Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx1111

Gramene 2018: unifying comparative genomics and pathway resources for plant research

Marcela K. Tello-Ruiz[1], Sushma Naithani[2], Joshua C. Stein[1], Parul Gupta[2], Michael Campbell[1], Andrew Olson[1], Sharon Wei[1], Justin Preece[2], Matthew J. Geniza[2], Yinping Jiao[1], Young Koung Lee[1,3], Bo Wang[1], Joseph Mulvaney[1], Kapeel Chougule[1], Justin Elser[2], Noor Al-Bader[2], Sunita Kumari[1], James Thomason[1], Vivek Kumar[1], Daniel M. Bolser[4], Guy Naamati[4], Electra Tapanari[4], Nuno Fonseca[4], Laura Huerta[4], Haider Iqbal[4], Maria Keays[4], Alfonso Munoz-Pomer Fuentes[4], Amy Tang[4], Antonio Fabregat[4], Peter D'Eustachio[5], Joel Weiser[6], Lincoln D. Stein[7], Robert Petryszak[4], Irene Papatheodorou[4], Paul J. Kersey[4], Patti Lockhart[8], Crispin Taylor[8], Pankaj Jaiswal[2] and Doreen Ware[1,9,*]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA, [2]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA, [3]Division of Biological Sciences and Institute for Basic Science, Wonkwang University, Iksan 54538, Korea, [4]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SD, UK, [5]Department of Biochemistry & Molecular Pharmacology, NYU School of Medicine, New York, NY 10016, USA, [6]Informatics and Bio-computing Program, Ontario Institute of Cancer Research, Toronto, M5G 1L7, Canada, [7]Adaptive Oncology Program, Ontario Institute for Cancer Research, Toronto M5G 0A3, Canada, [8]American Society of Plant Biologists, 15501 Monona Drive, Rockville, MD 20855-2768, USA and [9]USDA ARS NAA Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, Ithaca, NY 14853, USA

Received September 15, 2017; Revised October 20, 2017; Editorial Decision October 21, 2017; Accepted October 25, 2017

ABSTRACT

Gramene (http://www.gramene.org) is a knowledgebase for comparative functional analysis in major crops and model plant species. The current release, #54, includes over 1.7 million genes from 44 reference genomes, most of which were organized into 62,367 gene families through orthologous and paralogous gene classification, whole-genome alignments, and synteny. Additional gene annotations include ontology-based protein structure and function; genetic, epigenetic, and phenotypic diversity; and pathway associations. Gramene's Plant Reactome provides a knowledgebase of cellular-level plant pathway networks. Specifically, it uses curated rice reference pathways to derive pathway projections for an additional 66 species based on gene orthology, and facilitates display of gene expression, gene–gene interactions, and user-defined omics data in the context of these pathways. As a community portal, Gramene integrates best-of-class software and infrastructure components including the Ensembl genome browser, Reactome pathway browser, and Expression Atlas widgets, and undergoes periodic data and software upgrades. Via powerful, intuitive search interfaces, users can easily query across various portals and interactively analyze search results by clicking on diverse features such as genomic context, highly augmented gene trees, gene expression anatomograms, associated pathways, and external informatics resources. All data in Gramene are accessible through both visual and programmatic interfaces.

INTRODUCTION

The Gramene database is a versatile resource for querying, visualizing, analyzing, and comparing plant genome and pathway data across crops and model species. Our genomic data portal was produced in collaboration with Ensembl Genomes (1,2), and shares infrastructure, specialized software components and pre-computed data with Ensembl Plants. For example, the portal uses Ensembl's data model and analysis workflows to generate baseline genome annotations and perform genome-wide comparative anal-

*To whom correspondence should be addressed. Tel: +1 516 367 6979; Fax: +1 516 367 6851; Email: ware@cshl.edu; Doreen.ware@ars.usda.gov

Table 1. New data and functionalities in the Gramene database, by release (September 2015 - August 2017)

Gramene build #	48/b	49	50	51	52	53	54
Ensembl Plants build #	29	30	31	32	33	35	36
Release date	11/15	01/16	04/16	08/16	11/16	05/17	07/17
Genome assemblies		■		■			■
Gene annotations	■	■	■	■	■	■	■
Genetic variation		■	■	■		■	■
Gene trees		■	■	■	■	■	■
Sequence alignments	■		■	■	■		
Gene expression	■	■	■	■	■	■	■
Plant Reactome pathways	■		■		■		
Ensembl software update	■	■	■	■	■	■	■
Reactome software udpate	■					■	■
Analysis tools	■	■	■	■	■	■	■
BioMart builds	■	■	■	■	■	■	■
Archive resources	■	■	■	■	■	■	■
Website infrastructure	■	■	■		■	■	

Notes: The blue color indicates whether a given type of data or software feature was affected in a given release. This information is also available from our release notes (http://gramene.org/release-notes) and Supplementary Table S1.

ysis in order to construct phylogenetic trees, synteny, and whole-genome alignments for 44 plant reference genomes; the Ensembl genome browser to visualize and explore genomic data; and Ensembl online tools such as BLAST, BioMart, assembly converter, and the Variant Effect Predictor (VEP) for data analysis. The Plant Reactome database (http://plantreactome.gramene.org) (3,4), Gramene's pathways portal, is an ongoing development effort in collaboration with the human Reactome project (5). The Plant Reactome hosts pathway data for 67 species representing several model and crop plants, algae, and unicellular photoautotrophs. Rice is used as the reference species for manual curation of pathways and to derive gene orthology–based projections for other species, as described previously (3,4,6). Curated rice pathway data are integrated in the central pathway curation database and maintained by the human Reactome project. In addition, Gramene's gene pages and the Plant Reactome pathway browser make use of the whole-plant and seed anatomograms produced by the EMBL-EBI's Expression Atlas to display baseline gene expression.

All data in the Gramene database can be downloaded in graphical and/or tabular form in various standard formats and are also accessible programmatically via Application Programming Interfaces (APIs). Gramene also continues to provide access to archived legacy data (http://www.gramene.org/archive) (6) including Pathway Tools–based metabolic networks (7–9) from CyVerse (10).

In this article, we describe the new data and functionalities in the Gramene database introduced since our last NAR report (6). A summary is provided in Table 1.

NEW GRAMENE HOMEPAGE AND INTEGRATED SEARCH INTERFACE

A brand new website homepage with an integrated search interface, built on the React Javascript framework to support development of client-side visualizations, was released in May 2016 (build 53). The new homepage has a type-ahead search box to facilitate quick search and links to our various portal and resources. The genomic distribution of genes in the search results are summarized and individual genes are displayed in an expandable list with embedded views of gene structure, gene expression, conservation, associated pathways, and database cross-references, described later.

When a query only matches genes in well-annotated model organisms, the homology tab can be used to modify the search to show evolutionarily related genes. The main feature of the homology tab is an integrated visualization component that displays the expanded gene tree showing the gene of interest and its closest well-annotated homolog. Additional information is displayed as a track next to each leaf node or collapsed subtree. The default display mode shows regions of proteins included in the multiple sequence alignment, with color-coded InterPro domains. Users can pan and zoom the view or switch to a scrollable multiple-sequence alignment mode.

Because navigating search results or a gene tree with hundreds of family members from dozens of species can be overwhelming, and is unnecessary for most users, we implemented a filter to limit search results by a user-defined subset of species. Gene tree pruning removes excess gaps from the multiple sequence alignment, thereby eliminating visual clutter from less closely related species. These developments add value beyond the standard views by integrating diverse data in fast interfaces, while giving users the ability to focus on the species most relevant to their research. All code developed for the web service and the search interface are maintained on GitHub (https://github.com/warelab).

NEW AND UPDATED PLANT GENOMES

The current release of Gramene added five new fully sequenced reference plant genomes and updated four existing genomes (Figure 1A, Table 1, and Supplementary Table S1), bringing the total to 44. The new species include three dicots, *Beta vulgaris* subsp. *Vulgaris* (sugar beet), *Brassica napus* (oilseed rape) and *Trifolium pratense* (red clover), and two red algae, *Chondrus crispus* (Irish or carrageen moss) and *Galdieria sulphuraria,* an extremophilic unicellular species. All reference genome assemblies in Gramene are accessioned in the International Nucleotide Sequence Database Collaboration, INSDC (11), a policy that ensures provenance and interoperability with other online resources. The complete list of genomes in Gramene, shown in Figure 1A, includes 15 eudicots, 21 monocots, 1 basal angiosperm, and 7 non-flowering species. The four updated assemblies correspond to three staple food crops, *Zea mays* cv. B73 (maize), *Triticum aestivum* cv. Chinese Spring (wheat), and *Sorghum bicolor* cv. BTx623 (sorghum), and a wild rice indigenous to sub-Saharan Africa, *Oryza longistaminata*. Gramene researchers led the innovative effort to generate the new maize reference assembly (B73 RefGen_v4), constructed entirely from third-generation long-read sequence technology with the aid of an optical map (12).

The assembly was released with a new set of gene annotations generated by the MAKER-P pipeline (13) from 111 000 long-read Iso-Seq transcripts obtained by single-molecule sequencing (14). This approach more than doubled the number of alternative transcripts from 1.5 to 3.8 per gene, resolved gaps and assembly errors, corrected strand, consolidated gene models, and anchored previously unanchored genes (12). The new bread wheat as-

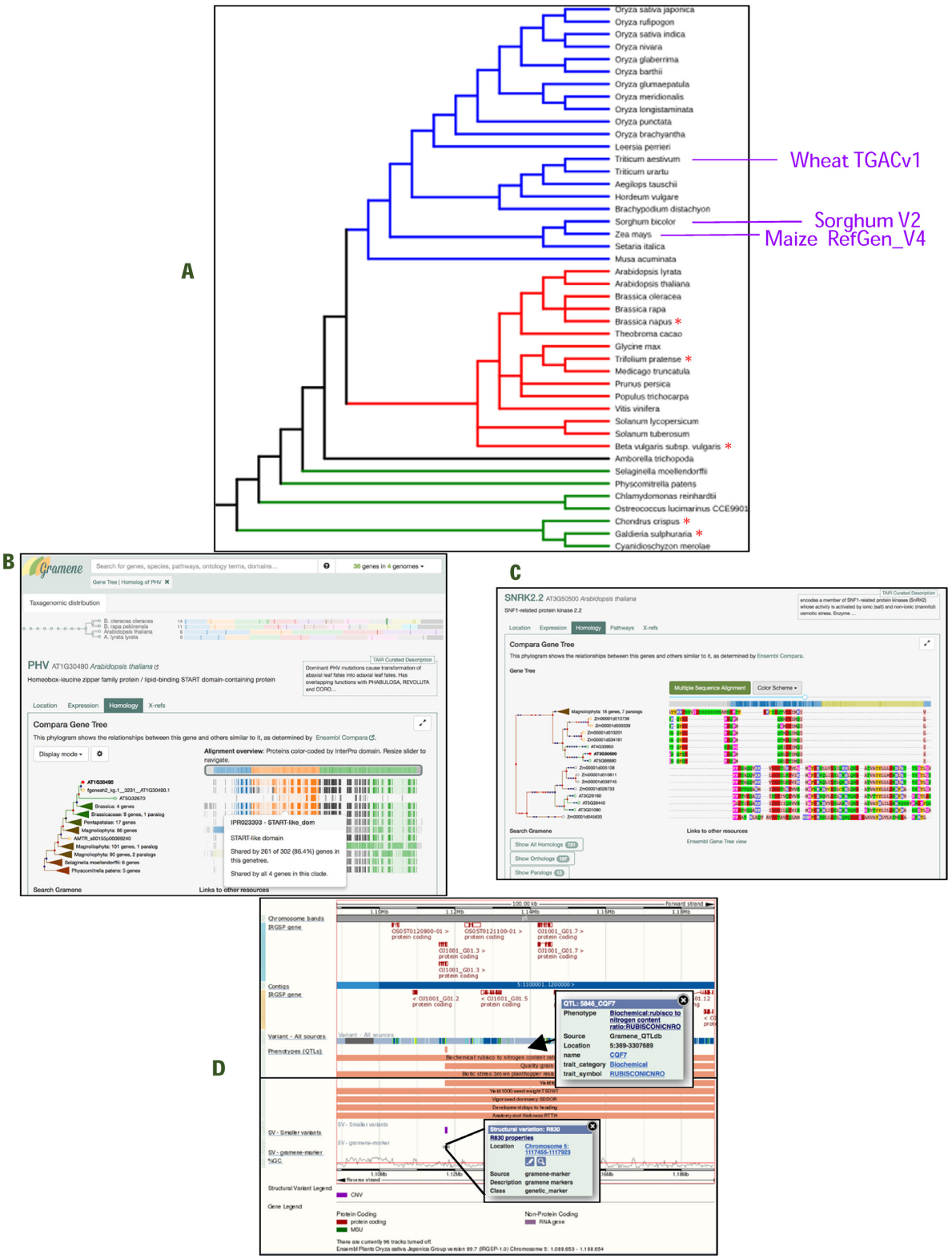

Figure 1. Visualization of query results in the new Gramene database search interface. (**A**) Species tree of 44 plant reference genomes available in Gramene build 54. Blue, red, black and green lines correspond respectively to 21 monocots, 15 dicots, the basal angiosperm *Amborella* and 7 lower plant species. New genome assemblies are marked with a red asterisk (*), and updated assemblies are highlighted in violet font (see also Table 1 and Supplementary Table S1). (**B**) Customized gene tree alignment views, color coded by InterPro domain. Filtered query results and gene tree branches are shown for selected species. Filter query results and gene tree branches are shown for selected species. (**C**) Detailed DNA sequence alignment. (**D**) *Oryza sativa* QTLs, SSRs, and RFLPs from legacy sets reincorporated to the IRGSP1 assembly (mapping courtesy of KeyGene). QTLs from Q-TARO (Yonemaru *et al*, 2010) are also available.

sembly, TGACv1, produced by the Earlham Institute (formerly TGAC), includes ∼99% of the genes remapped from the prior IWGSCv1 Chromosome Survey Sequence assembly (15). To facilitate transformation of genomic coordinates between assemblies, we continue to maintain an easy-to-use online assembly converter tool. For example, using this utility, it is now possible to traverse across the three maize B73 reference assemblies, RefGen_v2, RefGen_v3, and RefGen_v4 (http://ensembl.gramene.org/Zea_mays/Tools/AssemblyConverter?db=core).

UPDATED ANNOTATION AND COMPARATIVE GENOMICS

Genes and their encoded proteins are functionally annotated in Gramene using InterProScan (16), allowing comparison of InterPro functional domains and Gene Ontology terms within a gene-centered phylogenomic framework. Release 54 includes revised gene models for the updated maize, wheat and sorghum genome sequence assemblies, as well as new gene models for *Arabidopsis thaliana* (Araport 11; https://www.araport.org/data/araport11), although the genome assembly for the later was not updated.

Our phylogenomic analysis utilizes the Compara method (17) to reconstruct the evolutionary history of genes by clustering homologous genes into families, building gene trees with stable IDs for future reference, and classifying pairwise orthologous and paralogous relationships. Each node of a consensus tree is assigned a taxonomic date to identify points at which gene duplications occurred within evolving lineages, as well as specify the common ancestor in which the gene family emerged. For closely related species, synteny maps are constructed to reveal regions of conserved gene order (18,19). These maps are complemented by whole-genome alignments showing conserved intergenic, as well as genic, regions (19). These data can be visualized using the web-based Ensembl browser (20,21) and mined using BioMart (22,23), both of which are accessible programmatically. These data have been used for a wide range of applications, including the study of protein–protein interaction network evolution in *Arabidopsis* (24), development of methods for enrichment of candidate genes in genome-wide association studies in maize (25), and construction of genome-scale knowledge networks in wheat and barley (26). In addition, Compara analysis allows us to identify gene annotation errors (split gene models), suggest functional annotations for less well-annotated orthologs, and project mappings from curated rice pathways to other species.

NEW PLANT GENETIC DIVERSITY AND MUTAGENIZED SEQUENCE VARIATION

Gramene continues to provide single-nucleotide polymorphism (SNP) and/or structural variation data for 12 genomes (6,27–42; http://1001genomes.org). We recently added new variants for *A. thaliana*, Japonica rice, *O. glumaepatula*, sorghum and wheat.

Since our last NAR update, we added a combined total of over 8.8 million ethyl methanesulfonate (EMS)-derived mutations for sorghum and wheat. The sorghum dataset (41) includes ∼1.5 million EMS-induced G/C to A/T transition mutations, annotated from 252 M3 families selected from a 6400-mutant library in the BTx623 background. The wheat dataset comprises ∼7.4 million EMS-type variants derived from sequencing of tetraploid (2.8 million cv. 'Kronos') and hexaploid (4.6 million cv 'Cadenza') TILLING populations. Mutations were originally called in the IWGSC CSS scaffolds, as described in Krasileva *et al.* (42), and the high-confidence (*HetMC5HomMC3*) mutations were then projected onto the TGAC Chinese Spring scaffolds. The wheat EMS mutant data are also available in a specialized database generated by a joint project between the University of California Davis, Rothamsted Research, The Earlham Institute, and the John Innes Centre. Researchers and breeders can search this online database, identify mutations in different copies of their target genes, and request seeds for use in studies of gene function or improvement of wheat varieties via the project online search tools at http://www.wheat-tilling.com and http://dubcovskylab.ucdavis.edu/wheat-tilling. Seed requests can also be made from the UK SeedStor resource (https://www.seedstor.ac.uk/shopping-cart-tilling.php). The number of *O. glumaepatula* variants from the *Oryza* Genome Evolution project (Stein *et al.*, Submitted) is currently 4.9 million. For assembly updates of maize, wheat, and sorghum, variants were remapped to the latest genomic coordinate system (e.g. HapMap SNPs for maize, wheat and inter-homoeologous wheat variants). We have also incorporated rice QTLs from Gramene's archives (6) (http://archive.gramene.org/qtl) and the Q-TARO database (43; http://qtaro.abr.affrc.go.jp), as well as legacy rice SSR/RFLP data from our archives (http://archive.gramene.org/markers) remapped to the IRGSPv1 assembly.

We continue to analyze and assign putative functional and structural consequences to gene variants using the Ensembl VEP tool (4,44). Visualization of these consequences is provided in the context of transcript structure and protein domains. For many studies, we also provide information on the genotypes of individual plant accessions and their phenotypes.

PLANT GENE EXPRESSION ATLAS

The plant gene Expression Atlas (https://www.ebi.ac.uk/gxa/plant/experiments) contains transcriptomic data from 731 experiments in 18 plant species. These data have been manually curated, quality-controlled, and analyzed using standardized analysis pipelines (45). Baseline expression data from RNA-seq experiments from 14 plant species show expression levels of gene products under 'normal' conditions in various tissues (leaves, roots, etc.), developmental stages, cultivars, and ecotypes. The baseline expression profile of an individual gene across all tissue samples and growth stages, from EMBL-EBI Expression Atlas, can be accessed from the gene page in the Gramene database, as well as from the Plant Reactome Pathway Browser. Differential gene expression data, including responses to environmental stresses, genetic mutations, and bacterial infection, are available for over 2000 manually curated pairwise comparisons from 15 plant species, and include data from both microarray and RNA-seq experiments. At present, differential expression data can be viewed on the Expres-

sion Atlas website, and displayed on demand (not automatically) on the Gramene gene page and the Genome Browser as a genome feature track. Additional features include, visualization of GO, Pathway and Interpro domain enrichments in the given expression data. Baseline experiment page allows users to find genes with similar expression profiles. Moreover, to achieve adequate annotation of all experiments in Expression Atlas, the EMBL-EBI Experimental Factor Ontology (46) (EFO) has been supplemented with Plant Ontology (47,48) (PO) and Brenda Tissue and Enzyme Source Ontology (EFO). An automatic, scalable framework is used to propagate manually curated ontology annotations to matching sample attributes, ensuring that all new plant experiments loaded into Expression Atlas in the future will benefit from the existing ontology annotations without further manual curation. On a daily basis, an automatic analysis pipeline discovers new RNA-seq runs in 38 plant species in the European Nucleotide Archive, and then performs quality control, aligns them to the genome reference in Ensembl Plants, and quantifies gene and exon expression. The pipeline also re-aligns all runs in a given species when a new genome assembly is released. To date, 33 000 runs have been processed, and the results are available via the RNASeq-er API (http://www.ebi.ac.uk/fg/rnaseq/api), as well as a BioPython module and a CPAN Perl library. All experimental data in Expression Atlas are available for download as R objects from the corresponding experiment pages. Experiments can also be found via an ontology-powered search and retrieved from within the ExpressionAtlas Bioconductor package (https://www.bioconductor.org/packages/release/bioc/html/ExpressionAtlas.html). Finally, high-quality anatomical illustrations in Scalable Vector Graphics (SVG) format for the plant species in Expression Atlas have been created by a professional medical illustrator and deployed to highlight tissues with expression within the baseline expression views.

ENHANCED PLANT REACTOME PORTAL

The Plant Reactome database (3,4) (http://plantreactome.gramene.org) is a unifying resource for describing pathways associated with plant metabolism, development, and differentiation, and their regulation in response to various environmental stimuli. The Plant Reactome uses Japonica rice (*O. sativa*) as a reference for manual curation of pathways within the framework of plant cell architecture. In the current release, the database contains 241 reference pathways, including 229 curated and 12 predicted rice pathways; the latter were derived from projections of curated human pathways associated with evolutionarily conserved processes including DNA replication, vesicle transport, and translation. Since our last report, we added ∼25 new manually curated rice pathways, including choline biosynthesis, cardiolipin biosynthesis, glycine betaine biosynthesis, suberin biosynthesis, tryptophan biosynthesis, tyrosine biosynthesis, lysine degradation II, inter- and intra-cellular auxin transport, response to cold temperature, circadian rhythm, inflorescence development, and transition from vegetative to reproductive shoot apical meristem.

The reference set of rice pathways is used to generate gene orthology–based pathway projections for additional species (4). Since our last NAR report, the number of species available in the Plant Reactome doubled from 33 to 66, corresponding to all 44 species with sequenced reference genomes in Gramene and 22 additional species with publicly available sequenced genomes and/or transcriptomes (http://plantreactome.gramene.org/stats.html). The presence or absence of orthologs drives the projection of homologous reactions and pathway maps for a given species. Users can access pathway data in graphical and tabular format for any member species and compare the projected pathways with reference rice pathways.

As shown in Figure 2, our new 'Fireworks' pathway visualization platform displays hierarchical graphs of pathways, allows exploration of pathways by functional categories, links the Pathway Browser to various public external resources and genome databases that provide complementary information about various pathway entities, displays expression profiles of genes (fetched remotely from EMBL-EBI's gene Expression Atlas), and provides the option to display gene–gene interaction data from EBI's INTACT database via PSICQUIC web services. We also support upload, visualization, and analysis of user-defined transcriptome, proteome, metabolome and gene–gene interaction data, which are accessible for download in various standard formats.

INTEGRATED SEARCH

For each release of Gramene, we run pipelines to extract data and annotation terms from Ensembl, Reactome, Expression Atlas, and other external reference resources, transform them into JSON documents, and load them into MongoDB collections. The documents in the genes collection, initially generated from Ensembl core MySQL databases, were extended to include homology information from Ensembl Compara, relevant reactions and pathways from Plant Reactome, and associated ontology terms. Ontologies, the InterPro domain hierarchy, the pathway hierarchy, and the taxonomy tree have inherent structures in which parent terms are more general than their children. When a gene is associated with a specific term, we also include any ancestor terms from the corresponding ontology in our index, enabling users to find genes associated with any less specific, yet related, term. We also integrated InterPro domain annotations into gene tree documents to support their display in our gene tree browser. To support free text search and complex combinations of filters, we transformed the collection of gene documents for use in Apache Solr. To power the type-ahead search feature, we also prepared an index of suggestions derived from the various MongoDB collections. Each suggestion document includes fields that define a filter on the genes index. This design guides users to choose terms drawn from controlled vocabularies and encourages them to iteratively refine the search by adding or modifying filters. We use swagger (http://swagger.io) to define, document and serve our APIs, which access data in the MongoDB collections and Solr cores via http://data.gramene.org. We deploy customized installations of Ensembl and Reactome REST APIs to support

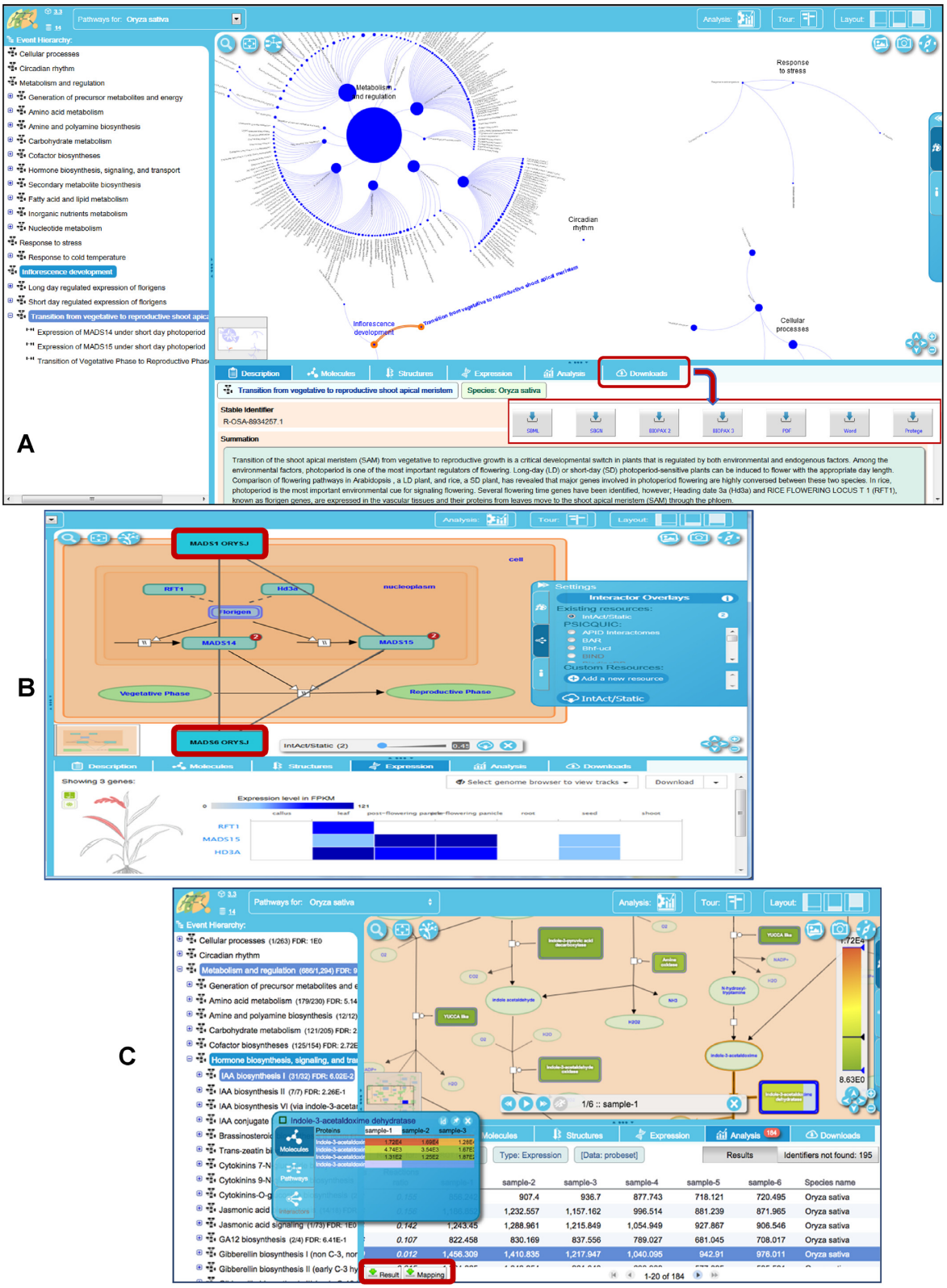

Figure 2. Plant Reactome pathway views and functionalities. (**A**) The 'Fireworks' pathway visualization platform displays a navigation panel (left) and a hierarchical graph of pathways in a pathway viewer window with various associated features, such as summation, download options, and links to external public resources providing complementary information on various pathway entities. (**B**) A view of the Plant Reactome pathway displaying an overlay of gene–gene interaction data. Two common interactors of the MADS15 and MADS14 transcription factors (MAD1 ORYS and MAD6 ORYS, shown in red boxes) were imported via web services from the IntAct database. (**C**) Analysis and visualization of user-uploaded gene expression data on the pathway browser, with options to explore full expression profiles of the homologs and download results.

the build pipeline and specific visualizations in the search interface.

COMMUNITY OUTREACH AND TRAINING

Gramene organizes community outreach activities aimed at training plant biologists at various stages of their careers, including high school students and faculty, undergraduate and graduate students, postdoctoral fellows, technical staff, database managers, senior researchers, and group leaders. We regularly publicize our database updates, meeting reports, new tools, and online or on-site training activities via the Gramene News blog, Facebook and Twitter (see http://gramene.org/outreach).

We organize regular online webinars and provide recorded video tutorials on the Gramene Youtube channel (https://goo.gl/ln9RLD) to train researchers to use our resources and bioinformatics tools, as well as to access data from public repositories to analyze using Gramene's comparative genomic and pathway tools. We conducted annual on-site workshops during the Plant and Animal Genome Conference and the Plant Biology meeting to introduce recent updates and new Gramene features to the plant community and provide personalized assistance to our users. We also collaborate with various plant researchers and public databases (e.g. AgBioData Consortium, GrapeIS (49), WheatIS) on development and improvement of data standards.

DISCUSSION AND FUTURE DIRECTIONS

In this report, we described two major improvements made to our website since our last NAR update: Gramene's new homepage with integrated search interface, and improved views and functionalities for Plant Reactome. Our current homepage supports simultaneous search of the contents of all Gramene portals, and produces summarized output with clickable links allowing detailed exploration. The Plant Reactome now utilizes the 'Fireworks' pathway visualization platform for pathway organization and navigation and provides users an option to overlay gene–gene interaction data (fetched remotely via APIs from external resources or by uploading their own data) on the pathway diagrams.

In the past 2 years, the number of species covered, with corresponding annotations, has increased across all portals of Gramene (Genome Browsers, Plant Reactome, Expression Atlas, outreach and training material, etc.). Gramene will continue to develop value-added genomic and pathway resources, and extend and develop the platform across all data types. In the future, we expect to surpass 75 annotated genomes and corresponding pathway projections in Plant Reactome, and to include expression datasets for these new species. In addition, we will produce sub-sites focused on important crop species with multiple sequenced genomes (i.e. pangenomes), such as maize, sorghum, rice, wheat, and grape, to provide pan-species views and facilitate comparative analysis. We will improve and develop analysis and visualization tools with special emphasis on inter- and intra-specific analysis of small- and large-scale data sets generated from global crops and emerging plant model species. We will continue to improve our integrated search interface with enhancements to the user interface, new visualizations for interactive search result summaries, and integrated gene tree display modes for exploration of functional conservation.

In addition, we will continue to train current and future plant biologists on how to generate accurate and reliable data sets, as well as how to use, re-use, and analyze genomic data. Gramene is working with plant researchers and other genomics communities to develop standard formats and guidelines for making heterogeneous data sets, including genes and genomes, germplasm, and high-throughput gene expression data, phenotypes, metabolomes, and proteomes, and annotations, accessible, reusable, and interoperable. We support the efforts of plant biologists to learn how to accurately annotate and format their data sets using community standards and ontology concepts. One of the key challenge is on scaling curation activities. To address this effort, Gramene has an ongoing collaboration with the American Society for Plant Biologists (ASPB) and its journals *Plant Physiology* and *The Plant Cell* to improve adoption of standards for data formatting and annotation during manuscript production. Validating and integrating data that are published in journals with data that reside in databases will improve the utility of both.

Through our monthly webinar series, and onsite workshops, we will continue to assist researchers in taking full advantage of the available genomic resources and seek their suggestions for improving our tools, data models, and resources. Our hope is that our users will apply these skills for construction of novel hypotheses, validation of existing knowledge, and characterization of all aspects of plant genes, including their functions, expression, roles in pathways, phenotypes, and associated functional and structural variations.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors are grateful to Gramene's users, researchers, and numerous collaborators for sharing datasets generated by their projects, and for providing valuable suggestions and feedback that have helped us to improve the overall quality of Gramene as a community resource. We also thank the Cold Spring Harbor Laboratory (CSHL) and the Dolan DNA Learning Center, the Center for Genome Research and Biocomputing (CGRB) at Oregon State University, and the Ontario Institute for Cancer Research (OICR) for infrastructure support. We thank Anker P. Sørensen from Keygene Inc. for remapping the rice QTLs from Q-TARO and Gramene's archive databases. We also thank Peter van Buren from Cold Spring Harbor Laboratory for system administration support, and David Croft (EBI) and Robin Haw (OICR) for technical support of Plant Reactome development. The funders had no role in study design, data analysis, or preparation of this manuscript.

FUNDING

National Science Foundation [IOS-1127112]; United States Department of Agriculture—Agricultural Research Service [58-1907-4-030 and 8062-21000-041-00D to D.W.]; United Kingdom Biotechnology and Biosciences Research Council [B/I008071/1, BB/J00328X/1, BB/P016855/1 to P.K.]. The in-kind infrastructure and intellectual support for the development and running the Plant Reactome is supported by the Reactome database project via a grant from the US National Institutes of Health [U41 HG003751 to L.S.]; EU grant [LSHG-CT-2005–518254]; 'ENFIN', the Ontario Research Fund; and EBI Industry Programme. Oregon State University (OSU) provided partial support for graduate students at OSU. Funding for open access charge: NSF award # 1127112 to the Gramene Project.
Conflict of interest statement. None declared.

REFERENCES

1. Kersey,P.J., Allen,J.E., Armean,I., Boddu,S., Bolt,B.J., Carvalho-Silva,D., Christensen,M., Davis,P., Falin,L.J., Grabmueller,C. *et al.* (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.*, **44**, D574–D580.
2. Kersey,P.J., Allen,J.E., Allot,A., Barba,M., Boddu,S., Bolt,B.J., Carvalho-Silva,D., Christensen,M., Davis,P., Grabmueller,C. *et al.* (2017) Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.*, doi:10.1093/nar/gkx1011.
3. Gupta,P., Naithani,S., Tello-Ruiz,M.K., Chougule,K., D'Eustachio,P., Fabregat,A., Jiao,Y., Keays,M., Lee,Y.K., Kumari,S. *et al.* (2016) Gramene Database: navigating plant comparative genomics resources. *Curr. Plant Biol.*, **7–8**, 10–15.
4. Naithani,S., Preece,J., D'Eustachio,P., Gupta,P., Amarasinghe,V., Dharmawardhana,P.D., Wu,G., Fabregat,A., Elser,J.L., Weiser,J. *et al.* (2017) Plant Reactome: a resource for plant pathways and comparative analysis. *Nucleic Acids Res.*, **45**, D1029–D1039.
5. 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.
6. Tello-Ruiz,M.K., Stein,J., Wei,S., Preece,J., Olson,A., Naithani,S., Amarasinghe,V., Dharmawardhana,P., Jiao,Y., Mulvaney,J. *et al.* (2016) Gramene 2016: comparative plant genomics and pathway resources. *Nucleic Acids Res.*, **44**, D1133–D1140.
7. Monaco,M.K., Sen,T.Z., Dharmawardhana,P.D., Ren,L., Schaeffer,M., Naithani,S., Amarasinghe,V., Thomason,J., Harper,L. and Gardiner,J. (2013) Maize metabolic network construction and transcriptome analysis. *Plant Genome*, **6**, doi:10.3835/plantgenome2012.09.0025.
8. Monaco,M.K., Stein,J., Naithani,S., Wei,S., Dharmawardhana,P., Kumari,S., Amarasinghe,V., Youens-Clark,K., Thomason,J., Preece,J. *et al.* (2014) Gramene 2013: comparative plant genomics resources. *Nucleic Acids Res.*, **42**, D1193–D1199.
9. 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.
10. Merchant,N., Lyons,E., Goff,S., Vaughn,M., Ware,D., Micklos,D. and Antin,P. (2016) The iPlant Collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biol.*, **14**, e1002342.
11. Nakamura,Y., Cochrane,G., Karsch-Mizrachi,I. and International Nucleotide Sequence Database, C. (2013) The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.*, **41**, D21–D24.
12. Jiao,Y., Peluso,P., Shi,J., Liang,T., Stitzer,M.C., Wang,B., Campbell,M.S., Stein,J.C., Wei,X., Chin,C.S. *et al.* (2017) Improved maize reference genome with single-molecule technologies. *Nature*, **546**, 524–527.
13. Campbell,M.S., Law,M., Holt,C., Stein,J.C., Moghe,G.D., Hufnagel,D.E., Lei,J., Achawanantakun,R., Jiao,D., Lawrence,C.J. *et al.* (2014) MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.*, **164**, 513–524.
14. Wang,B., Tseng,E., Regulski,M., Clark,T.A., Hon,T., Jiao,Y., Lu,Z., Olson,A., Stein,J.C. and Ware,D. (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.*, **7**, 11708.
15. International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, **345**, 1251788.
16. Jones,P., Binns,D., Chang,H.Y., Fraser,M., Li,W., McAnulla,C., McWilliam,H., Maslen,J., Mitchell,A., Nuka,G. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**, 1236–1240.
17. 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.
18. Schnable,P.S., Ware,D., Fulton,R.S., Stein,J.C., Wei,F., Pasternak,S., Liang,C., Zhang,J., Fulton,L., Graves,T.A. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.
19. Youens-Clark,K., Buckler,E., Casstevens,T., Chen,C., Declerck,G., Derwent,P., Dharmawardhana,P., Jaiswal,P., Kersey,P., Karthikeyan,A.S. *et al.* (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Res.*, **39**, D1085–D1094.
20. Flicek,P., Ahmed,I., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S. *et al.* (2013) Ensembl 2013. *Nucleic Acids Res.*, **41**, D48–D55.
21. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2014) Ensembl 2014. *Nucleic Acids Res.*, **42**, D749–D755.
22. Spooner,W., Youens-Clark,K., Staines,D. and Ware,D. (2012) GrameneMart: the BioMart data portal for the Gramene project. *Database (Oxford)*, **2012**, bar056.
23. Kinsella,R.J., Kahari,A., Haider,S., Zamora,J., Proctor,G., Spudich,G., Almeida-King,J., Staines,D., Derwent,P., Kerhornou,A. *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)*, **2011**, bar030.
24. Arabidopsis Interactome Mapping, C. (2011) Evidence for network evolution in an Arabidopsis interactome map. *Science*, **333**, 601–607.
25. Chen,C., DeClerck,G., Tian,F., Spooner,W., McCouch,S. and Buckler,E. (2012) PICARA, an analytical pipeline providing probabilistic inference about a priori candidates genes underlying genome-wide association QTL in plants. *PLoS One*, **7**, e46596.
26. Hassani-Pak,K., Castellote,M., Esch,M., Hindle,M., Lysenko,A., Taubert,J. and Rawlings,C. (2016) Developing integrated crop knowledge networks to advance candidate gene discovery. *Appl. Transl. Genom.*, **11**, 18–26.
27. Atwell,S., Huang,Y.S., Vilhjalmsson,B.J., Willems,G., Horton,M., Li,Y., Meng,D., Platt,A., Tarone,A.M., Hu,T.T. *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, **465**, 627–631.
28. Clark,R.M., Schweikert,G., Toomajian,C., Ossowski,S., Zeller,G., Shinn,P., Warthmann,N., Hu,T.T., Fu,G., Hinds,D.A. *et al.* (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science*, **317**, 338–342.
29. Fox,S.E., Preece,J., Kimbrel,J.A., Marchini,G.L., Sage,A., Youens-Clark,K., Cruzan,M.B. and Jaiswal,P. (2013) Sequencing and de novo transcriptome assembly of Brachypodium sylvaticum (Poaceae). *Appl. Plant Sci.*, **1**, doi:10.3732/apps.1200011.
30. Gan,X., Stegle,O., Behr,J., Steffen,J.G., Drewe,P., Hildebrand,K.L., Lyngsoe,R., Schultheiss,S.J., Osborne,E.J., Sreedharan,V.T. *et al.* (2011) Multiple reference genomes and transcriptomes for Arabidopsis thaliana. *Nature*, **477**, 419–423.
31. Fox,S.E., Geniza,M., Hanumappa,M., Naithani,S., Sullivan,C., Preece,J., Tiwari,V.K., Elser,J., Leonard,J.M., Sage,A. *et al.* (2014) De novo transcriptome assembly and analyses of gene expression during photomorphogenesis in diploid wheat Triticum monococcum. *PLoS One*, **9**, e96855.
32. International Barley Genome Sequencing, C., Mayer,K.F., Waugh,R., Brown,J.W., Schulman,A., Langridge,P., Platzer,M.,

Fincher,G.B., Muehlbauer,G.J., Sato,K. *et al.* (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.
33. Mace,E.S., Tai,S., Gilding,E.K., Li,Y., Prentis,P.J., Bian,L., Campbell,B.C., Hu,W., Innes,D.J., Han,X. *et al.* (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.*, **4**, 2320.
34. McNally,K.L., Childs,K.L., Bohnert,R., Davidson,R.M., Zhao,K., Ulat,V.J., Zeller,G., Clark,R.M., Hoen,D.R., Bureau,T.E. *et al.* (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 12273–12278.
35. Morris,G.P., Ramu,P., Deshpande,S.P., Hash,C.T., Shah,T., Upadhyaya,H.D., Riera-Lizarazu,O., Brown,P.J., Acharya,C.B., Mitchell,S.E. *et al.* (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 453–458.
36. Myles,S., Chia,J.M., Hurwitz,B., Simon,C., Zhong,G.Y., Buckler,E. and Ware,D. (2010) Rapid genomic characterization of the genus vitis. *PLoS One*, **5**, e8219.
37. Zhao,K., Wright,M., Kimball,J., Eizenga,G., McClung,A., Kovach,M., Tyagi,W., Ali,M.L., Tung,C.W., Reynolds,A. *et al.* (2010) Genomic diversity and introgression in O. sativa reveal the impact of domestication and breeding on the rice genome. *PLoS One*, **5**, e10780.
38. Zheng,L.Y., Guo,X.S., He,B., Sun,L.J., Peng,Y., Dong,S.S., Liu,T.F., Jiang,S., Ramachandran,S., Liu,C.M. *et al.* (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor). *Genome Biol.*, **12**, R114.
39. Tomato Genome Sequencing, C., Aflitos,S., Schijlen,E., de Jong,H., de Ridder,D., Smit,S., Finkers,R., Wang,J., Zhang,G., Li,N. *et al.* (2014) Exploring genetic variation in the tomato (Solanum section Lycopersicon) clade by whole-genome sequencing. *Plant J.*, **80**, 136–148.
40. Chia,J.M., Song,C., Bradbury,P.J., Costich,D., de Leon,N., Doebley,J., Elshire,R.J., Gaut,B., Geller,L., Glaubitz,J.C. *et al.* (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.*, **44**, 803–807.
41. Jiao,Y., Burke,J., Chopra,R., Burow,G., Chen,J., Wang,B., Hayes,C., Emendack,Y., Ware,D. and Xin,Z. (2016) A sorghum mutant resource as an efficient platform for gene discovery in grasses. *Plant Cell*, **28**, 1551–1562.
42. Krasileva,K.V., Vasquez-Gross,H.A., Howell,T., Bailey,P., Paraiso,F., Clissold,L., Simmonds,J., Ramirez-Gonzalez,R.H., Wang,X., Borrill,P. *et al.* (2017) Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E913–E921.
43. Regulski,M., Lu,Z., Kendall,J., Donoghue,M.T., Reinders,J., Llaca,V., Deschamps,S., Smith,A., Levy,D., McCombie,W.R. *et al.* (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res.*, **23**, 1651–1662.
44. Chen,Y., Cunningham,F., Rios,D., McLaren,W.M., Smith,J., Pritchard,B., Spudich,G.M., Brent,S., Kulesha,E., Marin-Garcia,P. *et al.* (2010) Ensembl variation resources. *BMC Genomics*, **11**, 293.
45. Petryszak,R., Keays,M., Tang,Y.A., Fonseca,N.A., Barrera,E., Burdett,T., Fullgrabe,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.
46. Malone,J., Holloway,E., Adamusiak,T., Kapushesky,M., Zheng,J., Kolesnikov,N., Zhukova,A., Brazma,A. and Parkinson,H. (2010) Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics*, **26**, 1112–1118.
47. Cooper,L. and Jaiswal,P. (2016) The Plant Ontology: A Tool for Plant Genomics. *Methods Mol. Biol.*, **1374**, 89–114.
48. 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.
49. Adam-Blondon,A.F., Alaux,M., Pommier,C., Cantu,D., Cheng,Z.M., Cramer,G.R., Davies,C., Delrot,S., Deluc,L., Di Gaspero,G. *et al.* (2016) Towards an open grapevine information system. *Hortic. Res.*, **3**, 16056.