

Plant Reactome: a resource for plant pathways and comparative analysis

Sushma Naithani¹, Justin Preece¹, Peter D'Eustachio², Parul Gupta¹,
Vindhya Amarasinghe¹, Palitha D. Dharmawardhana¹, Guanming Wu^{3,4}, Antonio Fabregat⁵,
Justin L. Elser¹, Joel Weiser³, Maria Keays⁵, Alfonso Munoz-Pomer Fuentes⁵,
Robert Petryszak⁵, Lincoln D. Stein³, Doreen Ware^{6,7} and Pankaj Jaiswal^{1,*}

¹2082 Cordley Hall, Department of Botany & Plant Pathology, Oregon State University, Corvallis, OR 97331, USA, ²Department of Biochemistry and Molecular Pharmacology, NYU School of Medicine, New York, NY 10016, USA, ³Ontario Institute of Cancer Research, Toronto, ON M5G 0A3, Canada, ⁴Oregon Health & Science University, Portland, OR 97239, USA, ⁵European Molecular Biology Laboratory - European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK, ⁶Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA and ⁷United States Department of Agriculture – Agriculture Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA

Received August 22, 2016; Editorial Decision October 04, 2016; Accepted October 27, 2016

ABSTRACT

Plant Reactome (<http://plantreactome.gramene.org/>) is a free, open-source, curated plant pathway database portal, provided as part of the Gramene project. The database provides intuitive bioinformatics tools for the visualization, analysis and interpretation of pathway knowledge to support genome annotation, genome analysis, modeling, systems biology, basic research and education. Plant Reactome employs the structural framework of a plant cell to show metabolic, transport, genetic, developmental and signaling pathways. We manually curate molecular details of pathways in these domains for reference species *Oryza sativa* (rice) supported by published literature and annotation of well-characterized genes. Two hundred twenty-two rice pathways, 1025 reactions associated with 1173 proteins, 907 small molecules and 256 literature references have been curated to date. These reference annotations were used to project pathways for 62 model, crop and evolutionarily significant plant species based on gene homology. Database users can search and browse various components of the database, visualize curated baseline expression of pathway-associated genes provided by the Expression Atlas and upload and analyze their Omics datasets. The database also offers data access via Application Programming Interfaces (APIs) and in various standardized pathway formats, such as SBML and BioPAX.

INTRODUCTION

Multi-faceted bioinformatic resources and databases can facilitate understanding of a plant's genetic complexity, structure, development, evolution and its response to environmental stress conditions. It is a prerequisite for developing new and improved varieties of crops to meet the growing demand for production and yield and challenges posed by population growth, global climate change, and biotic and abiotic stresses. The Gramene database (<http://gramene.org/>) (1) provides online resources to plant researchers for conducting a comparative analysis of plant genomes and pathways (1). The Plant Reactome database (<http://plantreactome.gramene.org/>), the Pathways portal of Gramene, hosts metabolic, transport, genetic, signaling and developmental pathways for 63 plant species including various crops, experimental models and other species important to understand the evolution of pathway networks (Table 1). The Plant Reactome (Supplementary Figure S1) was developed in collaboration with the Human Reactome project (2) and utilizes their data model to organize proteins, protein complexes, small molecules and macromolecular interactions into reactions and pathways in the context of their subcellular location and interactions to build a systems-level pathway network of a plant cell.

The Plant Reactome features *Oryza sativa* (rice) as a reference species. To construct the reference pathway network, metabolic pathways from RiceCyc (3) were bulk imported into the Reactome data structure using BioPAX Level-2 conversion tools. Subsequently, these pathways and their various components were curated, and pathway diagrams were manually drawn to meet the Reactome database

*To whom correspondence should be addressed. Tel: +1 541 737 8471; Fax: +1 541 737 3573; Email: jaiswalp@science.oregonstate.edu

Table 1. Taxonomic distribution of the 63 plant species represented in the Plant Reactome pathway database.

		Taxa	Pathways	Reactions	Gene products	Sequence source	Outgoing Link	Homology method			
Angiosperm	Dicot	Malvid	<i>Arabidopsis lyrata</i>	187	508	942	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
			<i>Arabidopsis thaliana</i>	187	508	944	Ensembl Gramene	TAIR	Ensembl Compara		
			<i>Brassica oleracea</i>	186	497	1417	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
			<i>Brassica rapa</i>	186	504	1443	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
			<i>Citrus sinensis</i>	197	528	2026	Phytozome	Phytozome	Planteome Inparanoid		
			<i>Gossypium raimondii</i>	197	532	1405	Phytozome	Phytozome	Planteome Inparanoid		
			<i>Populus trichocarpa</i>	188	508	1181	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
			<i>Theobroma cacao</i>	191	511	775	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
			Core Rosid	Fabid	<i>Arachis duranensis</i>	196	518	788	PeanutBase	PeanutBase	Planteome Inparanoid
		<i>Arachis ipaensis</i>			194	522	801	PeanutBase	PeanutBase	Planteome Inparanoid	
		<i>Cajanus cajan</i>			197	525	1009	LegumeInfo	LegumeInfo	Planteome Inparanoid	
		<i>Cicer arietinum</i>			198	528	1192	NCBI	NCBI	Planteome Inparanoid	
		<i>Fragaria vesca</i>			198	514	888	Phytozome	Phytozome	Planteome Inparanoid	
		<i>Glycine max</i>			187	499	1730	Ensembl Gramene	Ensembl Gramene	Ensembl Compara	
		<i>Jatropha curcas</i>			187	499	575	KDRI (Kazusa)	KDRI (Kazusa)	Planteome Inparanoid	
		<i>Malus domestica</i>			196	511	1771	PMID: 20802477	Phytozome	Planteome Inparanoid	
		<i>Manihot esculenta</i>			191	510	1014	Phytozome	Phytozome	Planteome Inparanoid	
		<i>Medicago truncatula</i>			188	497	1084	Ensembl Gramene	Ensembl Gramene	Ensembl Compara	
		<i>Phaseolus vulgaris</i>			199	534	977	Phytozome	Phytozome	Planteome Inparanoid	
		<i>Prunus persica</i>			188	507	797	Ensembl Gramene	Ensembl Gramene	Ensembl Compara	
		Myrtales			<i>Eucalyptus grandis</i>	190	510	825	Phytozome	Phytozome	Planteome Inparanoid
					<i>Vitis vinifera</i>	184	496	835	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
		Vitales			<i>Capsicum annuum</i>	198	511	1004	PMID: 24441736	N/A	Planteome Inparanoid
					<i>Coffea canephora</i>	198	530	926	PMID: 25190796	N/A	Planteome Inparanoid
		Asterid			<i>Mimulus guttatus</i>	189	512	703	Phytozome	Phytozome	Planteome Inparanoid
					<i>Solanum lycopersicum</i>	190	504	960	Ensembl Gramene	Ensembl Gramene	Ensembl Compara
			<i>Solanum tuberosum</i>	187	486	933	Ensembl Gramene	Ensembl Gramene	Ensembl Compara		
	<i>Aegilops tauschii</i>		182	472	841	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Brachypodium distachyon</i>		186	527	873	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Hordeum vulgare</i>		184	469	734	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Leersia perrieri</i>		193	525	861	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza australiensis</i> *		194	494	1485	OMAP/OGE	N/A	Planteome Inparanoid			
	<i>Oryza barthii</i>		193	536	900	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza brachyantha</i>		193	528	850	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza glaberrima</i>		195	533	910	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza glumaepatula</i>		195	537	918	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza granulata</i> *		196	510	2565	OMAP/OGE	N/A	Planteome Inparanoid			
	<i>Oryza kasalath</i>		157	249	298	PMID: 24578372	N/A	Planteome Inparanoid			
	<i>Oryza longistaminata</i> *		192	523	790	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza meridionalis</i>		186	487	789	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza minuta</i> *		197	516	1891	OMAP/OGE	N/A	Planteome Inparanoid			
	<i>Oryza nivara</i>		191	542	916	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza officinalis</i> *		195	524	1465	OMAP/OGE	N/A	Planteome Inparanoid			
	<i>Oryza punctata</i>		191	534	904	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza rufipogon</i>		193	544	923	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza sativa indica</i>		196	555	984	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Oryza sativa (REFERENCE)</i>		222*	1025	1173	Ensembl Gramene	Uniprot	Ensembl Compara			
	<i>Setaria italica</i>		191	531	962	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Sorghum bicolor</i>		192	536	929	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Triticum aestivum</i>		193	531	2159	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	<i>Triticum turgidum</i> *		197	528	2388	PMID: 23800085	N/A	Planteome Inparanoid			
	<i>Triticum urartu</i>	187	481	768	Ensembl Gramene	Ensembl Gramene	Ensembl Compara				
	<i>Zea mays</i>	191	528	1102	Ensembl Gramene	MaizeGDB	Ensembl Compara				
	Zingiberales	<i>Musa acuminata</i>	186	490	1141	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	Areciales	<i>Phoenix dactylifera</i>	191	497	922	PMID: 23917264	N/A	Planteome Inparanoid			
	Amborellae	<i>Amborella trichopoda</i>	189	499	617	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	Gymnosperm	<i>Picea abies</i>	191	477	1119	Congenie	Congenie	Planteome Inparanoid			
		<i>Pinus taeda</i>	181	434	1088	TreeGenes	N/A	Planteome Inparanoid			
	Lycopod	<i>Selaginella moellendorffii</i>	181	472	1108	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	Bryophyte	<i>Physcomitrella patens</i>	176	454	904	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	Chlorophyte	<i>Ostreococcus lucimarinus</i>	131	270	220	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
		<i>Chlamydomonas reinhardtii</i>	141	287	233	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			
	Rhodophyte	<i>Cyanidioschyzon merolae</i>	116	208	158	Ensembl Gramene	Ensembl Gramene	Ensembl Compara			

The columns, pathways, reactions and gene products show counts determined after projecting similar events based on the homology to *O. sativa* (rice) reference gene and pathway annotations. The conditional formatting of each respective column shows shades of red trending towards a higher count versus the shades of blue-trending towards lower counts. The conditional formatting ignored the reference *O. sativa* counts. Also provided for each species is the name of the external source or repository from where the gene sequence dataset was downloaded, the external source to which the pathway associated genes point to and the method applied for building the gene homology data. For more details see Supplementary file-1. * Transcriptome assembly data.

model standards. The dataset for 12 rice metabolic pathways was also inferred based on gene homology from the human Reactome (2) that includes evolutionary conserved regulatory processes, such as cell growth, cell cycle, DNA replication/repair, etc. As protein functions have not been systematically and experimentally characterized for large number of genes from many plant species, we use an integrated approach for curating the rice and non-rice literature to build the reference datasets. These reference data sets are then used to project homologous events in 62 model, crop and evolutionary significant plant species. The combination of pathway analysis tools, annotations, gene ex-

pression data and the search and browse functionalities enable plant researchers to review, reuse, discover and build stochastic models for further studies in their species of interest. Achieved by investigating the molecular interactions, gene expression under various conditions, and evolution of pathways, reactions and small molecules, thus moving the research beyond the realms of genome and transcriptome sequencing. The database extensively collaborates with various plant genomics projects such as Gramene's Ensembl genome portal (1), Phytozome (4), PeanutBase (5), TreeGenes (6), Genome Database for Rosaceae (7), SolGenomics Network (8), MaizeGDB (9), TAIR (10), AraPort

(11), Legume Information System (12) and Planteome (13) (Table 1). The collaboration extends to large online resource providers such as Gene Ontology (14), EMBL-EBI's Gene Expression Atlas (15), ChEBI (16), PubMed, UniProt (17) and NCBI, in order to share data and extend its linkages (Supplementary Figure S2). The database also provides access to data in various community-wide standardized pathway formats and via Application Programming Interfaces (APIs).

PLANT REACTOME RESOURCE

Since its first beta release in January 2013, Plant Reactome has grown to include primary reference annotations in the form of 222 pathways and 1025 reactions associated with 1173 rice gene products. We continue to curate rice pathways manually and also use automated methods and scripts that integrate data related to structure and function of proteins, their subcellular locations, the expression of protein-coding genes, and diversity of small molecule and metabolite from the external resources mentioned above. In addition to these reference rice pathways, the database provides gene homology-based pathway projections for 62 other plant species (Table 1) as described later. The Plant Reactome allows researchers to query and visualize pathways in species of their choice and analyze their genome-scale expression data for pathway enrichment analysis to identify differentially expressed pathways and genes. Users can also compare projected pathways with reference rice pathways and visualize curated baseline and differential expression data for all pathway-associated genes fetched programmatically from EMBL-EBI's Expression Atlas (15).

The database homepage, accessible from <http://plantreactome.gramene.org>, provides quick links to the pathway browser, data analysis tools and download options, user guide, video tutorials, database release summary, news and documentation about the project, data model, schema and APIs. (Supplementary Figure S1). The homepage also provides direct access to the primary database search feature.

We will use an example of the jasmonic acid (JA) signaling pathway and the *COII* (*CORONATINE INSENSITIVE1*) gene from the reference species rice and its projections on other angiosperm species like soybean, tomato, maize and non-angiosperms lycopod *Selaginella* and unicellular green algae *Chlamydomonas*. We will use this example to describe the development and various functionalities of the Plant Reactome resource including analysis tools for use in research, education, and training. Some studies on Plant *COII* gene family members are reported. The *A. thaliana COI* gene encodes an F-Box domain containing protein that binds to jasmonate-isoleucine (JA-Ile)/coronatine (COR) and acts as a jasmonic acid (JA) receptor, thus playing a role in the JA signaling pathway (18,19). COR is a phytotoxin produced by the fungus *Pseudomonas syringae* a well-known pathogen of crop plants (19). The rice *COII* gene is known for its role in immunity, plant development and senescence (20,21). Rice has three gene family members and the *COIIb* mutant stayed green in the dark-induced natural senescence and showed a decrease in spikelet fertility and grain weight leading to lower yield

(22). The tomato homolog is essential for seed maturation, immunity, and glandular trichome development (23,24), whereas soybean *COI* homolog complements and restores *Arabidopsis COI* function (25). Studies also showed that *COII* plays an interactive role with abscisic acid signaling in regulating the stomatal closure and ion channel activation in response to drought (26,27). To investigate if *COII* expression is affected by drought conditions, we selected a *Glycine max* (soybean) RNA-seq based drought transcriptome study as an example for OMICs data analysis.

PLANT REACTOME SEARCH

The quick search box is powered by the Solr search platform (<http://lucene.apache.org/solr/>), designed for highly scalable, load-balanced, and dynamically indexed full-text searches. The Plant Reactome implementation makes particular use of the auto-completion feature and faceted search result filters, which aid the user with search suggestions and the ability to narrow down results with precision. The quick search tool on the right-most upper side of Plant Reactome homepage (<http://plantreactome.gramene.org/>) allows users to search for any small molecules including metabolites, proteins, protein complexes, enzymes, reactions and pathways in all 63 species, including the reference species *Oryza sativa* (rice). For example, querying for the *COII* (Figure 1A) gene returns results showing that, if selected for reference *O. sativa*, there are five entities in the database that either bears the name *COII* or associated with the gene set called *COII*. The list of facets with check boxes in the left-hand menu allows additional query filters. One has an option to go further for a detailed view (Figure 1B) of the search on the *COII* set from rice. The top panel shows event hierarchy and provides links to the Pathway Browser (Figure 1B), and the bottom panel suggests associated cellular component locations, three rice gene products that are part of the *COII* set, reaction components, and a list of similar events projected in other species inferred by gene homology-based projection. In this query result, users also learn that three rice gene products are part of the *COII* set.

The additional small molecule search available from http://plantreactome.gramene.org/cgi-bin/small_molecule_search allows searches of various chemical compound entities. The query form allows the user to illustrate a molecule and submit searches to the remote ChEBI database of small molecules (16). Resulting matches connect to the same molecule in the Plant Reactome database. For example, a search for C₂H₄ finds a match to Ethene or Ethylene (CHEBI:18153) and a link to the pathways involving that molecule in the database.

PATHWAY BROWSER

The Plant Reactome utilizes the conceptual framework of a eukaryotic plant cell to annotate the interactions of small molecules including metabolites, enzymes, and regulators as reactions and pathways. The pathways, in turn, are organized around a hierarchical classification schema similar to the Gene Ontology (GO) that allows building relationships between pathways (e.g. grouping of similar pathways, parent-child relationship, connections between interacting

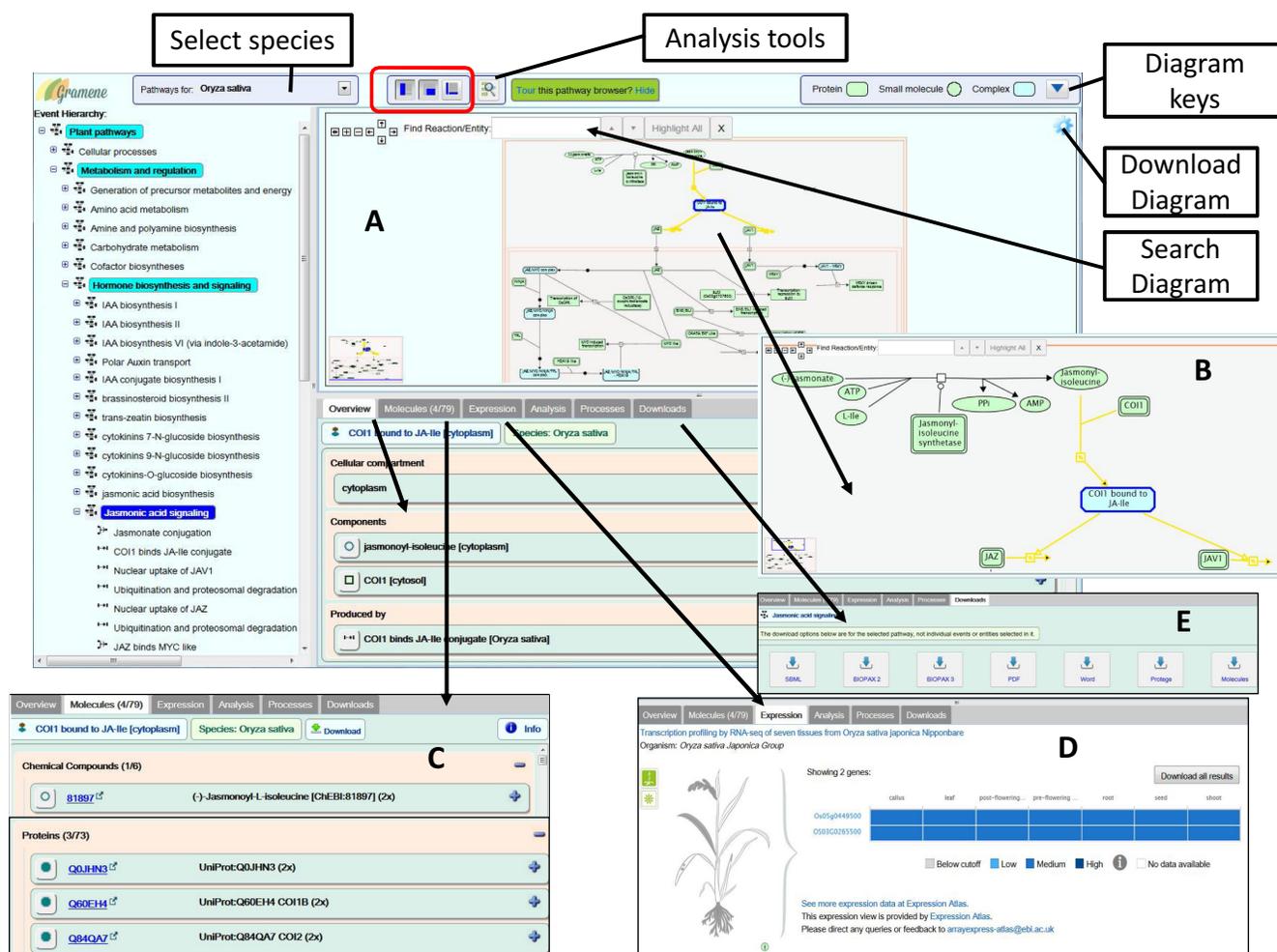


Figure 2. A view of the Plant Reactome Pathway Browser details. (A) An overview of the selected *O. sativa* (rice) jasmonic acid (JA) signaling pathway. Zoom in and out feature (red outlined section) allows adjusting the diagram views, such as the zoomed-in view (B) of the reaction involving jasmonic acid. (C) The Molecules tab provides a list of interacting molecules. (D) Depending on the selected entities in the diagram, the expression tab displays baseline (RNA-seq based) gene expression data analyzed and served by the Gene Expression Atlas. (E) Download the selected entities data in various standard formats, including SBML and BioPAX.

22 June 2016, release of the EMBL-EBI Gene Expression Atlas includes 671 gene expression dataset from fourteen plant species. Baseline expression data generated from the RNA-seq transcriptome studies provide estimates of transcript abundance for each gene across various experimental conditions, e.g. tissues, cell types and development stages. The differential expression data scored for each experimental study provides information on the differential expression of a given gene or a gene set between two conditions including the treatments as previously described (15).

All three panels in the pathway browser window are tightly linked and clicking on any of the panels changes the view of other panels displaying associated features and data. The icon pointing to the Analysis tools options is also available on top of the pathway browser page (Figure 2A) as well as the homepage of the project website and described in the respective section (Figure 3).

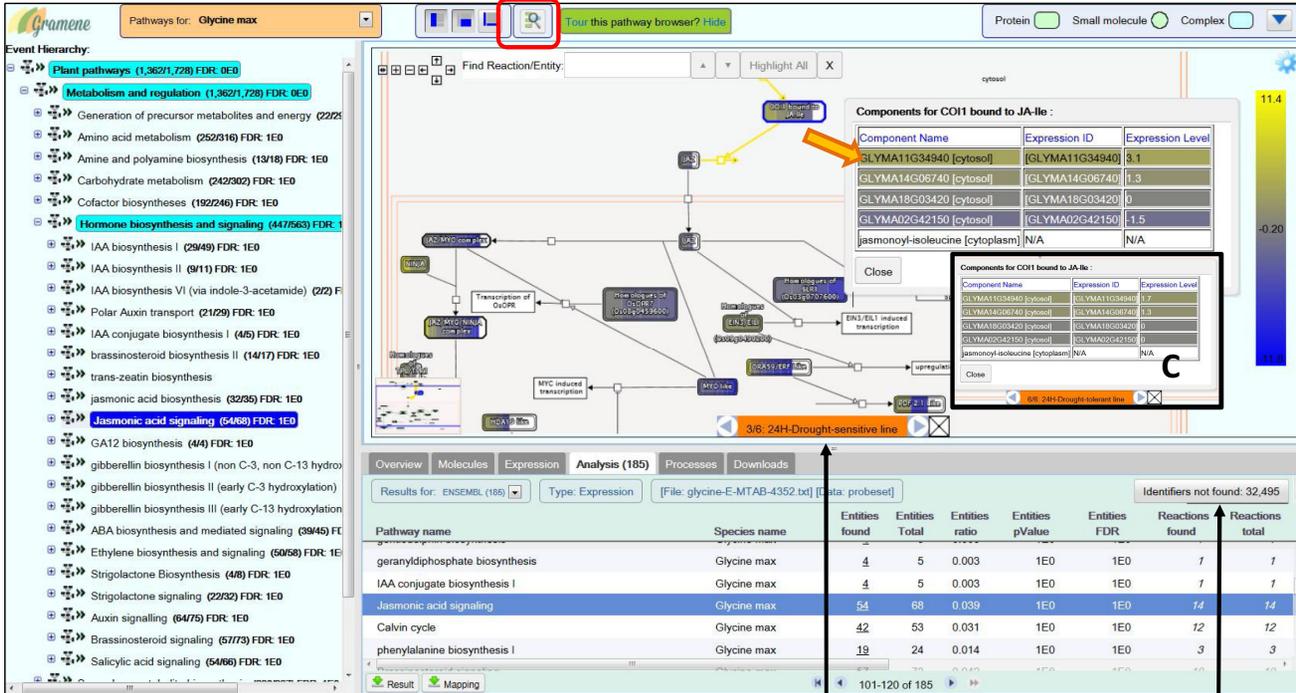
The information on selected entities data is downloadable in various standard formats, including SBML and BioPAX, underneath the Download tab (Figure 2E).

PATHWAY PROJECTIONS FOR PLANT GENOMES

The human Reactome database (2) allows the construction of gene homology-based pre-computed pathway network projections from a single reference species to other species. We used a similar approach, to construct pathway projections for 62 plant species (<http://plantreactome.gramene.org/stats.html>), which were generated using rice as a reference species (Table 1). Homology data for 38 of these species were accessed from Gramene's Ensembl Compara (1,28), while the remaining 24 species were generated in-house by comparing them with *O. sativa* reference annotations via the InParanoid homology clustering method developed by us (29–31). We extract these species homologs for *O. sativa* from a local InParanoid MySQL database. Both sets of homology data, formatted as flat files containing one-to-many pairs of homologous gene locus identifiers, are then fed to a Reactome pathway projection Perl script that generates new gene product, reaction, and pathway data in the Plant Reactome database. Table 1 provides a complete summary of the

OMICs Data Analysis Tool A

B



Events with Enrichment Statistics

Download pathway Enrichment data and mapped gene loci

Use arrows to view expression from different data point e.g. (C)

Data for unmapped gene loci

Figure 3. Plant Reactome pathway analysis tools interface: OMICs data analysis. (A) The analysis tool is accessible from the project website's homepage (<http://plantreactome.gramene.org>) by clicking on the Analyze Data icon or if the users are on the pathway browser click on the Analyze your data icon (red box). The analysis tools interface replaces the Pathway diagram view in the pathway browser. Users can input their data either by (a) uploading the data file (this figure) or (b) by copying the data into the text box (Supplementary Figure S3). After selecting the data, click on the Analyze button and the pathway, enrichment analysis data is returned in the tabulated form (B) in the Analysis tab. Select the event from the left-hand side panel to view the diagram with expression data. By hovering the computer mouse over the colored genes sets and clicking the blue info icon a list of mapped entities with painted expression values is displayed. Left or the right arrow next to the orange text box displays expression from a different data column (C).

projected pathways, reactions, and associated homologous gene counts. The counts and projections exclude the original 12 pathways initially projected for rice using the human Reactome annotations as described above. The availability of projected pathways and reactions are unlike the KEGG (32) or BioCyc-based (33) projections in that Plant Reactome will only display the individual reactions of a pathway associated with gene product activity. The projected data for any of these species is accessible to users with full functionalities described in this publication.

PATHWAY ANALYSIS TOOLS

Moving beyond differential gene expression, ontology-based enrichment and genome annotations, researchers are

often looking for ways to analyze their data for making inferences and building a hypothesis based on molecular interactions and pathways. Therefore, Plant Reactome provides two types of analysis tools for its users: (i) pathway enrichment and overlay visualization of user-provided data from OMICs experiments (Figure 3 and Supplementary Figure S3) and (ii) interspecies pathway comparison (Figure 4). Both tools described in the following sections provide data analysis, visual insights and data downloads. Users can access the Analysis Tools from the Plant Reactome homepage (Supplementary Figure S1) by clicking on the Analyze Data icon or from the icon in the header of the Pathway Browser (Figure 2).

Pathway analysis using OMICs Data

The Analysis Tools allow uploading, visualization and analysis of user-defined high-throughput OMICs data (e.g. transcriptome, proteome, metabolome, etc.) in the context of Plant Reactome pathways. The analysis tool does not accept raw data, and users need to process their data in a tab-delimited file format before uploading the data file to conduct pathway enrichment analysis (Figure 3A). The users can either upload the data file or copy the data in the text box (Supplementary Figure S3) and may choose to project their data against only the reference *O. sativa* (rice) or against another species. Before opting for the latter select the appropriate species from the drop-down menu, in this case, *Glycine max*. To illustrate the utility of the OMICs analysis tools, we used the pre-analyzed and publicly available RNA-seq based transcriptomic data (E-MTAB-4352; source: Gene Expression Atlas (15)) generated from the two accessions of *G. max* (soybean) that show different phenotype response to drought treatment (SUPPLEMENTARY FILE-3). After selecting the data, clicking on the Analyze button initiates the analysis/process. Depending on the size of the data, and mapped entities in the database (e.g. Gene IDs provided by the Source database (Table 1) and UniProt IDs for rice only, metabolite name or IDs, etc.), the pathway enrichment analysis data is returned in the tabulated form (Figure 3B). Other tabs continue to provide the data mentioned in Figure 2. Users would also notice that the event hierarchy panel on the left-hand side now shows the data statistics next to the pathway event because all the events recorded in the Plant Reactome database for the selected species, in this case, *G. max* are included in the analysis. Users can choose the pathway event from the left-hand side panel and the tabulated data in the Analysis tab. For example, the jasmonic acid signaling pathway showing 54 genes from the user input data mapped to the projected soybean JA signaling pathway with a total of 68 genes and 14 reactions. Users also get options to download the mapped, unmapped and analyzed data for later use in publications. If the input file has multiple data columns, by default, it displays data from the first data column in the overlaid views. Change it by clicking the left or the right arrows on either side of the orange highlighted text box at the bottom of the pathway diagram.

Visualize the expression data corresponding to the various genes mapped to a given pathway in the pathway diagram panel by selecting the gene set or an event and clicking the blue info icon. A dialogue box opens listing the gene homologs or gene family members mapped to the given event and overlays their expression profile (inset: Components for COI1 bound to JA-Ile). For example, at the 24-h time point after drought treatment in sensitive soybean accession (main figure) compared to the expression of same genes at the same time point after drought treatment in the tolerant soybean line (Figure 3C). Although the gene duplicates are likely to encode similar proteins, their expression and functions may differ across the tissues, organ and cells during plant development as well as in response to stimuli and environment (9,14,16–20).

Our analysis results suggest that the soybean *COI1* homolog GLYMA11G34940 mapped to the *COI1*-JA-Ile

complex shows about two-fold increased expression in response to the drought treatment in the tolerant line. The susceptible line shows downregulation of the other homolog GLYMA0-2G42150, whereas, GLYMA14G06740 and GLYMA18G03420 do not indicate any change in their expression when compared between the two lines. Although transcript abundance scored in this data may not provide a direct correlation to the translated protein function and abundance, the observation suggests that at least two *COI1* gene family members in soybean show altered expression at transcript level under drought response.

OMICs data analysis views like these are a simple way to begin building hypotheses searching for candidate genes that show altered transcript abundance in phenotypically different accessions in response to the same treatment. Such analyses have the potential to become powerful tools for deciphering and discovering candidate genes regulating the important agronomic traits like abiotic and biotic stress resistance or tolerance, development, anatomy, physiology, yield and nutrition quality. It applies in particular when these inferences are combined with ontology-based gene function annotations, overlapping genetic markers, SNPs, QTLs and syntenic alignments to closely related genomes, thus helping researchers formulate new hypotheses using a systems biology approach and narrowing down the scale of experimental tests.

Interspecific pathway comparison

The second type of analysis tool provided by the Plant Reactome allows comparison of events such as pathways and reactions between the reference species rice and any of the other species by way of mapped gene entities. See analysis steps shown in Supplementary Figure S4 using rice to grape comparison. In another example, selecting the species of choice, for example, *Zea mays* (maize) and clicking on the Compare button, the analyzed results are displayed and available for download from the Analysis tab of the pathway browser window described earlier. In this analysis, clicking on the hyperlinked event name in the event hierarchy panel or those in the analysis results table, for example, the jasmonic acid signaling pathway, will render a reference rice pathway diagram overlaid with homolog mappings from the selected species (Figure 4), in this case *Z. mays* (maize) with colored gene set boxes—suggesting either a respective rice gene homolog was found (yellow color) or not found (blue). In the case of multiple rice genes associated with the reaction, users may see both yellow and blue colors depending on the number of homolog matches between the species, which also correspond to the proportion of the yellow or blue color. Hovering the computer mouse over the colored box and clicking the blue info icon pops an inset showing the detailed gene list with color as shown in the Figure 4. Views such as these provide insights into the evolution of gene families, pathways, and reactions. For example, we observe that the jasmonic acid signaling pathway show likely conservation for all the curated events with at least one or more rice gene homologs present (yellow color) in the closely related maize, also a monocot (Figure 4A), including the components for *COI1* bound to JA-Ile. A similar comparison with tomato (Figure 4B), a dicot and a non-

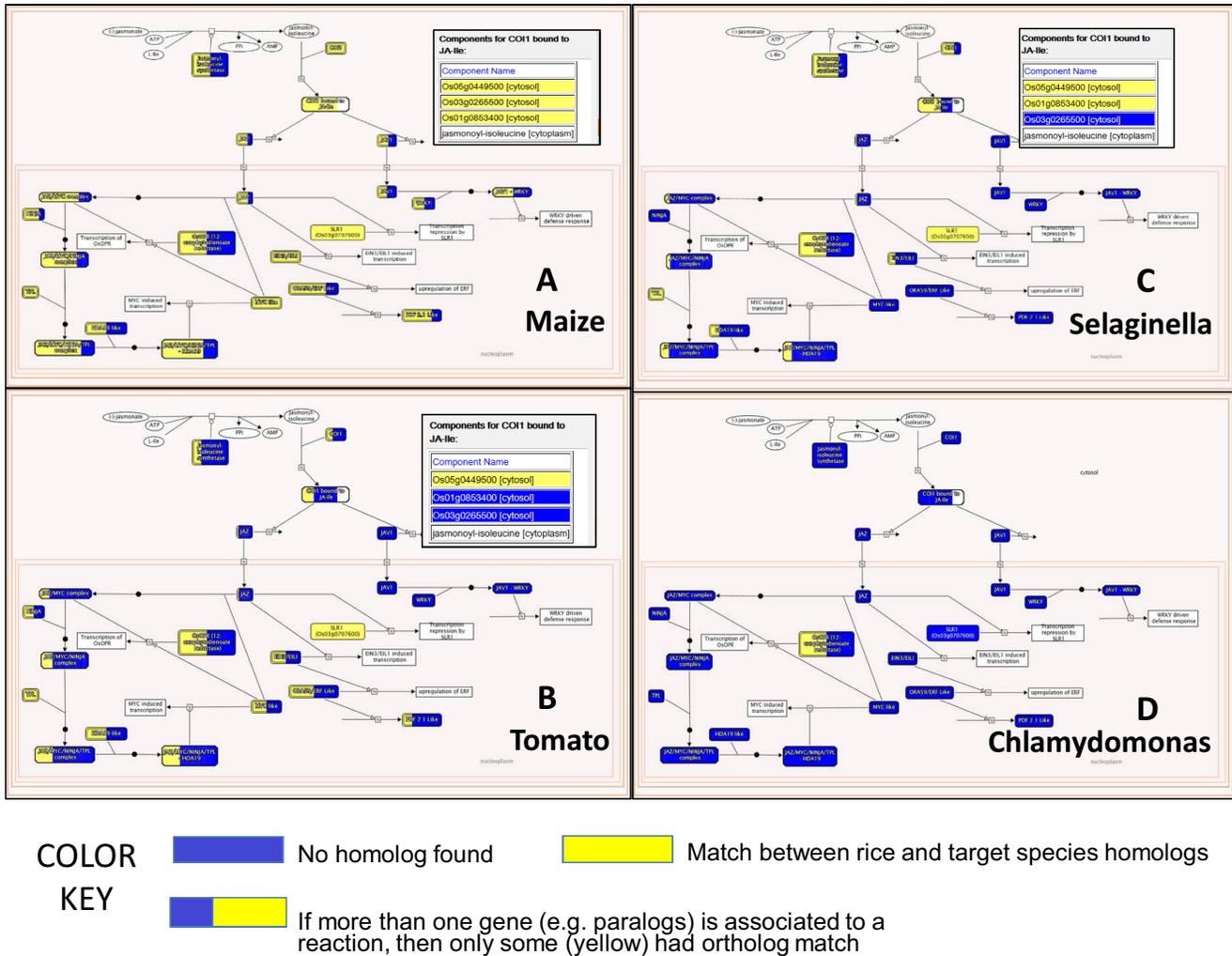


Figure 4. Plant Reactome pathway analysis tools interface: Interspecific pathway comparative views of the jasmonic acid signaling pathway when compared to angiosperms maize (A) and tomato (B), lycopod *Selaginella* (C) and unicellular green algae *Chlamydomonas* (D) based on the respective gene homology matches with the rice reference. Colored boxes suggest that either a respective rice gene homolog was found (yellow color) or not (blue). The reaction associated with multiple rice genes may display both yellow and blue colors depending on the number of homolog matches which also corresponds to the proportion of the yellow or blue color. The inset showing Components for COI1 bound to JA-Ile suggests which of the rice genes have homologs mapped from the respective species.

angiosperm basal vascular plant *Selaginella* (Figure 4C), a lycopod, show some degree of conservation, but also no homologs (all blue boxes) for certain reactions downstream of the JA-Ile complex formation stage. Similarly, *Chlamydomonas* (Figure 4D), green algae from an aquatic habitat, on the other hand, is probably missing homologs for a majority of the genes. These views provide clues for building a hypothesis that either certain parts of the JA-signaling pathway may lack support in tomato and *Selaginella* or need better annotation. A similar observation for *Chlamydomonas* suggests that the pathway may not be required for the unicellular algae to sustain in the aquatic habitat, even though the precursors for the biosynthesis of JA may exist in *Chlamydomonas* (34) and are known to be induced by salt stress. Though these inferences need to be tested, we strongly recommend researchers perform a second set of data mining and gene homology analysis elsewhere. De-

pending on the version and quality of the species-specific genome annotation and the gene sets acquired from the source archive (Table 1) as well as the applied stringency of the gene homology prediction methods (Supplementary file-1), data may show variation in the projection of homologous events.

DATA INTEGRATION, CURATION AND DATABASE DEVELOPMENT

The Plant Reactome was initiated in 2012 using rice as a reference species. All the curation is accomplished using manual and semi-automated methods. We integrate curated data in the central pathway curation database maintained by the human Reactome project (2). Thus we share the same small molecule data archive, data models, review process, database servers, curation and database production tools. The reference rice pathways were first created by adopting

metabolic pathways from RiceCyc (3), which were then imported in bulk into the central pathway curation database in the BioPAX level-2 format using import tools provided by the Reactome project. Subsequently, various other types of pathways involved in hormone signaling, development, transport, and stress response were curated by mining the published literature and using the Reactome Curator Tool. At time of release, the reference pathways checked into the central pathway database and marked for release are sliced out. These rice reference pathways serve as templates for computationally projecting systems-level pathway networks for 62 other species by identifying corresponding homologs of rice genes (Table 1) using the method described in the Supplementary File-1.

It is noteworthy that projected pathways based on the transcriptomes and some sequenced plant genomes are not available in any of the online annotated genomics resources including the Gramene Ensembl, Phytozome and NCBI databases (Table 1). For example, annotated leaf transcriptomes of wild *Oryza* species (provided by the OMAP project) (35) and the Kasalath rice genome (36) a representative of *O. sativa* AUS clade (Table 1), were used for deducing the projected pathways in the corresponding species using our in-house InParanoid clustered analysis (29–31). We managed to project pathways for these representative species because the Reactome data model allows inclusion of sequenced entities from a sequenced transcriptome, genome or for that matter a proteome. We included all the transcript isoforms from the assembled transcriptomes and did not make an effort to find the canonical longest isoform or the unigene. Similarly, polyploid gene annotations were not excluded and thus researchers may see an increased number of associated genes or transcript isoforms (Table 1) like the associated gene family members and paralogs for diploid species. Species such as *O. sativa* Aus Kasalath show lower projection counts, perhaps due to the lower coverage of the sequenced genome (36) or artifacts of the genome annotation quality. However, lower pathway counts for non-Angiosperm species are reflective of the evolutionary tree of life for plants.

DATA DOWNLOAD AND WEBSERVICES

We facilitate download of various data sets associated with the plant pathways (Figure 2). Users can download pathway data in community standard formats such as SBML, BioPAX level-2 and level-3, as well as in the Protégé OWL, PDF and Word formats. Users can also access a hierarchical list of pathways and reactions in a particular species, and download a list of small molecules and gene products participating in a particular pathway. Full downloads of the latest released database (in MySQL) are also available, along with tab-delimited files containing (i) all curated and projected gene product identifiers grouped by pathway and species, and (ii) all projected gene products grouped by pathway and reaction and indexed to curated *Oryza sativa* gene products (Supplementary Figure S1). These dataset are helpful if an external source is interested in establishing reciprocal links to the Plant Reactome entities as well as using the embedded identifiers for building Application Programming Interface (API)-based queries.

Application developers and advanced users may access Plant Reactome data via its API deployed with each release. This API is built on the work of the human Reactome project. It is used internally to provide data for the web application and is also provided externally as a service to the community. Many methods can be called over the HTTP protocol from a web browser, or directly from the command line; others can be invoked using client-side Javascript. Documentation for the API is available in the Reactome Developer Guide. Follow the same instructions for forming URLs, but replace the domain name 'reactome.org' suggested in the Reactome documentation with 'plantreactome.gramene.org', and then invoke the RESTful web service methods. For details consult Supplementary file-2.

OUTREACH AND TUTORIALS

To assist the plant research community in exploring resources and tools available at Plant Reactome, we regularly organize online webinars. Using case studies from a variety of crops and model plant species, we show how to use the available resources, data, and tools for plant comparative genomics and pathway analysis. The recorded webinars also serve as video tutorials and can be found on Gramene's YouTube channel (<https://goo.gl/qQ2Pjn>).

CONCLUSION

The Plant Reactome is a network of plant proteins linked to their molecular functions and interactors in the context of their locations within the plant cell. It serves as an archive of plant pathways (with all associated components) as well as a conceptual framework that can facilitate the analysis of genome-scale expression data. Thus, it enables researchers to discover previously unknown interactions that drive a plant's essential developmental genetic program and/or mitigate stress signals from the environment. The homology-based projections cover a broad taxonomic range within the plant kingdom including models and crop species from monocots and dicots represented by various grasses, fruits, legumes and bioenergy feedstock model crops, namely *Brachypodium*, *Setaria*, *Sorghum*, Poplar and *Jatropha*, as well as the basal angiosperm *Amborella trichopoda*. The database extends its pathway projections for non-Angiosperm species represented by gymnosperms *Pinus taeda* (Loblolly pine) and *Pinus abies* (Norway spruce), lycopod *Selaginella moellendorffii*, moss *Physcomitrella patens*, green algae *Chlamydomonas reinhardtii* and *Ostreococcus lucimarinus* and red algae *Cyanidioschyzon merolae*, which are included for their evolutionary importance (Table 1). Therefore, the Plant Reactome is a unique resource and provider of pathway collection and analysis tools.

Currently, the curation of reference pathways is ongoing, and we continue to develop projections for more plant species with each database release. We update our database 3–5 times per year, and post the current statistics and notes on new or updated data, tools and website content at <http://plantreactome.gramene.org/pages/content/release-summary>. Our plans include adopting the Reactome 3.5 or

later platform, staying synchronized with future Reactome code releases, projections for new plant species, and adding new pathways for response to biotic and abiotic stresses, development, transport, and metabolism.

We encourage feedback from the community of plant genomic researchers about the Plant Reactome data, our website, downloads, online webinars and recorded tutorials. We are especially looking for help in reviewing our curated pathways and identifying new plant species with genome and transcriptome data for which plant biology researchers would like to have pathway projections.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Cold Spring Harbor Laboratory's Gramene project members Dr. Marcela Tello-Ruiz, Joe Mulvaney, Andrew Olson and Kapeel Chougale for coordinating the updates to the Gramene database searches and its Ensembl Plant genome browser and the Dolan DNA Learning Center for help with the organization of Plant Reactome webinars. We also acknowledge Ontario Institute for Cancer Research for providing access to the Reactome central curation database and hosting our development and live infrastructure support. Christopher Sullivan from the Center for Genome Research and Biocomputing at Oregon State University provided timely support for local infrastructure needs. We also thank Robin Haw, Sheldon McKay and David Croft from the Reactome project for providing timely technical assistance on Plant Reactome development. The pathway curation help provided by Oregon State University undergraduate students Dylan Beorchia, Teague Green, Kindra Amoss and Christina Partipilo is greatly appreciated. The authors are grateful to our users, researchers and numerous collaborators for sharing datasets generated in their projects and for valuable suggestions and feedback on improving the overall quality of the database.

Authors contributions: P.J., D.W., P.D., L.S. envisioned and proposed the project. P.J. leads the Plant Reactome project as Co-PI of the Gramene Project led by the P.I. D.W. P.J., S.N., P.G., V.A., P.D.D. curated the pathways, reviewed by P.D. P.D., chief editor of the Human Reactome, also provided training for Plant Reactome curators. J.P. is the lead Bioinformatics software and database developer for the project. G.M. carried out the initial RiceCyc import and provided timely support on curation fixes, data standards, and data model. J.W. and A.F. provided the timely support on running the Plant Reactome functionalities and database build. L.S. provided the infrastructure support from OICR and human Reactome project. M.K., A.M.P.F., R.P. curated the Gene Expression Atlas data and provided programmatic access to the data visualization widget. J.E. provided local infrastructure support and contributed the InParanoid gene homology dataset. The manuscript was written by S.N., P.G., J.P., P.D., J.E. and P.J. and reviewed by everyone. Funding agencies had no role in the study design, data analysis, or preparation of the manuscript.

FUNDING

The Gramene database Award [NSF IOS-1127112] supported the project with in-kind infrastructure and intellectual support from the human Reactome database project [NIH: P41 HG003751, ENFIN LSHG-CT-2005-518254, Ontario Research Fund, and EBI Industry Programme]; The Planteome Project [NSF IOS:1340112] provided InParanoid gene homology dataset. Funding for open access charge: The Gramene database Award [NSF IOS-1127112]. *Conflict of interest statement.* None declared.

REFERENCES

- Tello-Ruiz, M.K., Stein, J., Wei, S., Youens-Clark, K., Jaiswal, P. and Ware, D. (2016) Gramene: a resource for comparative analysis of plants genomes and pathways. *Methods Mol. Biol.*, **1374**, 141–163.
- Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S. et al. (2016) The Reactome pathway Knowledgebase. *Nucleic Acids Res.*, **44**, D481–D487.
- Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D., McCouch, S., Ware, D. and Jaiswal, P. (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice (N. Y.)*, **6**, 15.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.
- Dash, S. and Cannon, E.K.S. (2016) Chapter 8 - PeanutBase and Other Bioinformatic Resources for Peanut A2 - Stalker, H. Thomas, In: Wilson, R.F. (ed.). *Peanuts*, AOCS Press, 241–252.
- Wegrzyn, J.L., Lee, J.M., Tearse, B.R. and Neale, D.B. (2008) TreeGenes: a forest tree genome database. *Int. J. Plant Genomics*, 412875.
- Jung, S., Ficklin, S.P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., Yu, J., Bombarely, A., Cho, I., Ru, S. et al. (2014) The Genome Database for Rosaceae (GDR): year 10 update. *Nucleic Acids Res.*, **42**, D1237–D1244.
- Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H. et al. (2015) The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. *Nucleic Acids Res.*, **43**, D1036–D1041.
- Andorf, C.M., Cannon, E.K., Portwood, J.L. II, Gardiner, J.M., Harper, L.C., Schaeffer, M.L., Braun, B.L., Campbell, D.A., Vinnakota, A.G., Sribalasu, V.V. et al. (2016) MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res.*, **44**, D1195–D1201.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M. et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.
- Krishnakumar, V., Hanlon, M.R., Contrino, S., Ferlanti, E.S., Karamycheva, S., Kim, M., Rosen, B.D., Cheng, C.Y., Moreira, W., Mock, S.A. et al. (2015) Araport: the Arabidopsis information portal. *Nucleic Acids Res.*, **43**, D1003–D1009.
- Dash, S., Campbell, J.D., Cannon, E.K., Cleary, A.M., Huang, W., Kalberer, S.R., Karingula, V., Rice, A.G., Singh, J., Umale, P.E. et al. (2016) Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.*, **44**, D1181–D1188.
- Cooper, L., Walls, R.L., Elser, J., Gandolfo, M.A., Stevenson, D.W., Smith, B., Preece, J., Athreya, B., Mungall, C.J., Rensing, S. et al. (2013) The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.*, **54**, e1.
- Gene Ontology, C. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
- Petryszak, R., Keays, M., Tang, Y.A., Fonseca, N.A., Barrera, E., Burdett, T., Füllgrabe, A., Fuentes, A.M., Jupp, S., Koskinen, S. et al. (2016) Expression Atlas update—an integrated database of gene and

- protein expression in humans, animals and plants. *Nucleic Acids Res.*, **44**, D746–D752.
16. Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P. and Steinbeck, C. (2016) ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
 17. UniProt, C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
 18. Yan, J., Li, H., Li, S., Yao, R., Deng, H., Xie, Q. and Xie, D. (2013) The Arabidopsis F-box protein CORONATINE INSENSITIVE1 is stabilized by SCFCO11 and degraded via the 26S proteasome pathway. *Plant Cell*, **25**, 486–498.
 19. Katsir, L., Schilmiller, A.L., Staswick, P.E., He, S.Y. and Howe, G.A. (2008) COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 7100–7105.
 20. Lee, H.Y., Seo, J.S., Cho, J.H., Jung, H., Kim, J.K., Lee, J.S., Rhee, S. and Do Choi, Y. (2013) *Oryza sativa* COI homologues restore jasmonate signal transduction in Arabidopsis coi1-1 mutants. *PLoS One*, **8**, e52802.
 21. Ye, M., Luo, S.M., Xie, J.F., Li, Y.F., Xu, T., Liu, Y., Song, Y.Y., Zhu-Salzman, K. and Zeng, R.S. (2012) silencing COI1 in rice increases susceptibility to chewing insects and impairs inducible defense. *PLoS One*, **7**, e36214.
 22. Lee, S.H., Sakuraba, Y., Lee, T., Kim, K.W., An, G., Lee, H.Y. and Paek, N.C. (2015) Mutation of *Oryza sativa* CORONATINE INSENSITIVE 1b (OsCOI1b) delays leaf senescence. *J. Integr. Plant Biol.*, **57**, 562–576.
 23. Li, C., Zhao, J., Jiang, H., Wu, X., Sun, J., Zhang, C., Wang, X., Lou, Y. and Li, C. (2006) The wound response mutant suppressor of prosystemin-mediated responses6 (spr6) is a weak allele of the tomato homolog of CORONATINE-INSENSITIVE1 (COI1). *Plant Cell Physiol.*, **47**, 653–663.
 24. Li, L., Zhao, Y., McCaig, B.C., Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E. and Howe, G.A. (2004) The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. *Plant Cell*, **16**, 126–143.
 25. Wang, Z., Dai, L., Jiang, Z., Peng, W., Zhang, L., Wang, G. and Xie, D. (2005) GmCOI1, a soybean F-box protein gene, shows ability to mediate jasmonate-regulated plant defense and fertility in Arabidopsis. *Mol. Plant Microbe Interact.*, **18**, 1285–1295.
 26. Harb, A., Krishnan, A., Ambavaram, M.M. and Pereira, A. (2010) Molecular and physiological analysis of drought stress in Arabidopsis reveals early responses leading to acclimation in plant growth. *Plant Physiol.*, **154**, 1254–1271.
 27. Munemasa, S., Oda, K., Watanabe-Sugimoto, M., Nakamura, Y., Shimoishi, Y. and Murata, Y. (2007) The coronatine-insensitive 1 mutation reveals the hormonal signaling interaction between abscisic acid and methyl jasmonate in Arabidopsis guard cells. Specific impairment of ion channel activation and second messenger production. *Plant Physiol.*, **143**, 1398–1407.
 28. Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R. and Birney, E. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, **19**, 327–335.
 29. Myburg, A.A., Grattapaglia, D., Tuskan, G.A., Hellsten, U., Hayes, R.D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D. *et al.* (2014) The genome of *Eucalyptus grandis*. *Nature*, **510**, 356–362.
 30. Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P. *et al.* (2011) The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.*, **43**, 109–116.
 31. Sonhammer, E.L. and Ostlund, G. (2015) InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.*, **43**, D234–D239.
 32. Kanehisa, M. (2016) KEGG Bioinformatics Resource for Plant Genomics and Metabolomics. *Methods Mol. Biol.*, **1374**, 55–70.
 33. Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **44**, D471–D480.
 34. Arisz, S.A. and Munnik, T. (2011) The salt stress-induced LPA response in *Chlamydomonas* is produced via PLA2 hydrolysis of DGK-generated phosphatidic acid. *J. Lipid Res.*, **52**, 2012–2020.
 35. Wing, R.A., Ammiraju, J.S., Luo, M., Kim, H., Yu, Y., Kudrna, D., Goicoechea, J.L., Wang, W., Nelson, W., Rao, K. *et al.* (2005) The oryza map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Biol.*, **59**, 53–62.
 36. Sakai, H., Kanamori, H., Arai-Kichise, Y., Shibata-Hatta, M., Ebana, K., Oono, Y., Kurita, K., Fujisawa, H., Katagiri, S., Mukai, Y. *et al.* (2014) Construction of pseudomolecule sequences of the aus rice cultivar Kasalath for comparative genomics of Asian cultivated rice. *DNA Res.*, **21**, 397–405.