furthermore, tissue-specific basal expression data and/or differential expression data (transcriptomics data from the EMBL-EBI's Expression Atlas project) from several plant species [3] can be viewed for individual genes or in their genomic context via the Genome Browser. The Genome Browser also facilitates upload, analysis, and visualization of user-defined high-throughput omics data (*e.g.*, transcriptomics and genetic diversity data). BLAST search can be conducted directly from the gene, transcript and variant summary pages, or using the BLAST tool accessible on the Tools page (<http://ensembl.gramene.org/tools.html>). Another tool, the Variant Effect Predictor (VEP), allows users to predict — either online or offline — the functional effect of genetic variants on gene regulation and encoded gene products [4]. For data mining, our BioMart tool enables complex queries of sequence, annotation, homology and variation data, and provides an additional gateway into the Genome Browsers [5].

The Pathways portal of Gramene, the Plant Reactome (<http://plantreactome.gramene.org>) [1, 6], was developed in collaboration with the Human Reactome project [7] and hosts pathways for 67 plant species (Gramene #release 52). In the Plant Reactome database, rice (*O. sativa*) serves as the reference species for curation of plant metabolic, transport, signaling, genetic-regulatory, and developmental pathways. The curated rice pathways are then used to derive orthology-based pathway projections for other species [6]. For a few species, users can also view baseline tissue-specific expression profile of the genes (transcriptomics data fetched remotely from the Expression Atlas project) associated with pathways in the the Pathway Browser [1, 6]. The Plant Reactome also allows users to i) upload and analyze omics datasets in the context of plant pathways, thereby fostering discoveries about the roles of genes of interest and their interacting partners, and ii) compare of pathways between the reference species rice and any projected species, as described recently [1, 6].

2. DATABASE NAVIGATION

2.1. Genes and Genome Browsers

Users can access various portals, tools, data, and release notes from the navigation panel located on the left-hand side of the Gramene homepage (<http://gramene.org>). Recently, we added a new dedicated search page (<http://search.gramene.org>) to support quick access to database contents (Genome Browser and Plant Reactome), bulk data downloads (FTP site), analysis tools (BLAST and BioMart), archived data, outreach and training material, a news blog, and external collaborators (*e.g.*, EBI Atlas) (Fig. 1). This interface offers interactive views of the search results both in aggregate form and in the context of a gene. The search bar, located at the top of this new user interface, facilitates queries for genes, pathways, gene ontology terms, and protein domains. In addition, the search bar provides suggestions for closely related terms and allows scientists to find genes by selecting among auto-suggested filters. Search results include a summary and associated interactive graphical depiction of a gene's structure, associated genomic location features and links to functional annotations in

the Genome Browser and external sites (e.g., PhytoMine and AraPort), a phylogenetic gene tree and associated homology elements, external references, and expression data from EMBL-EBI's Atlas.

The Genomes icon on the search page is hyperlinked to the Gramene's Genome Browser page (http://www.gramene.org/genome_browser), which lists the available genomes for various plant species. Users can select their species of interest and open the Genome Browser window to display a gene or genomic region (e.g., Fig. 1B shows the rice gene *Os06g0611900* in the *O. sativa* Japonica Genome Browser). From this page, users can access several other details pertaining to a gene's structure, function, and evolution, including synteny, gene trees, genetic variation data, and expression data (Fig. 1). As an example, Fig. 1C shows a view of synteny between rice chromosome 6 containing *Os06g0611900* and the corresponding region on *Zea mays* chromosomes 6 and 9. Such comparisons of syntenic genomic regions among plant genomes are particularly useful for identifying evolutionarily conserved co-linear regions, functional orthologs, and genetic markers [8].

The gene-based Compara trees provide information about the evolutionary history of genes in context of speciation. Gramene generates gene trees and alignments of orthologous and paralogous genes using the Ensembl Gene Tree method [9]. Users can access Compara trees from the Genome Browser page (the navigation panel located on the left-hand side). For example, the view of a gene-tree for rice *Os05g0113900* (Fig. 1D) shows gene duplication (red node), speciation events (blue node), and gene alignments (green lines) in various species.

2.2. Pathways and Pathway Browser

We recently described in detail the development of the Plant Reactome database, the pathway portal of Gramene, and its various functionalities [6]. Users can access the Plant Reactome database from the Gramene homepage or search page. Alternatively, users can directly access the Plant Reactome homepage (<http://plantreactome.gramene.org>), which provides links to the quick search, Pathway Browser, data analysis tools, video tutorials, user guide, data download, data model, database release summary, news, APIs, etc. By clicking on 'Browse pathways' or by searching for a pathway or any entity associated with that pathway users can access the default Pathway Browser for reference species rice (*O. sativa*). Fig. 2 provides an example of the Pathway Browser showing the 'abscisic acid biosynthesis and ABA-mediated signaling pathway' from the reference species rice (*O. sativa*). The left-hand side panel of the Pathway Browser shows the list of the available pathways (arranged hierarchically based on enzyme function and ontological concepts) to facilitate easy navigation. The top right-hand side panel shows a pathway diagram comprising of various types of macromolecular interactions (reactions) between proteins, protein complexes, and small molecules in the context of their subcellular locations. The bottom panel (below the pathway diagram) provides a pathway summary and data associated with various pathway entities, with hyperlinks to external databases (e.g. Fig. 2B) that provide further details on their structure, function, and expression. If desired, users can select species of their interest from the available options. At present the Plant Reactome (Gramene release #52) hosts ~240 metabolic, signaling, regulatory, and genetic rice reference pathways, omics data and

pathway comparison analysis tools, and orthology-based projections to over 78,000 gene products in 67 species.

The Plant Reactome allows users to select pathways of interest and visualize the baseline expression data using the Pathway Browser (Fig. 2C), and also provides links to differential expression data and additional detailed information about related experiments within the species (Fig. 2D). Data related to selected entities can be downloaded as PDF, Word, BioPAX, and SBML files.

2.3. Plant Gene Expression Atlas

Users can access the plant Expression Atlas from the Gramene search page. The plant gene Expression Atlas (<https://www.ebi.ac.uk/gxa/plant/experiments>) was developed by our collaborators at EMBL-EBI [3]. Currently, it contains transcriptomics data from 17 plant species corresponding to 698 experiments including baseline tissue-specific expression data and differential expression data (Fig. 3A). Manually curated, baseline expression data from RNA-Seq experiments are available from 12 plant species, showing expression levels of gene products under 'normal' conditions in various tissues (leaves, roots, etc.). As of release #52, the baseline expression profile of an individual gene across all tissue samples and growth stages from EMBL-EBI Expression Atlas (Fig. 3B) can be accessed from the gene page in the Gramene database (Fig. 3C), as well as from the Plant Reactome Pathway Browser page (Fig. 2C). Differential gene expression data are available from 13 plant species, and include datasets from both microarray and RNA-Seq experiments. At present, differential expression data can be viewed on the Expression Atlas website, and projected on demand (not automatically) onto the Gramene gene page and the Genome Browser.

Furthermore, users can select a given experiment and view the baseline expression profile of all available genes included in that dataset online at the plant gene Expression Atlas page (Fig. 3A), or offline after downloading the data (Fig. 3B). The plant gene Expression Atlas page has a widget for loading the expression data on the Gramene/Ensembl Genome Browser. User can select a gene and a cultivar, treatment or organismal part (Fig. 3D), and then open the Gramene/Ensembl Genome Browser. The resultant Genome Browser window shows the expression value of the selected gene, mapped onto the corresponding genomic region (Fig. 3E). Because data for all genes in a sample are automatically loaded into the Genome Browser, users can view the expression of all other genes in a sample by simply scanning the Genome Browser (by zooming in/out or selecting a different chromosome).

The Plant Reactome Pathway Browser automatically pulls out the baseline expression data from the EMBL-EBI Atlas database (Fig. 2C).

2.4. Analysis and Visualization Tools

For the processing of preloaded data, as well as data uploaded by users, Gramene hosts a number of analysis and visualization tools. The BLAST tool allows users to query orthologs from multiple target species using gene or protein sequence (Fig. 1E). The Genome Browser provides multiple options for displaying various types of preloaded data, such as genomic variation data (*i.e.* single-nucleotide polymorphisms [SNPs], ESTs, structural variants, etc.),

and also allows uploading of user-defined genomic data (*i.e.* transcriptomes, proteomes, long non-coding RNAs, methylomes, etc.) as described recently [1, 2, 10]. Users can also perform variant effect predictor (VEP) analysis to determine the functional consequence of genomic variations, such as SNPs and indels, on genes, transcripts, protein sequences, and regulatory regions using the VEP tool accessible from the Tools link on the Gramene homepage and Genome Browser pages. The online version of the VEP permits analysis of up to 700 variants (1 per row of a VCF file) in a single run. The analysis of large datasets can be performed offline by downloading the VEP Tool (<http://www.ensembl.org/info/docs/tools/vep/index.html>) and using command-line protocols and Perl scripts [4]. For some species, including tomato, pre-analyzed data are available from Gramene, and the consequences of genetic variants can be accessed online (Fig. 4).

Plant Reactome also allows users to upload and visualize omics data (*e.g.* transcriptome, proteome, metabolome, etc.) in the context of the plant pathways and to compare pathways between the reference species rice and any other species. Users have the option to download the results of the analysis, along with pathway diagram images as described recently [6].

To learn how to effectively mine data and use the resources and tools available at the Gramene database, open access video tutorials and training material are available via Gramene's outreach portal (<http://gramene.org/outreach>; Fig. 1F) and Gramene's YouTube channel (<https://goo.gl/qQ2Pjn>).

3. CONCLUSION

Gramene strives to provide plant researchers and breeders with the most updated and rich annotated data, tools, and user-friendly resources to support comparative plant genomics and pathway analysis. The contents, tools, and webpage of Gramene are updated three to five times annually. In each release, we add new genome assemblies, update assembly versions and annotations, and add new manually curated and projected pathways. We recommend that our users acquaint themselves regularly with our release notes and new updates to the database. We also host monthly webinars on various topics and welcome suggestions from users.

Acknowledgments

The work is supported by the Gramene database award (NSF IOS-1127112) and in-kind infrastructure and intellectual support from the Reactome database project (NIH: P41 HG003751 and U54 GM114833, ENFIN LSHG-CT-2005-518254, Ontario Research Fund, and EBI Industry Programme). The authors are grateful to Gramene's users, researchers, and numerous collaborators for sharing their datasets, valuable suggestions, and feedback, all of which have helped us to improve the overall quality of the resources and tools.

References

1. Tello-Ruiz MK, et al. Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.* 2016; 44(D1):D1133–40. [PubMed: 26553803]
2. Tello-Ruiz MK, et al. Gramene: A Resource for Comparative Analysis of Plants Genomes and Pathways. *Methods Mol Biol.* 2016; 1374:141–63. [PubMed: 26519404]
3. Petryszak R, et al. Expression Atlas update--an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* 2016; 44(D1):D746–52. [PubMed: 26481351]

4. McLaren W, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016; 17(1):122. [PubMed: 27268795]
5. Spooner W, et al. GrameneMart: the BioMart data portal for the Gramene project. *Database (Oxford)*. 2012; 2012:bar056. [PubMed: 22374386]
6. Naithani S, et al. Plant Reactome: a resource for plant pathways and comparative analysis. *Nucleic Acids Res.* 2016
7. Croft D, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 2011; 39(Database issue):D691–7. [PubMed: 21067998]
8. Youens-Clark K, et al. Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 2011; 39(Database issue):D1085–94. [PubMed: 21076153]
9. Vilella AJ, et al. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 2009; 19(2):327–35. [PubMed: 19029536]
10. Monaco MK, et al. Gramene 2013: comparative plant genomics resources. *Nucleic Acids Res.* 2014; 42(Database issue):D1193–9. [PubMed: 24217918]
11. Aflitos SA, et al. Introgression browser: high-throughput whole-genome SNP visualization. *Plant J.* 2015; 82(1):174–82. [PubMed: 25704554]

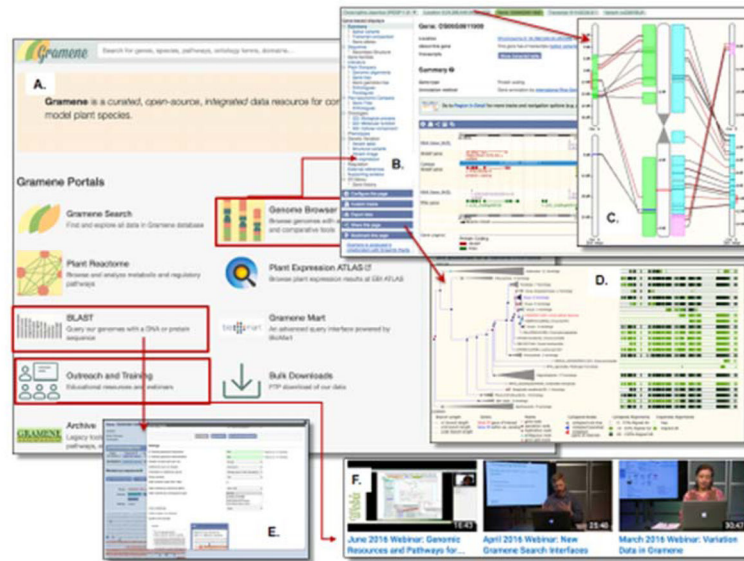


Fig. 1. Gramene's Genome Browser and associated data, tools, and features. (A) The Gramene search page serves as a springboard for accessing various tools, functionalities and genomic resources via clickable hyperlinks. (B) *O. sativa* Genome Browser showing location of a rice gene (*Os06g0611900*) and associated genomic data. The Genome Browser page also provides links for accessing i) synteny, e.g., synteny between the *O. sativa* subsp. Japonica locus *Os06g0611900* on chromosome 6 and the corresponding homologous regions on the *Z. mays* chromosomes 6 and 9 (C); and ii) Compara gene tree and expanded alignments with homologous and paralogous genes (D). From the search page, users can access the BLAST tool (E), training material including exercises, recorded webinars, etc. (F), GrameneMart and bulk data.

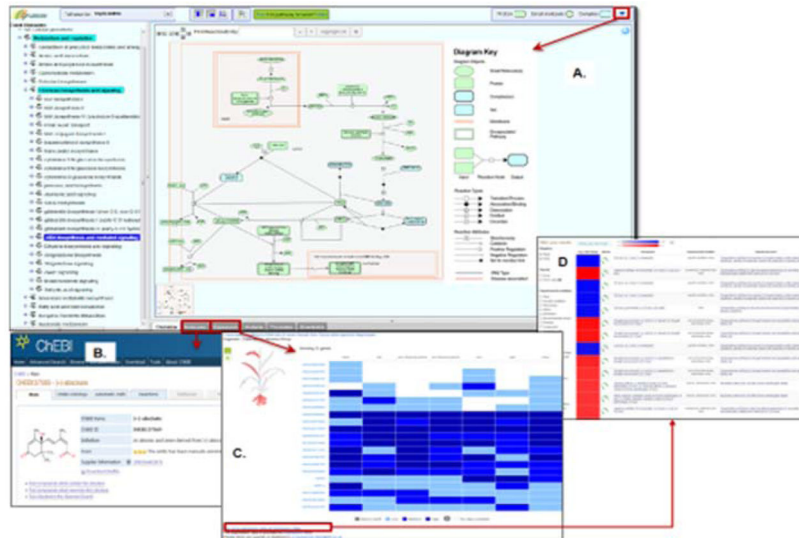


Fig. 2.

A view of the Plant Reactome Pathway Browser page. (A) An example of the ‘abscisic acid biosynthesis and mediated signaling pathway’ page from *O. sativa*. The left-hand side panel shows a hierarchical list of the available rice pathways, and the right-hand side upper panel shows a pathway diagram (depicted in the context of the plant cell’s subcellular structure) along with a diagram key. The ‘Zoom’ feature helps users visualize pathway diagrams more clearly. The lower panel, below the pathway diagram, provides options for exploring pathway-associated features, such as a list of proteins and small molecules with external links, e.g., to ChEBI (B); baseline tissue-specific expression data from EMBL-EBI’s Expression Atlas (C); and access to differential expression data at EMBL-EBI’s Atlas (D).

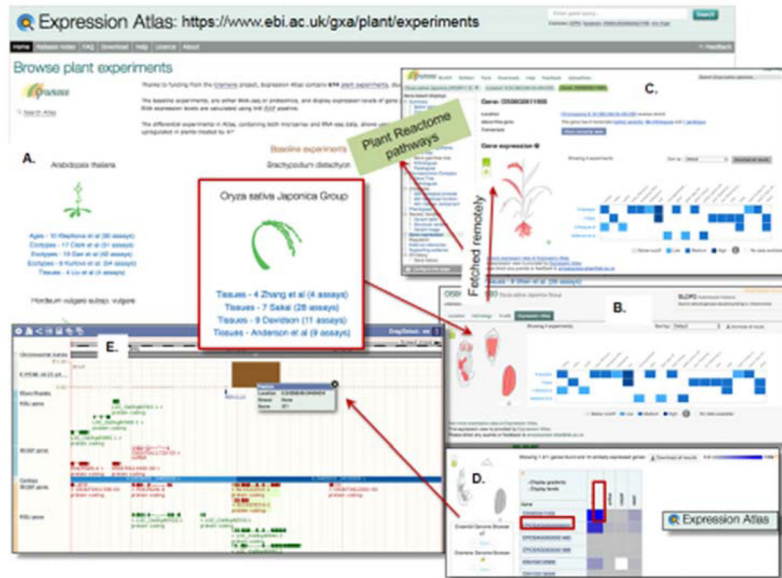


Fig. 3. Data and tools available in the plant gene Expression Atlas. (A) The plant gene Expression Atlas homepage (<https://www.ebi.ac.uk/gxa/plant/experiments>) showing a list of the available expression data/experiments from various plant species. (B) Baseline tissue-specific expression profiles of rice genes. (C) Baseline expression profiles of individual genes across multiple tissues can also be viewed from the gene page in Gramene, as well as from Plant Reactome Pathway Browser (see Fig. 2). Users can select baseline expression data from gene Expression Atlas (D) and open it in the Gramene Genome Browser page (E).

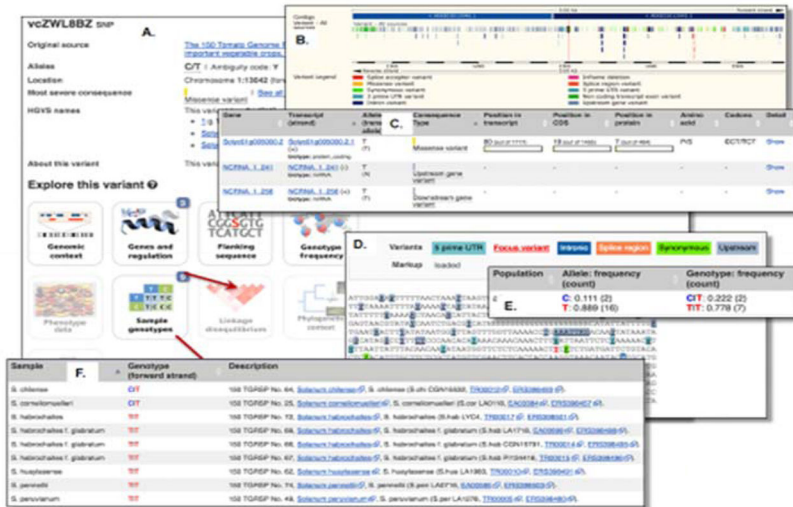


Fig. 4. An example of genomic variation data analysis at Gramene. (A) Genetic variations available in tomato, e.g., missense variation in *vcZWL8FA* [11]. (B) Genomic variations (SNP data for *ZWL8FA*) mapped to the Genome Browser. Variant effect predictor analysis for genes (C) and regulatory elements (D). (E) Allele frequency in the studied population. (F) Sample genotypes of 21 tomato accessions.