Published online 8 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D1133–D1140***
doi: 10.1093/nar/gkv1179

Gramene 2016: comparative plant genomics and pathway resources

Marcela K. Tello-Ruiz[1], Joshua Stein[1], Sharon Wei[1], Justin Preece[2], Andrew Olson[1], Sushma Naithani[2], Vindhya Amarasinghe[2], Palitha Dharmawardhana[2], Yinping Jiao[1], Joseph Mulvaney[1], Sunita Kumari[1], Kapeel Chougule[1], Justin Elser[2], Bo Wang[1], James Thomason[1], Daniel M. Bolser[3], Arnaud Kerhornou[3], Brandon Walts[3], Nuno A. Fonseca[3], Laura Huerta[3], Maria Keays[3], Y. Amy Tang[3], Helen Parkinson[3], Antonio Fabregat[3], Sheldon McKay[4], Joel Weiser[4], Peter D'Eustachio[5], Lincoln Stein[4], Robert Petryszak[3], Paul J. Kersey[3], Pankaj Jaiswal[2] and Doreen Ware[1,6,*]

[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]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK, [4]Informatics and Bio-computing Program, Ontario Institute of Cancer Research, Toronto, M5G 1L7, Canada, [5]Department of Biochemistry & Molecular Pharmacology, NYU School of Medicine, New York, NY 10016, USA and [6]USDA ARS NAA Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, Ithaca, NY 14853, USA

Received September 25, 2015; Accepted October 13, 2015

ABSTRACT

Gramene (http://www.gramene.org) is an online resource for comparative functional genomics in crops and model plant species. Its two main frameworks are genomes (collaboration with Ensembl Plants) and pathways (The Plant Reactome and archival BioCyc databases). Since our last NAR update, the database website adopted a new Drupal management platform. The genomes section features 39 fully assembled reference genomes that are integrated using ontology-based annotation and comparative analyses, and accessed through both visual and programmatic interfaces. Additional community data, such as genetic variation, expression and methylation, are also mapped for a subset of genomes. The Plant Reactome pathway portal (http://plantreactome.gramene.org) provides a reference resource for analyzing plant metabolic and regulatory pathways. In addition to ∼200 curated rice reference pathways, the portal hosts gene homology-based pathway projections for 33 plant species. Both the genome and pathway browsers interface with the EMBL-EBI's Expression Atlas to enable the projection of baseline and differential expression data from curated expression studies in plants. Gramene's archive website (http://archive.gramene.org) continues to provide previously reported resources on comparative maps, markers and QTL. To further aid our users, we have also introduced a live monthly educational webinar series and a Gramene YouTube channel carrying video tutorials.

INTRODUCTION

Modern agriculture faces global challenges including food security and adaptation to climate change. In this context, bioinformatics resources can facilitate better understanding of the complexity, structure and evolution of plants. Gramene provides online resources for visualizing and comparing plant genomes and biological pathways. Community-based gene annotations constitute the primary sources of annotations accompanying each reference genome, and are supplemented with functional classification and phylogenomics-based comparisons. In addition, we annotate and display variation data, largely obtained through collaboration with large-scale re-sequencing and genotyping initiatives. The project is also committed to consolidating plant pathway databases by applying both manual curation and automated methods.

Gramene is powered by several platform infrastructures that are linked via the Drupal content management system to provide a unified user experience. Our genome browser (http://ensembl.gramene.org/genome_browser) takes advantage of the Ensembl infrastructure (1) to provide an interface for exploration of genome features,

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

functional ontologies, variation data and comparative phylogenomics. For the past six years, Gramene has partnered with the Plants division of Ensembl Genomes (http:/plants.ensembl.org) to jointly produce a common genome browser, mirrored yet independently hosted by the United States and the United Kingdom, respectively. This USA-European collaboration has facilitated the timely adoption of the software updates and novelty tools that frequently accompany Ensembl releases (1).

Gramene developed a comprehensive framework for metabolic, signaling and genetic networks at the systems level for plants using the Reactome data model, analysis and visualization platform (2), also known as The Plant Reactome. The current version of The Plant Reactome database website (http://plantreactome.gramene.org/) has a new user interface and features 238 rice pathways (~80% of which were manually curated), which also serve as reference for projection of orthologous pathways for 33 plant species with sequenced and annotated plant genomes.

Gramene's software platforms provide region-specific (e.g. genome browser) or pathway-specific (e.g. Plant Reactome) data downloads. In addition, project data are available for customizable downloads from the GrameneMart (3,4), bulk downloads via File Transfer Protocol (ftp://ftp.gramene.org/pub/gramene), and programmatic access via public MySQL (5) and a new RESTful API (http://data.gramene.org/) together with REST APIs from Plant Reactome and Ensembl. This article summarizes Gramene updates since the last report in this journal (3) through the 47th release of the Gramene database in August 2015. The website, database and its contents continue to be updated five times per year; changes can be followed at the Gramene news portal (http://www.gramene.org/blog) and by browsing the site's release notes (http://www.gramene.org/release-notes).

NEW PLANT GENOMES

Gramene added 12 new fully sequenced reference genomes (Supplementary Table S1), bringing the total to 39. The composition of selected species reflected community needs, balanced with other considerations such as assembly quality, and placement within the plant phylogenetic tree, as well as collaborative support. In addition, all reference genome assemblies are accessioned with the International Nucleotide Sequence Database Collaboration (INSDC; http://www.insdc.org) (6), a data sharing policy standard adopted by Gramene and Ensembl Plants. As shown in Figure 1A and Supplementary Table S1, the list of included species provides broad taxonomic representation across green plants, with greater representation in particular clades, such as rice, wheat and the Brassicas. In all, the set includes the genomes of 21 monocots, 12 core eudicots, 1 basal angiosperm and 5 non-flowering plants, which collectively serve both crop and model organism research communities. Five of the newly added species are wild relatives of rice: *Oryza meridionalis*, *O. glumaepatula*, *O. nivara*, *O. rufipogon*, *O. longistaminata* and the rice outgroup *Leersia perrieri*, which were sequenced by an international consortium of researchers known as I-OMAP (7). With the addition of these species, Gramene now hosts data for almost half of the 23 extant species in the *Oryza* genus, providing an unprecedented resource for rice research. Deep species representation is also provided within the *Triticaceae*, including the first chromosome-scale assembly of hexaploid bread wheat (8,9), now the largest genome (16 Gbp) in Gramene. Newly added eudicot genomes include three nutritionally and economically important species: the cole crops (*Brassica oleracea*) (10), cocoa (*Theobroma cacao*) (11) and peach (*Prunus persica*) (12). Two other new species are of particular interest in diverse fields of research. The addition of *Amborella trichopoda* (13,14), basal to Angiosperms (flowering plants), should aid in efforts to elucidate archaic and derived genomic features within the angiosperm tree of life. The genome of one of the smallest known free-living eukaryotes, a unicellular green alga, *Ostreococcus lucimarinus* (15), has attracted intense research interest due to its central role in the oceanic carbon cycle.

ANNOTATION AND COMPARATIVE GENOMICS

A principal aim of Gramene is to integrate and draw connections between genomes through consistent functional annotation and comparative analyses. For each genome, the community-recognized gene annotation set is characterized for InterPro domains, assigned Gene Ontology (GO) and Plant Ontology terms (Supplementary Table S2) (16) and cross-referenced to source databases. Genomes are mapped by whole genome alignment between selected pairs of species (17,18). In the past year, the number of available pairwise alignments increased from 64 to 133, with particular emphasis on comparisons within the rice and wheat-related species (Supplementary Table S3). In addition, protein-coding genes from all species were used to build phylogenetic gene trees using the Ensembl Compara Gene Trees pipeline (19). In addition to providing inferred evolutionary histories of gene families, this pipeline identifies orthologous relationships between genes, which are subsequently used to build synteny maps between sufficiently related species (Supplementary Table S4) (5). The gene tree pipeline also outputs contiguous 'split' gene models frequently associated with either the artifacts of misannotation or evolutionary adaptations. We provide the split gene data as species-wise downloadable files with each release (Supplementary Table S5).

Advances in genomics technologies and research over the last several years have provided new types of data for Gramene to display, particularly in the realms of gene expression, regulation and epigenetics. RNA-seq alignment tracks, representing whole transcriptome expression in diverse tissues, conditions or genetic backgrounds, are now displayed for wheat, maize (20) and all rice-related species, either as publicly accessible data tracks or importable gene expression Atlas data, as described later. Methylation site data, derived from bisulphite sequencing of two maize genetic backgrounds (21) are now displayed in the maize genome browser (Figure 1B). New data and software advances have also contributed to improved annotation tracks in maize, including long non-coding RNAs (21) and protein coding genes annotated with MAKER-P (22,23).

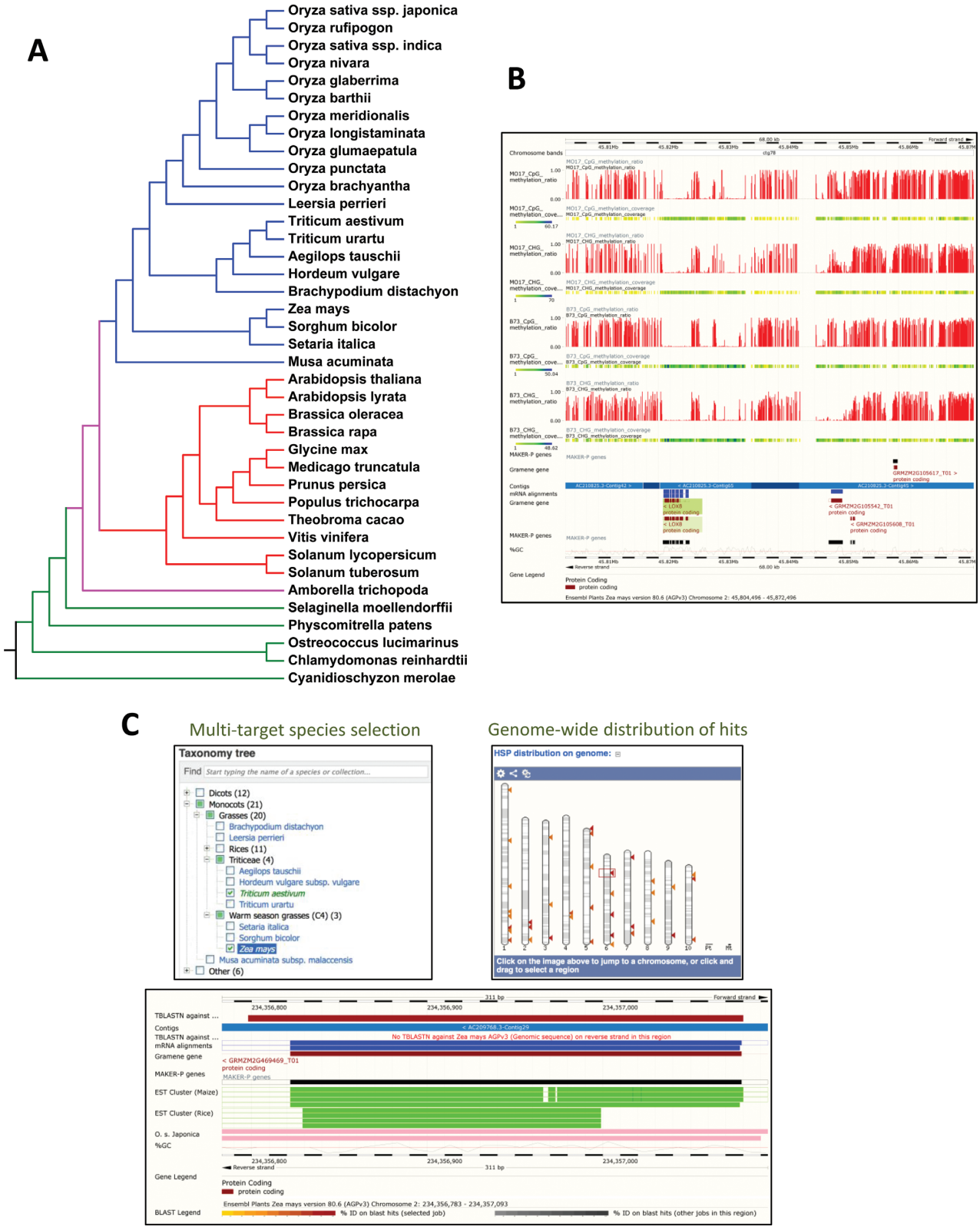

Figure 1. New genome visualizations and improvements to the Gramene database. (**A**) Species tree of 39 plant reference genomes available in Gramene. (**B**) Visualization of maize methylomic variation (Regulski *et al.*, 2013). (**C**) Improvements to BLAST include multiple target species selection, process tracking system and data integration with genome browser visualizations.

PLANT GENETIC DIVERSITY AND SEQUENCE VARIATION

Gramene provides SNP and/or structural variation data sets for 11 genomes (Supplementary Table S6 (24–39)). The new bread wheat variation data include over 1.5 million SNPs from the Wheat HapMap study (36), about 725 000 non-redundant SNPs imported from CerealsDB (Axiom, iSelect and KASP array sets; http://www.cerealsdb.uk.net) and over 10.3 million inter-homoeologous variants. As part of the wheat genome analysis, we provide homoeologous SNPs (i.e. variants between each two of the A, B and D wheat component genomes (40) (Kersey *et al.*, this issue)), as well as their corresponding sequence alignments against the *Hordeum vulgare* genome in the wheat and barley genome browsers. Sorghum variation was enriched with over 6.5 million SNPs genotyped in 45 *Sorghum bicolor* lines by aligning them against the BTx623 reference genome, plus two *S. propinquum* lines reported by Mace *et al.* (35), and about 265 000 SNPs genotyped in 378 lines from the US Sorghum Association Panel (SAP) (34). The Panzea 2.7 genotyping-by-sequencing (GBS) data set (http://www.panzea.org) consisting of almost 720 000 SNPs typed in 16 718 maize and teosinte lines was newly added to our maize genetic variation collection. The barley variation collection benefitted from the addition of about 5 million SNPs from 90 Morex × Barke individuals and about 6 million variants from 84 Oregon Wolfe barley individuals, determined by POPSEQ (41), in addition to ∼2600 markers associated with the SNPs from the Illumina iSelect 9k barley SNP chip (42). The new variation data for tomato include over 71 million SNPs obtained from whole-genome sequencing of 84 tomato accessions and related species (32). SNPs identified in the de-novo transcriptome studies on *T. monococcum* (∼500 000 SNPs) (29) and *Brachypodium distachyon* (∼340 000 SNPs) (27) are also being provided. We continue to assign putative functional and structural consequences to variants on the genes analyzed with the Ensembl variant effect predictor (VEP) tool (43). Consequences can be visualized 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.

BETTER BLAST

Each Gramene release brings new features through advances in the Ensembl software infrastructure. As of July 2015, Gramene makes use of Ensembl version 80, and we improved our plant-specific Basic Local Alignment Search Tool (BLAST; Figure 1C). BLAST improvements include better integration of input and output with the genome browser (e.g. direct input from a DNA/protein sequence view page; results linked to browser views including genome-wide distribution of hits, local alignments, associated genes, etc.), multi-species target selection and tracking system for job progress and completion.

PLANT PATHWAYS

The pathway portal of Gramene (http://www.gramene.org/pathways) provides resources for comparative analysis of plant pathways across several model and crop plant species. The Plant Reactome (http://plantreactome.gramene.org) a database of plant metabolic and regulatory pathways, was developed collaboratively by Gramene and the Human Reactome project (44). The Reactome data model organizes gene products, small molecules and macromolecular interactions into reactions and pathways to build a systems-level framework of an eukaryotic organism and to illustrate how it is influenced by intrinsic developmental cues and in response to various signals originating in the surrounding environment.

The Plant Reactome features *O. sativa* as a reference species. To build the reference pathway database, cellular-level metabolic networks from RiceCyc (45) were imported into the Reactome data structure via BioPax v2.0. The database was further enriched by adding a small set of highly conserved signaling and regulatory pathways (e.g. cell cycle, DNA replication, transcription, translation, etc.) that were projected from the curated human Reactome based on gene homology. From this initial platform, we continue to curate metabolic and regulatory pathways in rice. Recently, we began to curate and link signaling and regulatory pathways. The signaling pathways involve various plant hormones, such as brassinosteroid, auxin, ethylene, abscisic acid and strigolactone. Currently, the Plant Reactome database contains ∼240 rice reference pathways of which 200 were manually curated.

The pathways are organized based on a hierarchical classification (Figure 2A). The detailed view of a pathway shows a graphical display of the various associated reactions, enzymes, metabolites and cofactors in the context of their subcellular locations (Figure 2B). Typically, details and summary of various entities (e.g. overview, molecules, structures, download links, etc.) are provided below the pathway diagram, as shown in Figure 2. Links are also provided to Gramene gene loci and several external resources like UniProt, PIR, ChEBI, PubChem, PubMed and GO. For advanced users, the data are also accessible for download in various formats from the download tab and via APIs.

We developed gene homology-based pathway projections for 33 plant species (http://plantreactome.gramene.org/stats.html) by using *O. sativa* ssp. *japonica* as a reference. Gene IDs and corresponding UniProt IDs of genes from curated rice pathways were used as a reference to identify orthologs from plant species with a sequenced genome and/or transcriptomes, even when the data might not have been served from the Gramene browser. For example, we applied Inparanoid clustered orthology data on annotated leaf transcriptomes of wild *Oryza* species provided by the OMAP project (46) and the *O. sativa* Aus Kasalath genome (47), which are not hosted by Gramene.

In addition to search and browsing of pathways, the new version of Plant Reactome allows analysis of expression data (e.g. transcriptome, proteome, metabolome, etc.). Users can upload custom data to conduct pathway enrichment analysis and/or compare rice reference pathways with projected pathways on another plant species. To illustrate the utility of the pathway enrichment tool, we uploaded tissue-specific expression data from *Oryza sativa* cv Nipponbare (Figure 2B) provided by the EMBL-EBI Expression Atlas project (48) (Petryszak *et al.*, this issue).

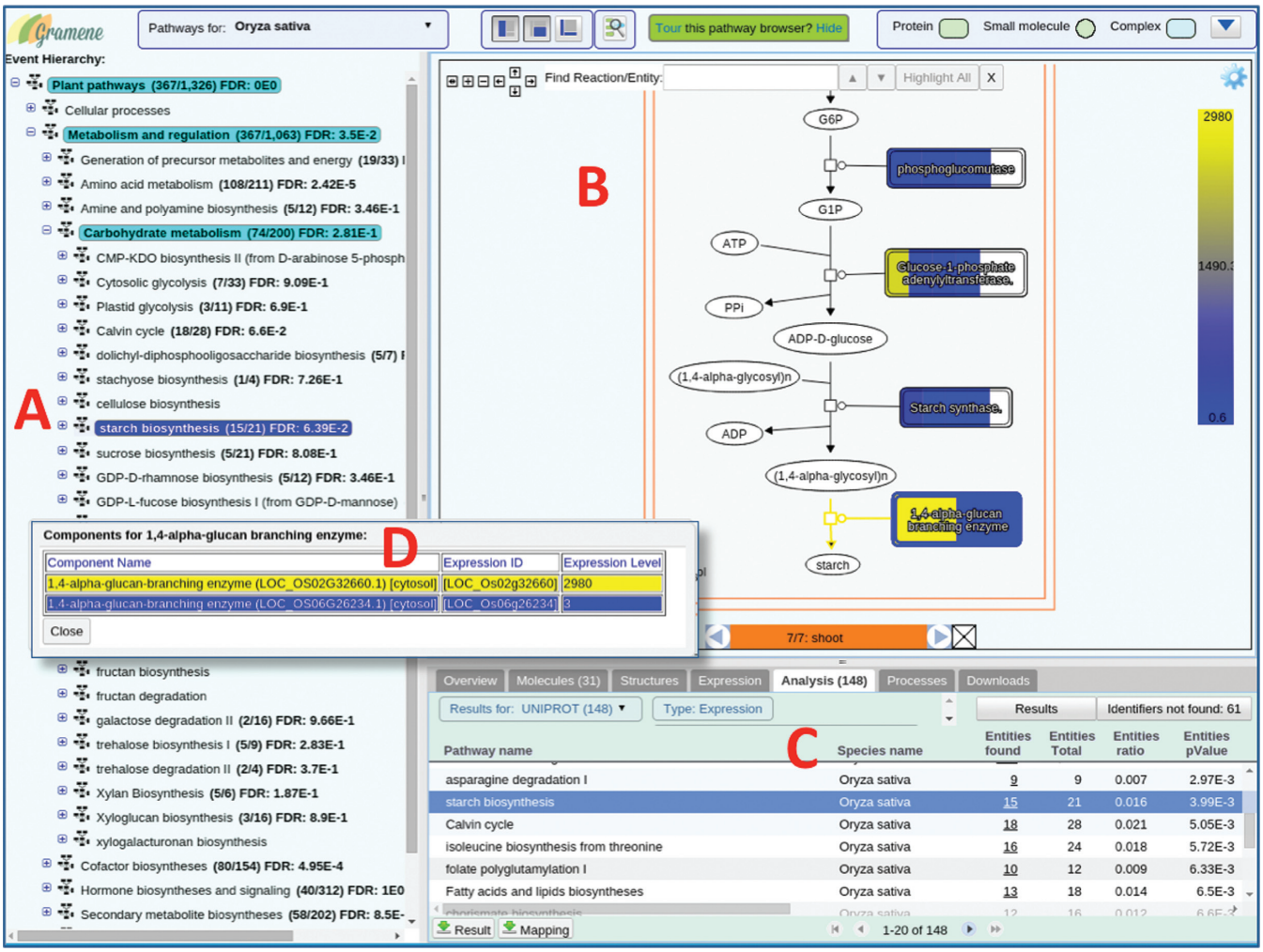

Figure 2. A view of Gramene's Plant Reactome Pathway browser (http://plantreactome.gramene.org/) showing the gene expression data overlay feature of the Analysis Tools. The left panel (**A**) hierarchically lists pathways from reference species *Oryza sativa* along with the number of gene entities mapped from the gene expression data and the FDR value. The upper-right panel (**B**) shows the network of entities from the starch biosynthesis pathway painted with the EBI Atlas baseline expression profiles (Petryszak *et al.*, 2013) of associated genes across seven tissue types from *O. sativa* cv. Nipponbare (http://www.ebi.ac.uk/gxa/experiments/E-MTAB-2037). Gene products colored in blue and yellow indicate low and high relative expression values, respectively. Diagram entities simultaneously displaying multiple colors signify a defined set or complex of multiple gene products. Display options at the top of the browser allow users to toggle the display of left and right panels and view a diagram key defining pathway entities, reaction types and attributes. If the users are in the expression analysis panel as show here, they will see a data analysis table (**C**) that provides a list of mapped pathways, counts of genes and reactions and statistical parameters such as *P*-value, FDR and average expression. Users have options to download the analyzed data table. The pop-up dialogue box in this example (**D**) displays the constituent gene products for one such entity, along with their expression ID and level.

Gramene continues to provide access to the previously described Pathway Tools-based metabolic networks (3,45,49) from the iPlant Collaborative website (http://pathway.iplantcollaborative.org).

PLANT GENE EXPRESSION ATLAS

In collaboration with the EMBL-EBI's Expression Atlas (http://www.ebi.ac.uk/gxa) project (48) (Petryszak *et al.*, this issue), we provide information about gene expression assayed in samples of different cell types, organism parts, developmental stages, diseases and other conditions. These data include microarray and RNA-sequencing studies imported from ArrayExpress, selected for having at least three biological replicates and other quality measures. To systematize individual studies into a common framework, metadata was manually curated and annotated with ontology terms prior to processing using standardized analysis methods described by Petryszak *et al.* (this issue). Atlas data can be queried for both the baseline and differential expression of genes. Baseline expression data, provided only for carefully selected RNA-sequencing based studies, report transcript abundance estimates for each gene in all samples (control, treated tissues, cell types, collection of germplasms). Differential expression data report statistically significant differential gene expression in manually curated pairwise comparisons between two sets of biological replicates – the 'reference' set (e.g. 'healthy' or 'wild type') and a 'test' set (e.g. 'diseased' or 'mutant'). As of August 2015, Atlas contains 389 experiments in 11 species of plants (http://www.ebi.ac.uk/gxa/plant/experiments; e.g.

rice, wheat, maize and Arabidopsis), including seven baseline studies reporting expression in tissues, strains and cultivars. Atlas's ability to display expression across all tissues and all baseline studies, in a juxtaposed manner, in a single intuitive interface makes it easy for the user to spot correlated patterns of expression across multiple 'omics' studies. Atlas offers a number of visualizations, e.g. 'enrichment' of GO terms, Plant Reactome pathways and InterPro domains within each pairwise comparison. The user interface also allows researchers to select the experimental condition and the genes and to upload the expression data as a data track on the plant genome browsers provided by Gramene and Ensembl Plants.

NEW PUBLIC WEB SERVICE

Gramene now provides a public web service at http://data.gramene.org/ that collects and links data from Gramene's Ensembl and Plant Reactome resources, exposing them through a RESTful API alongside the Ensembl and Reactome REST APIs. This service is the basis for development of our web-based search, analysis and visualization tools, but is open to all developers.

In order to provide a comprehensive gene search service, we decided to integrate Gramene's Ensembl genes with pathway, Interpro domain, and plant and GO associations using MongoDB. Gramene's genes document collection (GDC) contains basic information on each gene (e.g. ID, name, description, species, genomic location, cross references, etc.), as well as references to Compara gene trees, associated GO and PO terms (50), Interpro domains and pathways. Separate document collections are maintained for each of these structured annotation types to facilitate customization and extension of the database in the future. After incorporation of some additional indexing fields, fields from the GDC are loaded into a Solr core. Auxiliary MongoDB collections are also loaded into Solr so that we can suggest relevant filters given a partial query string. We use Solr for its advanced querying and facet-counting features and MongoDB for its flexible support for structured documents.

Using this integrated service, it is possible to compose complex queries across all hosted genomes and obtain a variety of summary statistics for the genes in the result set. The web services provided through http://data.gramene.org are documented with Swagger and open to all developers.

WEBINARS AND VIDEO-TUTORIALS

In November 2014, Gramene began to offer monthly webinars focused on providing an overview of resources and functionalities available to users of the Gramene database for comparative plant genomics, as well as explaining how users can visualize and analyze data of their choice using various built-in tools. The recorded webinars are available on Gramene's YouTube channel (https://goo.gl/qQ2Pjn). Examples include webinars focused on specific plant species (e.g. Arabidopsis, rice, maize, etc.) and/or specific resources (e.g. genome browser, Plant Reactome, GrameneMart, expression data analysis, phylogenetic comparisons and gene trees, etc.). We are open to suggestions for webinar topics via e-mail (webinars@gramene.org), and we benefit from user feedback via post-webinar surveys (e.g. https://www.surveymonkey.com/r/3J2J2M9).

DISCUSSION AND FUTURE PERSPECTIVE

Currently in its 15th year, the Gramene database has become an unparalleled resource for integrated genomic and pathway information for major model and crop plants. Since the last NAR update in January 2014, Gramene added 12 new plant genome assemblies to a total of 39 fully sequenced reference genomes, many of which were also updated with newer assembly versions and/or annotations. We added ∼100 million new genetic variants to a net total of over 190 million plant genetic and structural variants from 11 plant species, scored on a diversity of germplasm accessions. Thus, the bulk of the new variation data set is comprised of over 70 million SNPs determined in 84 tomato accessions (http://www.tomatogenome.net) (32), over 12 million bread wheat SNPs (including 10 million of inter-homoeologous wheat variants, Wheat HapMap SNPs and wheat SNPs from the CerealsDB database), as well as maize SNPs typed in 16 718 maize and teosinte lines from the Panzea 2.7 GBS set (http://www.panzea.org/#!genotypes/cctl). In addition, we expanded our data collection describing the transcriptome and epigenome of economically important species like rice and maize.

In the next year, besides projected growth of existing genomic and genetic data, we aim to consolidate our phenotypic data collection by updating our QTL information, and storing and visualizing genome-wide association studies data.

Further user interface and search improvements include development of a comparative genomics-based search interface, available at http://search.gramene.org/ (beta version) that queries and connects data from the public web services described above. This enables users to find genes of interest by asking specific questions—rather than performing a general text search—and to see the distribution of results across all genomes in Gramene. The speed with which results are returned enables an 'exploratory' navigation style; when an interesting gene is identified, it can be used as the basis of further searches, e.g. to see all available homologs or find genes with similar domain structures. In its alpha release, the search interface is a JavaScript web application built using ReactJS interface library with visualizations developed in D3 and derived from the KBase platform. The web application is open source, with source code public and managed at Github. Releases are built using Node Package Manager ('npm') and deployed automatically using Travis-CI.

We will continue to interact, develop and provide value-added data and analysis tools to plant biologists, breeders, geneticists, molecular biologists, biochemists and bioinformatics researchers. With special emphasis on inter an intra-specific analysis of small and large-scale data sets generated from global crop and emerging plant model species, we expect that researchers will benefit by building hypotheses and/or validating the existing or new knowledge about novel genes, their function, expression, role in pathways, phenotypes and associated functional and structural variations.

Our monthly webinar series will continue to consider the specific needs of the plant community, addressing trending topics and targeting communities working on particular species that would most benefit from direct assistance for working with Gramene's data and resources.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Gramene's users, researchers and numerous collaborators for sharing data sets generated in their projects, valuable suggestions and feedback on improving the overall quality of Gramene as a community resource. We would also like to 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 also thank Peter van Buren from Cold Spring Harbor Laboratory for system administration support; David Croft (EBI) and Robin Haw (OICR) for support on Plant Reactome development; undergraduate students Dylan Beorchia, Teague Green and Kindra Amoss (OSU) for their help in pathway curation; Samuel Fox and Matthew Geniza (OSU) on gene expression Atlas curation and data analysis.

FUNDING

National Science Foundation [IOS-0703908 and IOS-1127112]; United States Department of Agriculture - Agricultural Research Service [58-1907-4-030 and 1907-21000-030-00D to D.W.]; European Community's 7th Framework Programme (FP7/2007-2013; Infrastructures) [contract # 283496 to P.K.]; United Kingdom Biotechnology and Biosciences Research Council [BB/J000328X/1, I008071/1 and H531519/1 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 [P41 HG003751 to L.S.]; EU grant [LSHG-CT-2005-518254]; 'ENFIN', Ontario Research Fund; EBI Industry Programme. The funders had no role in the study design, data analysis or preparation of the manuscript. Funding for open access charge: "Gramene - Exploring Function through Comparative Genomics and Network Analysis" [NSF award #1127112].
Conflict of interest statement. None declared.

REFERENCES

1. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
2. Croft,D. (2013) Building models using Reactome pathways as templates. *Methods Mol. Biol.*, **1021**, 273–283.
3. 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.
4. Spooner,W., Youens-Clark,K., Staines,D. and Ware,D. (2012) GrameneMart: the BioMart data portal for the Gramene project. *Database: J. Biol. Databases Curation*, bar056.
5. 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.
6. 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.
7. Jacquemin,J., Bhatia,D., Singh,K. and Wing,R.A. (2013) The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr. Opin. Plant Biol.*, **16**, 147–156.
8. International Wheat Genome Sequencing, C. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, **345**, 1251788.
9. Chapman,J.A., Mascher,M., Buluc,A., Barry,K., Georganas,E., Session,A., Strnadova,V., Jenkins,J., Sehgal,S., Oliker,L. *et al.* (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.*, **16**, 26.
10. Parkin,I.A., Koh,C., Tang,H., Robinson,S.J., Kagale,S., Clarke,W.E., Town,C.D., Nixon,J., Krishnakumar,V., Bidwell,S.L. *et al.* (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol.*, **15**, R77.
11. Motamayor,J.C., Mockaitis,K., Schmutz,J., Haiminen,N., Livingstone,D. 3rd, Cornejo,O., Findley,S.D., Zheng,P., Utro,F., Royaert,S. *et al.* (2013) The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.*, **14**, r53.
12. International Peach Genome, I., Verde,I., Abbott,A.G., Scalabrin,S., Jung,S., Shu,S., Marroni,F., Zhebentyayeva,T., Dettori,M.T., Grimwood,J. *et al.* (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.*, **45**, 487–494.
13. Amborella Genome,P. (2013) The Amborella genome and the evolution of flowering plants. *Science*, **342**, 1241089.
14. Chamala,S., Chanderbali,A.S., Der,J.P., Lan,T., Walts,B., Albert,V.A., dePamphilis,C.W., Leebens-Mack,J., Rounsley,S., Schuster,S.C. *et al.* (2013) Assembly and validation of the genome of the nonmodel basal angiosperm Amborella. *Science*, **342**, 1516–1517.
15. Palenik,B., Grimwood,J., Aerts,A., Rouze,P., Salamov,A., Putnam,N., Dupont,C., Jorgensen,R., Derelle,E., Rombauts,S. *et al.* (2007) The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 7705–7710.
16. Kersey,P.J., Staines,D.M., Lawson,D., Kulesha,E., Derwent,P., Humphrey,J.C., Hughes,D.S., Keenan,S., Kerhornou,A., Koscielny,G. *et al.* (2012) Ensembl Genomes: an integrative resource for genome-scale data from non-vertebrate species. *Nucleic Acids Res.*, **40**, D91–D97.
17. Kent,W.J., Baertsch,R., Hinrichs,A., Miller,W. and Haussler,D. (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 11484–11489.
18. Harris,R.S. (2007) *Improved pairwise alignment of genomic DNA*. PhD Thesis. Pennsylvania State University.
19. 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.
20. Erhard,K.F. Jr, Talbot,J.E., Deans,N.C., McClish,A.E. and Hollick,J.B. (2015) Nascent transcription affected by RNA polymerase IV in Zea mays. *Genetics*, **199**, 1107–1125.
21. 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.
22. 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.
23. Law,M., Childs,K.L., Campbell,M.S., Stein,J.C., Olson,A.J., Holt,C., Panchy,N., Lei,J., Jiao,D., Andorf,C.M. *et al.* (2015) Automated update, revision, and quality control of the maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. *Plant Physiol.*, **167**, 25–39.
24. 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.
25. 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.
26. 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.
27. 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**.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. Jordan,K.W., Wang,S., Lun,Y., Gardiner,L.J., MacLachlan,R., Hucl,P., Wiebe,K., Wong,D., Forrest,K.L., Consortium,I. *et al.* (2015) A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol.*, **16**, 48.
37. 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.
38. Gore,M.A., Chia,J.M., Elshire,R.J., Sun,Q., Ersoz,E.S., Hurwitz,B.L., Peiffer,J.A., McMullen,M.D., Grills,G.S., Ross-Ibarra,J. *et al.* (2009) A first-generation haplotype map of maize. *Science*, **326**, 1115–1117.
39. 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.
40. Brenchley,R., Spannagl,M., Pfeifer,M., Barker,G.L., D'Amore,R., Allen,A.M., McKenzie,N., Kramer,M., Kerhornou,A., Bolser,D. *et al.* (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, **491**, 705–710.
41. Mascher,M., Muehlbauer,G.J., Rokhsar,D.S., Chapman,J., Schmutz,J., Barry,K., Munoz-Amatriain,M., Close,T.J., Wise,R.P., Schulman,A.H. *et al.* (2013) Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.*, **76**, 718–727.
42. Comadran,J., Kilian,B., Russell,J., Ramsay,L., Stein,N., Ganal,M., Shaw,P., Bayer,M., Thomas,W., Marshall,D. *et al.* (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.*, **44**, 1388–1392.
43. 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.
44. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The Reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.
45. 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*, **6**, 15.
46. 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.
47. 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.
48. Petryszak,R., Burdett,T., Fiorelli,B., Fonseca,N.A., Gonzalez-Porta,M., Hastings,E., Huber,W., Jupp,S., Keays,M., Kryvych,N. *et al.* (2014) Expression Atlas update–a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **42**, D926–D932.
49. 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**.
50. 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.