

## Gramene: A Resource for Comparative Analysis of Plants Genomes and Pathways

Marcela Karey Tello-Ruiz, Joshua Stein, Sharon Wei, Ken Youens-Clark, Pankaj Jaiswal, and Doreen Ware

### Abstract

Gramene is an integrated informatics resource for accessing, visualizing, and comparing plant genomes and biological pathways. Originally targeting grasses, Gramene has grown to host annotations for economically important and research model crops, including wheat, potato, tomato, banana, grape, poplar, and *Chlamydomonas*. Its strength derives from the application of a phylogenetic framework for genome comparison and the use of ontologies to integrate structural and functional annotation data. This chapter outlines system requirements for end users and database hosting, data types and basic navigation within Gramene, and provides examples of how to (1) view a phylogenetic tree for a family of transcription factors, (2) explore genetic variation in the orthologues of a gene with a known trait association, and (3) upload, visualize, and privately share end user data into a new genome browser track.

Moreover, this is the first publication describing Gramene's new web interface—intended to provide a simplified portal to the most complete and up-to-date set of plant genome and pathway annotations.

**Key words** Plant genome, Reference genomes, Comparative genomics, Phylogenetics, Gene homology, Synteny, Genetic variation, Structural variation, Plant pathways

---

### 1 Introduction

Gramene (<http://www.gramene.org>; Fig. 1) is a curated online resource for comparative functional genomics in socioeconomically important crops and research model plant species, currently hosting over 30 completely sequenced plant reference genomes (Table 1; [1–31]). Each plant genome features community-based gene annotations provided by primary sources and enriched with supplemental annotations from cross-referenced sources, functional classification, and comparative phylogenomics analysis performed in-house. Increasing amounts of genetic and structural variation data derived both from data repositories and through collaboration with large-scale resequencing and genotyping initiatives are also available for visualization and analysis (Table 2; [32–40]).

Gramene: A comparative resource for plants | Gramene

Gramene: A comparative resour...

www.gramene.org

Google

Search

Home Release Notes Collaborators Contact About Archive Citing Gramene Outreach

Navigation


- Current Release (40)
- Search
- Genomes
- Pathways
- BLAST
- Gramene Mart
- News
- Archive (Build 39)
- Download
- Web Services
- Contact
- Tools

Recent blog posts

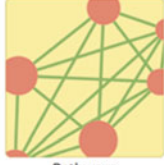
- Gramene at the Maize Genetics Conference 2014
- PAG 2014: Some Highlights
- Gramene workshop at the Maize Genetics Conference 2014 — See you in Beijing!
- Tenure-track faculty position in Plant-Microbe Interactions at Oregon State University
- ABA Biosynthesis and Signaling

More

## Gramene: A comparative resource for plants



Genomes



Pathways

Gramene is a curated, open-source, integrated data resource for comparative functional genomics in crops and model plant species. Our goal is to facilitate the study of cross-species comparisons using information generated from projects supported by public funds. Gramene currently hosts annotated whole genomes in over two dozen plant species and partial assemblies for almost a dozen wild rice species in the Ensembl browser, genetic and physical maps with genes, ESTs and QTLs locations, genetic diversity data sets, structure-function analysis of proteins, plant pathways databases (BioCyc and Plant Reactome platforms), and descriptions of phenotypic traits and mutations.

### Gramene Portals

- Genome Browser:** Browse gene annotations & diversity data
- BLAST:** Align DNA & protein sequences
- Plant Reactome:** Browse metabolic & regulatory pathways
- Pathways databases:** BioCyc based cellular metabolic networks for 10 plant species
- Gramene Mart:** Customized data queries
- Bulk downloads**
- ARCHIVE - Markers, Proteins and Ontology databases, QTLs, Comparative Maps**

**Fig. 1** Gramene's homepage

Furthermore, plant pathway databases generated by applying both manual curation and automated methods complement available sequence-based gene annotations. Most advantageous to plant researchers and bioinformaticians is that, by using a core set of consistently applied protocols, Gramene offers a reference resource for basic and translational research in plants.

Gramene is driven by several platform infrastructures or modules that are linked to provide a unified user experience. Its Genome Browser (<http://ensembl.gramene.org>; Fig. 2) takes advantage of the Ensembl project's infrastructure [41] to provide an interface for exploring genome features, functional ontologies, variation

**Table 1**  
**Plant reference genome sequences in Gramene build 41 (May 2014)**

Species	Reference genome status	Assembly/gene space annotation	Literature references
<i>Aegilops tauschii</i> (goatgrass, wheat D-genome progenitor)	Complete draft	GCA 000347335.1/2013-12-BGI	Jia et al. [1]
<i>Amborella trichopoda</i>	Complete draft	AMTRI.0 (GCA_000471905.1)/2014-01-AGD	Amborella Genome Project [2]; Chamala et al. [3]
<i>Arabidopsis lyrata</i>	Complete draft	v.1.0/2008-12-Araly1.0	Hu et al. [4]
<i>Arabidopsis thaliana</i>	Complete draft	TAIR10/2010-09-TAIR10	Arabidopsis Genome Initiative [5]
<i>Brachypodium distachyon</i>	Complete draft	v1.0/2010-02-Brachy1.2	The International Brachypodium Initiative [6]
<i>Brassica rapa</i> (Chinese cabbage)	Complete draft	IVFCAASv1/bra_v1.01_SP2010_01	Wang et al. [7]
<i>Chlamydomonas reinhardtii</i> (green algae)	Complete draft	v3.1 (GCA_000002595.2)/2007-11-ENA	Merchant et al. [8]
<i>Cyanidioschyzon merolae</i> (red algae)	Complete draft	ASM9120v1/2008-11-ENA	Matsuzaki et al. [9]
<i>Glycine max</i> (soybean)	Complete draft	V1.0 (GCA_000004515.1)/JGI-Glyma-1.1	Schmutz et al. [10]
<i>Hordeum vulgare</i> (barley)	Complete draft	030312v2/IBSC 1.0	The International Barley Genome Sequencing Consortium [11]
<i>Medicago truncatula</i>	Complete draft	GCA 000219495.1/2011-11-EnsemblPlants	Young et al. [12]
<i>Musa acuminata</i> (banana)	Complete draft	MA1/2012-08-Cirad	D'Hont et al. [13]
<i>Oryza barthii</i>	Complete draft	O.barthii_v1 (GCA_000182155.2)/OGE-MAKER	OGE/OMAP (NSF award #1026200)
<i>Oryza brachyantha</i>	Complete draft	Oryza brachyantha.v1.4b (GCA_000231095.2)/OGE-MAKER	Chen et al. [14]
<i>Oryza glaberrima</i>	Complete draft	AGII.1/2011-05-AGI	OGE/OMAP (NSF award #1026200)
<i>Oryza glumaepatula</i>	Complete draft	Oryza_glumaepatula_v1.5 (GCA_000576495.1)/OGE-MAKER	OGE/OMAP (NSF award #1026200)

(continued)

**Table 1**  
**(continued)**

<b>Species</b>	<b>Reference genome status</b>	<b>Assembly/gene space annotation</b>	<b>Literature references</b>
<i>Oryza meridionalis</i>	Complete draft	Oryza_meridionalis_v1.3 (GCA_000338895.2)/OGE-MAKER	OGE/OMAP (NSF award #1026200)
<i>Oryza nivara</i>	Complete draft	Oryza_nivara_v1.0 (GCA_000576065.1)/OGE-MAKER	OGE/OMAP (NSF award #1026200)
<i>Oryza punctata</i>	Complete draft	Oryza_punctata_v1.2 (GCA_000573905.1)/OGE-MAKER	OGE/OMAP (NSF award #1026200)
<i>Oryza sativa ssp. indica</i>	Complete draft	ASM465v1 (GCA_000004655.2)/2010-07-BGI	Yu et al. [15]; Zhao et al. [16]
<i>Oryza sativa ssp. japonica</i> (rice)	Complete draft	IRGSP-1.0/MSU7 (IRGSP-1.0)	International Rice Genome Sequencing [17]; Kawahara et al. [18]
<i>Physcomitrella patens</i> (moss)	Complete draft	ASM242v1/2011-03-Phypal.6	Rensing et al. [19]
<i>Populus trichocarpa</i> (poplar)	Complete draft	JGI 2.0/2010-01-JGI	Tuskan et al. [20]
<i>Prunus persica</i> (peach)	Complete draft	Prupe1_0 (GCA_000346465.1)/2013-03	Verde et al. [21]
<i>Selaginella moellendorffii</i> (spikemoss)	Complete draft	v1.0/2011-05-ENA	Banks et al. [22]
<i>Setaria italica</i> (foxtail millet)	Complete draft	JGIv2.0/JGIv2.1	Bennetzen et al. [23]; Zhang et al. [24]
<i>Solanum lycopersicum</i> (tomato)	Complete draft	SL2.40/ITAG2.3	Tomato Genome Consortium [25]
<i>Solanum tuberosum</i> (potato)	Complete draft	3.0/SolTub 3.0	Potato Genome Sequencing Consortium [26]
<i>Sorghum bicolor</i>	Complete draft	Sorbi1/2007-12-JGI (Sbi1.4)	Paterson et al. [27]
<i>Triticum aestivum</i> (bread wheat)	Complete chromosome survey	IWGSP1/MIPS2.1	International Barley Genome Sequencing Consortium et al. [11]; Brenchley et al. [28]

(continued)

**Table 1**  
**(continued)**

Species	Reference genome status	Assembly/gene space annotation	Literature references
<i>Triticum urartu</i> (einkorn wheat, A-genome progenitor)	Complete draft	ASM34745v1 (GCA 000347455.1)/2013-04-BGI	Ling et al. [29]
<i>Vitis vinifera</i> (grape)	Complete draft	IGGP_12X/2012-07-CRIBI	Jaillon et al. [30]; Myles et al. [31]
<i>Zea mays</i> (corn)	Complete draft	B73 RefGen AGPv3/5b+	Schnable et al. [31]; Wei et al. [31]
<i>Leersia perrieri</i> (chr. 3s)	Partial	454.pools.2012Feb/2012-10-CSHL	OGE/OMAP (NSF award #1026200)
<i>Oryza granulata</i> (chr. 3s)	Partial	454.pools.2012Feb/2012-10-CSHL	OGE/OMAP (NSF award #1026200)
<i>Oryza longistaminata</i> (chr. 3s)	Partial	OGE.2012Jul/2012-10-CSHL	OGE/OMAP (NSF award #1026200)
<i>Oryza minuta BB</i> (chr. 3s)	Partial	BAC.Sanger.1.1 (May 2011)/ CSHLv3.1	OGE/OMAP (NSF award #1026200)
<i>Oryza minuta CC</i> (chr. 3s)	Partial	BAC.Sanger.1.1 (May 2011)/ CSHLv3.1	OGE/OMAP (NSF award #1026200)
<i>Oryza officinalis</i> (chr. 3s)	Partial	BAC.Sanger.1.1 (May 2011)/ CSHLv3.1	OGE/OMAP (NSF award #1026200)
<i>Oryza rufipogon</i> (chr. 3s)	Partial	454.pools.1.1 (Jul 2010)/CSHL	OGE/OMAP (NSF award #1026200)

data, and comparative phylogenomics. Since 2009 Gramene has partnered with the Plants division of Ensembl Genomes [42] to jointly produce this resource, each benefiting from the other's proximity to research communities in the USA and Europe, respectively. This collaboration has also facilitated timely adoption of innovative tools and software updates that accompany frequent version releases by the Ensembl project [41].

Since the last edition of this volume in 2007 [43], Gramene has also become a portal for pathway databases developed and curated internally or mirrored from external sources. Two pathway platforms are currently supported: (1) Gramene's Pathway Tools (<http://pathway.gramene.org>; Fig. 3) to emphasize the annotation of metabolic and transport pathways [44–46], and (2) Plant Reactome (<http://plantreactome.gramene.org>; Fig. 4) to facilitate the annotation of metabolic and regulatory pathways. The Pathway Tools

**Table 2**  
**Genetic and structural variation data in Gramene build 41 (May 2014)**

Species	Variants	Source	Studies
<i>Arabidopsis thaliana</i>	14,234,197 SV: 13,667	250K SNPs×1179 accessions 1001 genomes project: 411 resequenced accessions	Atwell et al. [32]
<i>Brachypodium distachyon</i>	327,988	394K SNPs×3 accessions of <i>Brachypodium sylvaticum</i>	Fox et al. [33]
<i>Oryza glaberrima</i>	7,704,409	Resequenced 20 accessions & 19 accessions of its wild progenitor ( <i>Oryza barthii</i> )	OGE/OMAP (NSF award #1026200)
<i>Oryza sativa ssp. japonica</i>	3,332,525	160K SNPs×20 accessions 1311 SNPs×395 accessions ~4 M BGI Japonica vs. Indica SNPs	McNally et al. [34] Zhao et al. [35] Yu et al. [36] NCBI dbSNP OGE/OMAP (NSF award #1026200)
<i>Oryza sativa ssp. indica</i>	4,747,883	~4 M BGI Japonica vs Indica SNPs	Yu et al. [36] NCBI dbSNP OGE/OMAP (NSF award #1026200)
<i>Sorghum bicolor</i>	257,153 SV: 64,507	265K SNPs×336 SAP lines Structural variants from Database of Genomic Variants archive (dGVA)	Morris et al. [37] Zheng et al. [38]
<i>Vitis vinifera</i>	457,404	Resequencing USDA germplasm collection	Myles et al. [31]
<i>Zea mays</i>	50,719,843	HapMap1: NAM founder lines HapMap2: 103 pre-domesticated & domesticated lines	Gore et al. [39] Chia et al. [40]

platform [47] supports the implementation of pathway databases in the BioCyc collection [48] to which Gramene has contributed MaizeCyc [45], RiceCyc [44], SorghumCyc [46], and BrachyCyc [46]. In addition, this resource mirrors six databases for *Arabidopsis* (AraCyc [49]), medicago (MedicCyc [50]), poplar (PoplarCyc [51]), potato (PotatoCyc [52]), coffee (CoffeaCyc [52]), and tomato (Lycocyc [52]), as well as the MetaCyc [48] and PlantCyc [51] reference databases (Fig. 3). The Plant Reactome is based on the Reactome data model and visualization platform [53]. It currently hosts manually curated rice and *Arabidopsis* pathways, and gene homology-driven inferred pathway projections for the maize and *Arabidopsis thaliana* reference genomes. It will continue to grow with the addition of data for new species and broader coverage of molecular interactions.

Search: All species for  Go

e.g. Carboxy\* or chx28

**Popular genomes**

- Arabidopsis thaliana** TAIR10
- Oryza sativa Japonica (Rice)** IRGSP-1.0
- Triticum aestivum** IWGSP1
- Hordeum vulgare** 030312v2
- Zea mays** AGP-v3
- Physcomitrella patens** ASM242v1

**All genomes**

-- Select a species --

[View full list of all Ensembl Plants species](#)

**What's New in Release 40**

- Updated genomes
  - Updated gene models for *Oryza sativa* from rap-db.
- New data
  - Gene models for *Triticum aestivum* (bread wheat) chromosome survey sequences from MIPS.
  - DNA-DNA alignments between bread wheat and *Oryza sativa*, as well as *Brachypodium distachyon*.
  - DNA-DNA alignments between *Hordeum vulgare* (barley) and the wild wheat progenitors (*Triticum urartu* and *Aegilops tauschii*).

**Did you know...?**

The **Assembly Converter Tool** allows the conversion of a variety of annotation formats between different versions of a genome assembly.

**Bread wheat genome and gene annotation released!**

The bread wheat genome in Ensembl Plants is the chromosome survey sequence for *Triticum aestivum* cv. Chinese Spring generated by the **International Wheat Genome Sequencing Consortium**. The gene models are provided by MIPS (version 2.0). A total of 108,569 protein coding genes have been predicted.

See also the wheat homepage at [Z. UGI](#)

[Read more about the assembly, annotation and analysis of bread wheat provided by Ensembl Plants...](#)

**Organelle Annotation**

For annotations relating to Organelles, see the [organelles page](#)

**Fig. 2** Gramene Ensembl Genome Browser homepage (Ensembl software v75)

The Genomes and Pathway modules enable species-specific and cross-species data downloads for discrete region(s), gene(s) or gene feature(s) via the Genome Browser, and pathway-centered downloads via the Pathways portal and Plant Reactome. In addition, project data is available for customizable downloads from the GrameneMart utility (<http://ensembl.gramene.org/Tools/Blast?db=core> [54]), nucleotide and protein sequence alignments via BLAST (<http://ensembl.gramene.org/Tools/Blast?db=core>), bulk downloads via file transfer protocol (FTP) at Gramene (<ftp://ftp.gramene.org/pub/gramene> and Ensembl Genomes (<http://ensembl.gramene.org/info/web-site/ftp/index.html>), and programmatic access via Ensembl's REST application programming interface (API) and public MySQL (<http://www.gramene.org/web-services> [55]). Since March 2013, the website, database, and its contents are being updated quarterly and updates can be followed from the Gramene news portal (<http://www.gramene.org/blog>) and by browsing the site's release notes (<http://www.gramene.org/release-notes>).

This chapter summarizes updates to the Gramene website and database since reported in the last edition of this volume [43].



Gramene: Plant Biochemical Pathway Database and Home to RiceCyc, MaizeCyc, BrachyCyc and SorghumCyc.

Gramene: Plant Biochemical Pat...

pathway.gramene.org

Google

**Gramene**

Pathways Home | Search Pathways | Omics Viewer | Omics Validator | Downloads | Help | Tutorial | FAQs | Release notes | Plant Reactome (beta)

**Plant Metabolic Pathways**

The pathways section in the Gramene databases is home for RiceCyc, MaizeCyc, BrachyCyc and SorghumCyc, the pathway databases for rice, maize, *Brachypodium*, and sorghum, respectively. It also provides mirrors of pathway databases from *Arabidopsis*, tomato, potato, pepper, coffee, *Medicago*, *E. coli*, and the MetaCyc and PlantCyc reference databases, and might enable comparative analysis again upon software upgrade that supports the most current version of ptools software 17.0 by the original sources, such as the [SolGenomics Network](#) (SGN). In addition to search and browse functions, the database allows users to find genes mapped to respective reactions and pathways and draw inetspecific comparison between the pathways.

**Pathways Browse and Other Options**

Click on the species specific links such as **browse** to go through the list of pathways; **summary** to get a summarized overview. Click on the **more info** link to learn more details on the respective pathway database.

<p>RiceCyc ver 3.3 <i>Oryza sativa japonica</i> Strain: Nipponbare <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>AraCyc* ver 10.0 <i>Arabidopsis thaliana</i> Strain: Columbia <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>EcoCyc* ver 17.0 <i>Escherichia coli</i> Strain: K-12 MG1655 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>
<p>SorghumCyc ver 1.1 <i>Sorghum bicolor</i> Strain: BTx623 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>MedicCyc* ver 1.0.1 <i>Medicago truncatula</i>, Barrelclover Unavailable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>MetaCyc* ver 17.0 Reference Pathway Database Strain: not applicable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>
<p>MaizeCyc ver 2.2 <i>Zea mays</i> Strain: B73 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>PoplarCyc* ver 5.0 <i>Populus trichocarpa</i> (and other <i>Populus</i> species and hybrids) Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>PlantCyc* ver 7.0 Plant Metabolic Pathway Database Strain: not applicable <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>
<p>BrachyCyc ver 2.0 <i>Brachypodium distachyon</i> Strain: Bd21 <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	<p>PotatoCyc* ver 1.0.1 <i>Solanum tuberosum</i>, Potato Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	
	<p>CoffeaCyc* ver 1.1.1 <i>Coffea canephora</i>, Coffee Strain: n/a <a href="#">Browse</a>   <a href="#">Summary</a>   <a href="#">More Info</a></p>	

**Fig. 3** Gramene's BioCyc pathways homepage. <http://pathway.gramene.org>

## 2 Materials

### 2.1 Hardware and Software Requirements for Users

A computer with internet access and a standard web browser such as Mozilla/Firefox, Internet Explorer, Chrome, or Safari.

### 2.2 Gramene System Components

Gramene is a web-based application that allows users to search and view biological data, making use where appropriate of graphics viewers such as the Ensembl genome browser or the Pathway Tools Omics viewer. Data is maintained in distinct relational databases (MySQL), and users connect to the site using a standard web browser. User queries for static (HTML) and dynamic content are negotiated by the Apache web server via the Solr search platform and a middleware layer written in Perl. Bulk downloads of data are provided through FTP sites at Gramene and Ensembl Genomes.



The image shows a screenshot of the Plant Reactome homepage. The browser window title is "Reactome" and the address bar shows "plantreactome.gramene.org/". The page has a green and white color scheme. At the top, there is a navigation menu with links for Home, About, Documentation, Tools, and Contact Us. Below this is a search bar with a magnifying glass icon and a list of search examples: LOC\_Os01g45760.1, YUC4, cytokinins, and glucose. There are buttons for "Advanced search", "Browse Pathways", and "Retrieve SBML".

The main content area is divided into several sections:

- About Plant Reactome:** A text block stating that Plant Reactome (now in beta) is a freely accessible plant pathway database. It describes how pathways are constructed through manual curation and computational projections from other databases like MetaCyc and Human Reactome. It also mentions cross-referencing to other bioinformatics databases.
- Featured pathway: Ethylene biosynthesis and signaling:** A complex metabolic pathway diagram showing the conversion of methionine to ethylene through various enzymatic steps, including the role of ACCase and ACCase2.
- Explore...:** A section titled "a sampling of metabolic and regulatory pathways in *Oryza sativa*:" listing several pathways such as Proline biosynthesis I, Lysine biosynthesis I, Jasmonic acid biosynthesis, Cytosolic glycolysis, COP11 Mediated Vesicle Transport, Gibberellin biosynthesis II, and Translation elongation.
- News and Notes:** A section with several bullet points:
  - New rice pathways:** Announcing 32 new rice pathways, including ethylene biosynthesis and signaling.
  - Gramene Release 40:** Announcing the 40th database release.
  - Plant Reactome will be presented at PAG XXII during the Gramene Workshop:** Announcing the presentation at the Plant & Animal Genome XXII conference in San Diego, CA, USA, from Jan 11-15.
  - Introductory Video Tutorial:** A link to a video presentation describing the main features of the Plant Reactome.
  - Useful Links:** A list of links to related databases: Gramene Plant Pathway Database, KEGG Pathway database, WikiPathways: Plants Portal, MetaCyc, Plant Metabolic Network, and Reactome.

At the bottom of the page, there are logos for Gramene, OSU, NSF, and the Ontario Institute for Cancer Research (OICR).

Fig. 4 Gramene's Plant Reactome homepage. <http://plantreactome.gramene.org>

### 2.3 Local Installation of Gramene

The minimum hardware configuration required for a local installation of Gramene consists of a desktop or server with a multicore CPU, 4GB of memory, and 500GB of disk space. Installation inside a virtual machine is possible. A recent distribution of Linux is required, such as Redhat/CentOS 6 or Ubuntu 12.x. Software packages required include Apache web server (*see Note 1*), Perl, PHP, MySQL, OpenJava, Drupal, and Apache Solr. Many of these can be installed via the distribution package management system

(yum, apt-get). Solr can be downloaded from the Apache Solr website. For specific installation instructions, contact Gramene developers at [feedback@gramene.org](mailto:feedback@gramene.org).

---

## 3 Methods

### 3.1 Basic Navigation of the Gramene Website

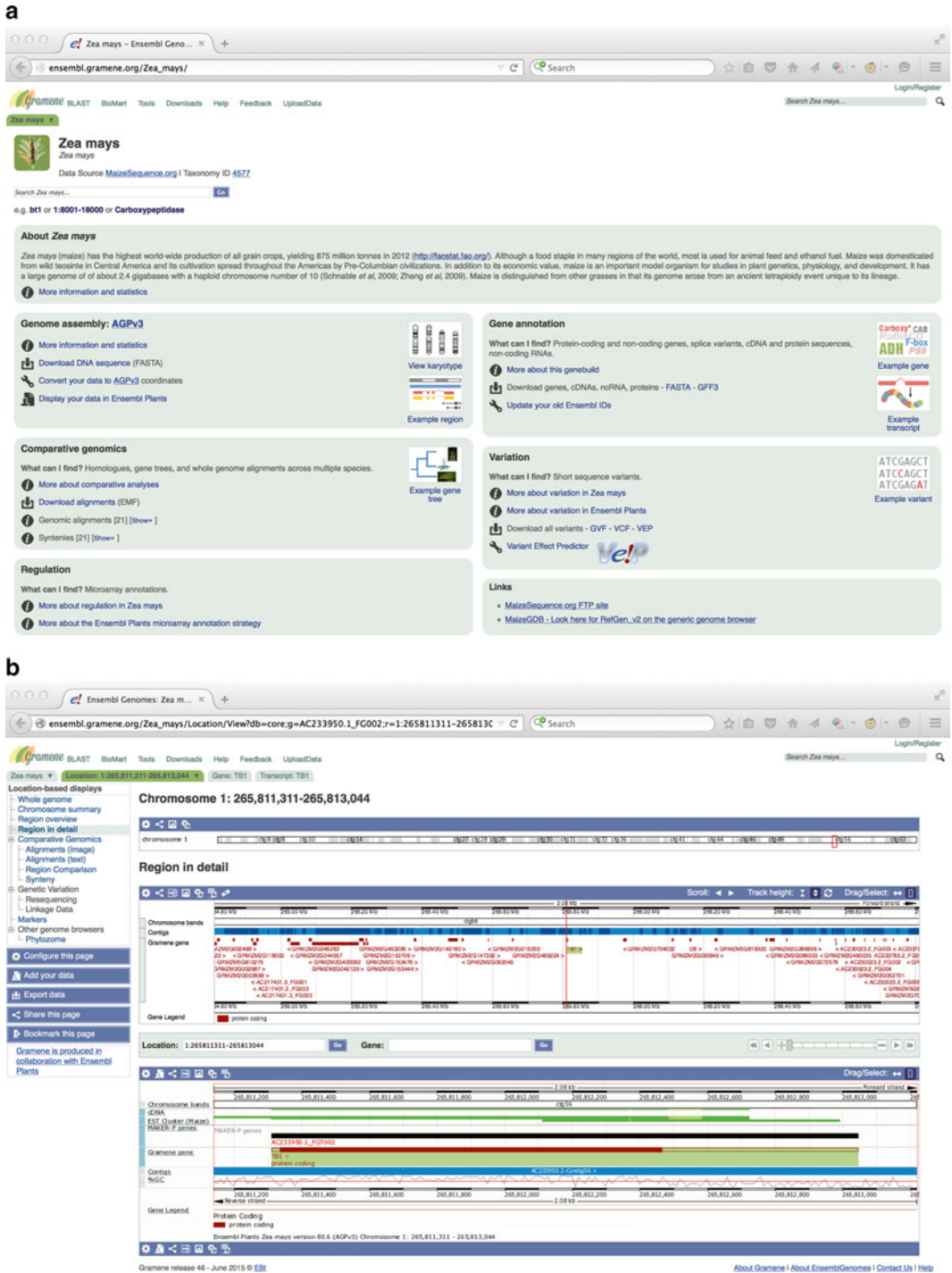
Gramene is powered by multiple modular platforms. The main entry point for the system is through the front web homepage (<http://www.gramene.org>; Fig. 1). Every Gramene page contains the main navigation bar and module-specific navigation bars if applicable, a simple search form that can be refined to interrogate the different data modules, a link to the homepage and another to our Contact page. The main navigation bars are found at the top and left side of each Gramene content page and constitute the main entry point to the search module (Search), genome browser (Genomes), pathway databases (Pathways), bulk data sets (Download), information about the project (About) and its collaborators (Collaborators), outreach events and educational materials (Outreach), and legacy data and resources (Archive). The Contact link is set up to provide the user a feedback page where the URL from the page that the user was viewing at the time of the response is automatically included in the message. The interfaces within Gramene are interactive, providing the user with links to external reference databases as well as links to internal modules within Gramene.

### 3.2 Example Uses of Gramene

Within the constraints of this chapter, it would not be possible to go through all of the Gramene interfaces. Instead, these examples provide sample queries and walk through using Gramene to obtain information and facilitate genomic research. These and additional examples focused on comparative analysis of plant metabolic and regulatory pathways as well as plant gene expression analysis are available from Gramene's Outreach page ([www.gramene.org/outreach](http://www.gramene.org/outreach)). Herein we will only demonstrate one of many possible ways to explore the Gramene website to address a given query and encourage users to discover other ingenious ways to solve them.

*Exercise 1. View a phylogenetic tree for a family of transcription factors*

In this exercise, we will navigate a phylogenetic tree for plant genes in the TCP family of transcription factors centered on the maize gene *tb1* (teosinte branched 1 [56]). We will then generate a list of homologues (i.e., orthologues and paralogues) for this gene, highlight species-specific homologues with particular Gene Ontology (GO) annotations in the tree, and download images and tables with the results. On the Gramene homepage, type “tb1” in the search box on the top of the page. Once the results appear, narrow your target by specifying “*Zea mays*” under Species. The resulting link will conduct you to the maize *tb1* gene page of the Ensembl genome



**Fig. 5** Gramene Ensembl genome browser pages: (a) Species and (b) Location (e.g., maize *tb1* gene)

**a**

Ensembl Genomes: Zea m...  
ensembl.gramene.org/Zea\_mays/Gene/Summary?db=core&g=AC23950.1\_FG002&v=1-26581311-265813

Gene: TBI1 AC23950.1\_FG002

Description: Transcription factor TEGOSANTE BRANCHED 1 [Source:UniProtKB/Swiss-Prot;Acc:Q89W92]

Location: Chromosome 1\_265,811,311-265,813,044 forward strand.

About this gene: This gene has 1 transcript (nucleotide), 87 orthologues and 17 paralogues.

Transcripts: [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
TBI1	AC23950.1_FG002	1754	276aa	Protein coding	A0A06CXX00 Q89W92	

Summary

Name: TBI1 (UniProtKB Gene Name)

UniProtKB: This gene has proteins that correspond to the following UniProtKB identifiers: Q89W92

Gene type: Protein coding

Annotation Method: Gene annotation by Gramene through an automated, evidence-based method

Go to Region in Detail for more tracks and navigation options (e.g. zooming)

Genome area: [Genomic track showing exons and introns]

Genes: [TBI1](#)

Transcripts: [TBI1](#)

Variation Legend: [Protein variant](#) [Synonymous variant](#) [3 prime UTR variant](#) [Upstream gene variant](#) [Downstream gene variant](#) [Intergenic variant](#)

Configuring the display: Tip: use the "Configure this page" link on the left to show additional data in this region.

Gramene release 48 - June 2019 © EBI

**b**

Ensembl Genomes: Zea m...  
ensembl.gramene.org/Zea\_mays/Transcript/SupportingEvidence?db=core&g=AC23950.1\_FG002&v=1-26581311-265813

Transcript: TBI1 AC23950.1\_FG002

Description: Transcription factor TEGOSANTE BRANCHED 1 [Source:UniProtKB/Swiss-Prot;Acc:Q89W92]

Location: Chromosome 1\_265,811,311-265,813,044 forward strand.

About this transcript: This transcript has 1 exon, is annotated with 15 domains and features, is associated with 19 variations and maps to 1 oligo probe.

Gene: This transcript is a product of gene AC23950.1\_FG002 [Hide transcript table](#)

Supporting evidence

Genome area: [Detailed supporting evidence track showing various evidence types]

Transcript support:

- EST evidence
- Protein evidence
- RNA-seq evidence
- Other evidence

Legend: [protein evidence](#) [EST evidence](#)

Annotations: [non-coding antisense](#), [evidence start / ends within exon / CDS](#), [evidence extends beyond exon / CDS](#), [part of evidence distributed in transcript structure](#), [part of evidence missing from transcript structure](#), [evidence extends beyond the end of the transcript](#)

**Fig. 6** Gene-centric Gramene Ensembl genome browser pages: (a) Gene, (b) Transcript, and (c) Variation pages (e.g., maize *tbi1* gene; SNP variant PZE01264848659)

**c**

Ensembl Genomes: Zea mays x

ensembl.gramene.org/Zea\_mays/Variation/Explore?db=core;g=AC233950.1\_FG002;r=1:265811311-265813044;t=AC233950.1\_FG...

Gramene BLAST BioMart Tools Downloads Help Feedback UploadData Search Zea mays...

Zea mays Location: 1:265,811,311-265,813,044 Gene: TB1 Transcript: TB1 Variation: PZE01264848659

Variation displays

- Explore this variation
- Genomic context
  - Genes and regulation
  - Flanking sequence
- Genotype frequency
- Individual genotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Gramene is produced in collaboration with Ensembl Plants

**PZE01264848659 SNP**

Original source

Alleles **G/A** | Ambiguity code: R

Location Chromosome 1:265811476 (forward strand) | [View in location tab](#)

Most severe consequence **Missense variant** | [See all predicted consequences \(Genes and regulation\)](#)

HGVS names **+** This variation has 3 HGVS names - click the plus to show

About this variant This variant overlaps 1 [transcript](#).

**Explore this variation**

Genomic context

Genes and regulation

Genotype frequency

Individual genotypes

Linkage disequilibrium

Phenotype data

Citations

Phylogenetic context

Flanking sequence

**Fig. 6** (continued)

browser. Note the four distinct tabs on the top of this page: Species (Fig. 5a), Location (Fig. 5b), Gene (Fig. 6a), and Transcript (Fig. 6b); an additional Variation page (Fig. 6c) may be accessed for species with variation data in Gramene (*see* Table 2). Each of these tab views will be discussed in more detail below. Common to the Location, Gene, Transcript, and Variation pages, as well as the views available therein, are customizable tracks, links to internal pages, and contextual links to data sources outside of Gramene. Actions enabled for each of those pages and their embedded views include (1) configuring and resizing, (2) uploading and managing user-provided data for graphic display, (3) exporting or downloading data, and (4) sharing pages and images. For example, you may customize the tracks on display by selecting the “Configure this page” instruction on the left side navigation bar or upon clicking on the “Configure this image” icon on the top left corner of an image. A new browser window will pop up listing all available data tracks for the browser view that you wish to customize. Data tracks are grouped by category; click on a category to see the complete list of available tracks for that category (e.g., “mRNA and protein alignments” may include tracks for EST clusters, cDNAs, and protein features from various species, sources, or methods). A track gets activated for display on the browser by clicking on the square preceding its name



and selecting a desired “track style”. Favorite tracks may be set and the order of tracks may be changed. Save your selections and close the pop-up window by clicking on the check mark on the top right corner. The browser will automatically refresh itself and your selected tracks should now be visible.

Gramene Ensembl Genome Browser pages:

1. The Species page (*Zea mays* for this example; Fig. 5a) contains detailed information about the reference genome assembly and gene annotation; comparative genomics data including phylogenetic gene trees, whole-genome alignments, and synteny views; gene regulation (microarray) data; genetic and structural variation; and links to download data sets in bulk.
2. The Location page (Chr 1: 265,811,311–265,813,044 in the B73 maize AGPv3 assembly; Fig. 5b) offers several scalable views on the left side navigation bar, e.g., karyotype or whole-genome view, chromosome summary, region overview, region in detail (expanded red box from the region overview), as well as comparative genomics views, which include multi-species alignments, region comparisons, and synteny views. Semantic zooming is available for each “region” view.
3. The Gene page (TBI; Fig. 6a) provides a summary of data available for a given gene, as well as an extensive list of features including splice variants (*see* also Transcript page), exon/intron marked-up sequence, associated ontology terms and literature references, external references, comparative genomic alignments, expandable gene trees, orthologues and paralogues, and genetic/structural variation. The Plant Compara Gene Trees are derived from a pre-computed phylogenetic analysis of protein-coding genes from all Gramene species, plus several representative animal genomes used as outgroups. The Pan-taxonomic Compara Gene Trees sample species more broadly across taxa represented by the Ensembl Genomes project, including bacteria, fungi, protists, and metazoa, and include only a subset of representative plant species held within the Gramene database.
4. The Transcript page (TBI-201; Fig. 6b) includes sequence data, external cross-references (including oligo probe sources), supporting protein/EST evidence, GO associations, variation, and protein domains and features. If variation data is available for a given gene, each variant will have its own Variation page (*see* below). The Variations table under the Protein Information category provides a complete list of the transcript’s variants with alleles, functional consequence, relative position in the protein’s amino acid sequence, ambiguity code, and actual affected codons/amino acids, if any. Moreover, for species like *Arabidopsis* in which the same set of variants have been genotyped in different populations, tabular and graphic “population comparisons” are available from the Transcript page.



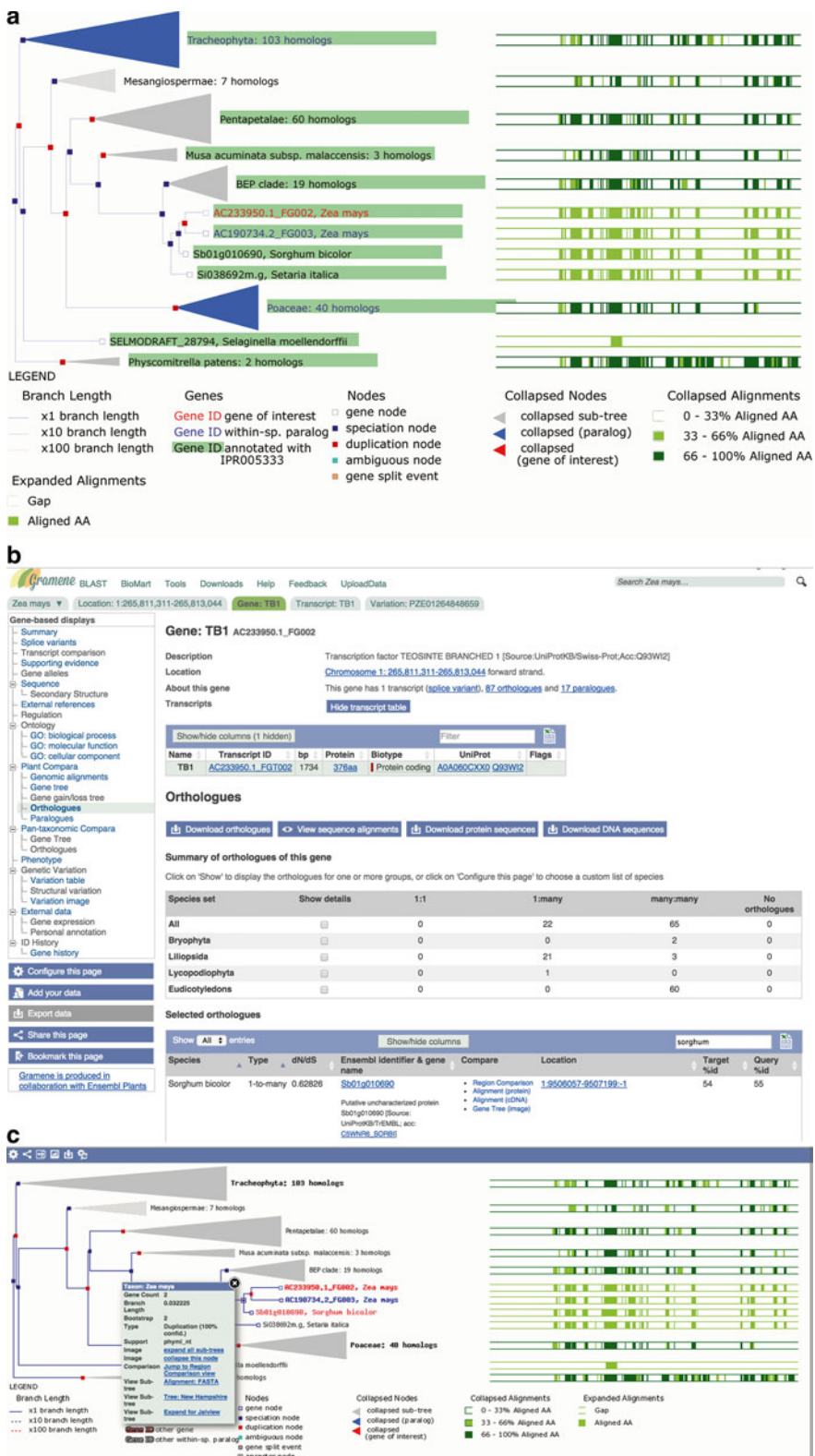
5. The Variation page (e.g., PZE01264848659 for *tb1*; Fig. 6c) includes the variant's genomic context, functional consequences in all transcripts, individual genotype data, as well as allele/genotype frequency by population tables. Note that if several transcripts are available for a given gene, the same variant may have different functional consequences in each transcript as per its relative location in the corresponding protein product.

To view the phylogenetic tree for the TCP family of transcription factors, go to the *tb1* Gene page and click on the “Gene tree (image)” view. In the “Highlight annotations” table, both InterPro and GO terms are enabled by default; uncheck the box for GO terms to show only InterPro domains. From the list, select IPR005333, which is the InterPro ID for the complete TCP domain (*see Note 2* for visualizing the complete protein domain structure of the maize *tb1* gene). Figure 7a displays the collapsed view of the tree with all the clades highlighted. Click on “View fully expanded tree” from the “View options” at the bottom of the page. Except for a handful of genes, all the genes in the tree image will light up because of the prevalence of the TCP domain. We may also highlight orthologues and paralogues between two species. For example, let's find the sorghum orthologue with highest similarity to maize *tb1*. From the maize gene's page, select “Plant Comparison Orthologues” and enter “sorghum” in the “Filter” box on the top right corner of the orthologues table (Fig. 7b). In the “Compare” column, click on the “Gene tree (image)” link for the sorghum orthologous gene (Sb01g010690), and upon full expansion of the tree, you will see TB1 and SB01G010690 highlighted in different shades of red, maize within-species paralogues in different shades of blue, and sorghum paralogues highlighted in black (Fig. 7c shows the collapsed view of the highlighted tree). Note that by clicking on any speciation tree node, a pop-up inset will appear with various parameters describing the tree, as well as options to selectively collapse nodes and view a sub-tree in other formats like FASTA.

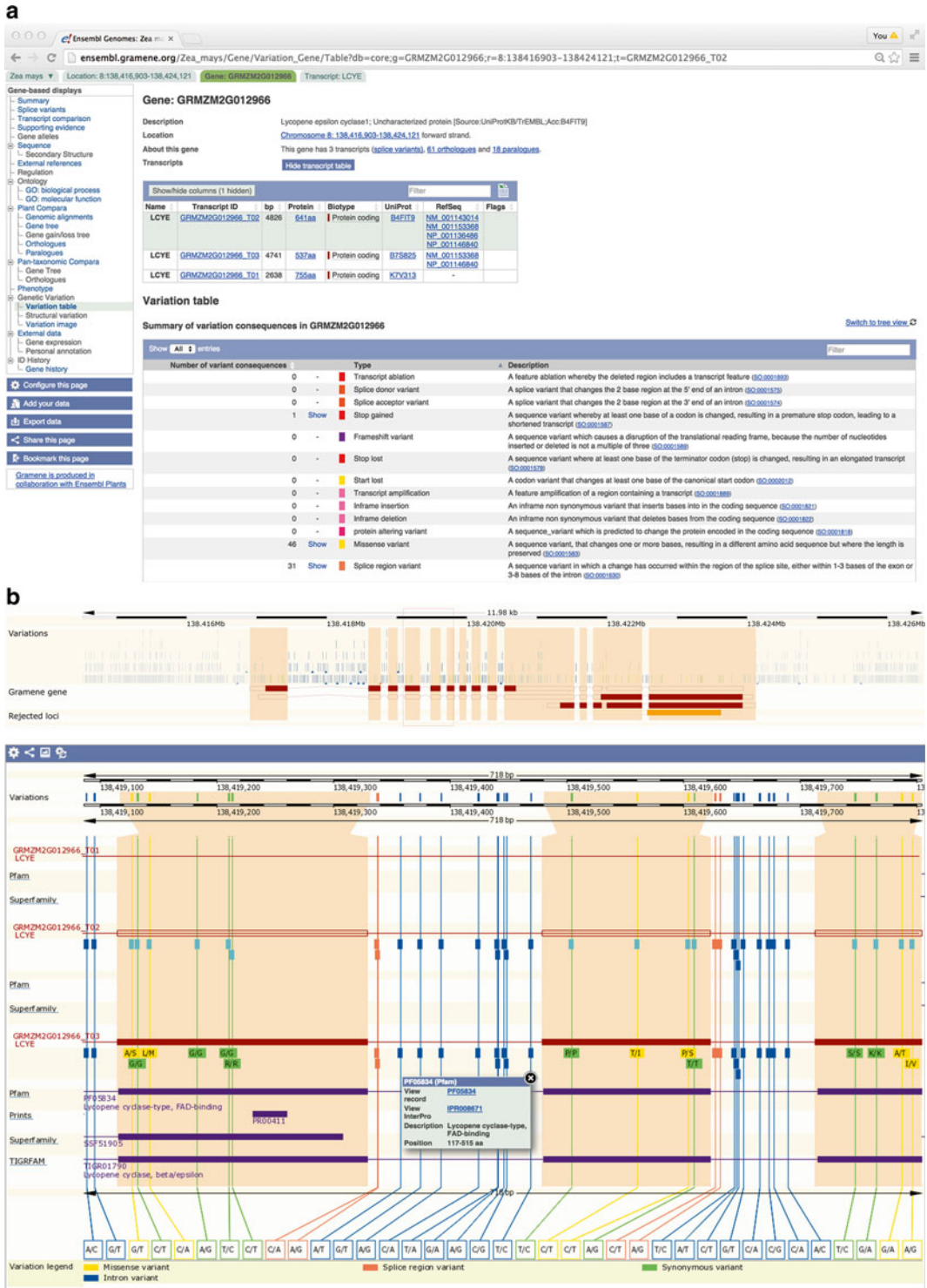
*Exercise 2. Explore genetic variation in the rice orthologues of a maize gene with a known trait association*

We will now explore genetic variation in the rice orthologues of the maize lycopene epsilon gene (*lcyE*). Specifically, we will determine whether the non-synonymous substitution mapping to nucleotide 210 relative to the start codon of the transcript with the longest genomic span (LCYE-201 or GRMZM2G012966\_T03), which was found to be associated with provitamin A accumulation in the maize kernel [57], is also present in its rice orthologues.

Go to the maize *lcyE* gene page as done for the *tb1* gene. From the gene and transcript pages, you may visualize all its genetic variants in tabular (“Variation table” option in the left side navigation bar; Fig. 8a) or graphic form in their genomic context



**Fig. 7** Exercise 1: visualization of a phylogenetic tree for the TCP family of transcription factors centered on maize TB1. **(a)** Collapsed view of the tree highlighting all gene products that include the TCP InterPro domain (IPR005333). **(b)** Filtered view of sorghum orthologues of TB1. **(c)** Gene tree image for TB1 highlighting its sorghum orthologues (e.g., SB01G010690) and within-species paralogues (maize and sorghum, respectively); speciation nodes shown in *black*, duplication nodes shown in *red*. Also shown is *inset* that pops up upon clicking on any node



**Fig. 8** Exercise 2: exploring genetic variation in a rice gene while looking for conservation of a maize SNP variant (PZE08137569063) associated with provitamin A accumulation in the kernel [57]. (a) Genetic variation in the maize *lycE* gene in tabular form, and (b) graphic form

(“Variation image” option; Fig. 8b). The Variation table groups variants by functional consequence; by clicking on the “Show” option for a given category (e.g., missense variant), you will get a list of the variants with other data like genomic position, alleles, relative amino acid position (if affected), etc. The Variation image displays the same information as the table in graphic form, plus the relative location of known protein domains. However, if you know the SNP identifier, the simplest way to find all the available information for a given variant is to select the “Variations” option under “Protein Information” as it lists all variants by identifier. The SNP variant associated with provitamin A accumulation identified by Harjes *et al* [57] is PZE08137569063. As shown in the “Genotype frequency” (Fig. 9a) view available from the Variation page, this variant has alleles G and T with variable genotype frequencies in 13 maize or teosinte populations, including HapMap2 (“Zmays”).

Now, let’s identify the closest rice orthologues of the *lcyE* gene by proceeding as described above in the “Gene tree (image)”. The “Orthologues” view allows users to download all or a selected set of orthologous genes (by using the filter box on the top right corner of the table), as well as to view and download the corresponding protein sequences and/or pairwise alignments. To download nucleotide sequences or download all the genetic variants for the orthologues, users could go to each individual gene’s page and proceed as described above (i.e., go to the Variation table/image and download the data directly from the table/image or click on “Export data” option on the left sidebar menu). Alternatively, users may download for each species the same DNA/protein sequence and variation data using, respectively, the “Plant Genes” and “Plant Variation” databases in the GrameneMart utility (<http://ensembl.gramene.org/biomart/mart-view/>; Fig. 9b). Users may visually compare all the species in the gene tree by selecting “Gene tree (alignment)” in the left sidebar menu or view pairwise genomic alignments with the “Genomic alignments (text)” option. Alternatively, users may use a multiple alignment program such as ClustalW to visually compare the rice orthologous gene sequences, and realize that (1) this site has not been found to be polymorphic in *O. sativa Japonica* and *Indica*, (2) there is a sequence gap around this position in *O. glaberrima*, and (3) the ancestral G allele is the one present at this position in the *O. nivara*, *O. glumaepatula*, and *O. punctata* orthologous genes.

Further genomic analysis may be performed with the Ensembl “Tools” available at <http://ensembl.gramene.org/tools.html> and other links from Gramene’s archival Diversity pages at <http://archive.gramene.org/diversity/tools.html>.

*Exercise 3. Upload, visualize, and share your own data into a new genome browser track*

The Ensembl genome browser allows users to upload their own data and visualize it on a custom track. Data may be formatted in various file formats including GFF, GTF, BED, BAM, VCF,



a

Zea mays Location: 8:138,416,835-138,417,835 Variation: PZE08137569063

**Variation displays**

- Explore this variation
- Genomic context
  - Genes and regulation
  - Flanking sequence
- Genotype frequency**
- Individual genotypes
- Linkage disequilibrium
- Phenotype Data
- Phylogenetic Context
- Citations
- External Data

**PZE08137569063 SNP**

Original source Alleles T/G | Ambiguity code: K

Location Chromosome 8:138417335 (forward strand) | [View in location tab](#)

Most severe consequence Missense variant | [See all predicted consequences \(Genes and regulation\)](#)

Synonyms Panzea\_2.7GBS S8\_138883026

HGVS names This variation has 4 HGVS names - click the plus to show

About this variant This variant overlaps 3 transcripts.

**Genotype frequency**

Frequency data (13)

Population	Allele: frequency (count)		Genotype: frequency (count)		
BREAD	T: 0.041 (450)	G: 0.959 (10642)	TTT: 0.019 (105)	GIG: 0.938 (5201)	GIT: 0.043 (240)
IBM	T: 0.351 (253)	G: 0.649 (467)	TTT: 0.339 (122)	GIG: 0.636 (229)	GIT: 0.025 (9)
IBM_parent	T: 0.351 (253)	G: 0.649 (467)	TTT: 0.339 (122)	GIG: 0.636 (229)	GIT: 0.025 (9)
Margaret_Smith_lines	T: 0.594 (253)	G: 0.406 (173)	TTT: 0.563 (120)	GIG: 0.376 (80)	GIT: 0.061 (13)
NAM	T: 0.706 (8104)	G: 0.294 (3380)	TTT: 0.690 (3962)	GIG: 0.279 (1600)	GIT: 0.031 (180)
NAM_F1	T: 0.630 (29)	G: 0.370 (17)	TTT: 0.304 (7)	GIG: 0.043 (1)	GIT: 0.652 (15)
NAM_common_parent	T: 1.000 (110)	G: 0.000 (0)	TTT: 1.000 (55)		
NAM_parent	T: 0.441 (82)	G: 0.559 (104)	TTT: 0.441 (41)	GIG: 0.559 (52)	
Spanish_inbred_lines	T: 0.421 (69)	G: 0.579 (95)	TTT: 0.341 (28)	GIG: 0.500 (41)	GIT: 0.159 (13)
W22_x_Teosinte_BC2S3	T: 1.000 (182)	G: 0.000 (0)	TTT: 1.000 (91)		
Zmays	T: 0.262 (44)	G: 0.738 (124)	TTT: 0.250 (21)	GIG: 0.726 (61)	GIT: 0.024 (2)
fine_mapping	T: 1.000 (2)	G: 0.000 (0)	TTT: 1.000 (1)		
teosinte_inbred_lines	T: 0.406 (13)	G: 0.594 (19)	TTT: 0.312 (5)	GIG: 0.500 (8)	GIT: 0.188 (3)

b

Gramene BLAST BioMart Tools Feedback

Search all species...

New Count Results URL XML Perl Help

**Dataset 43 / 5512746 SNPs**  
Oryza sativa Japonica (Rice) variations (IRGSP-1.0 (IRGSP-1.0))

**Filters**

Ensembl Gene ID(s) [Max 500 advised] : [ID-list specified]

**Attributes**

Variation Name  
Chromosome name  
Position on Chromosome (bp)  
Strand  
Variant Allele  
Population Name  
Population Size  
Population Genotype  
Genotype Frequency  
Ensembl Gene ID  
Ensembl Transcript ID  
Consequence to transcript

**Dataset**  
[None Selected]

**Please restrict your query using criteria below**

REGION:

GENERAL VARIATION FILTERS:

GENE ASSOCIATED VARIATION FILTERS:

Ensembl Gene ID(s) [Max 500 advised] OS01G0581400

No file selected.

Consequence type

splice\_acceptor\_variant  
splice\_donor\_variant  
stop\_lost  
coding\_sequence\_variant  
missense\_variant

**Fig. 9 (a)** Genotype frequency data for PZE08137569063 in 13 maize or teosinte populations from HapMap2 and the Panzea 2.7 GBS data release Gramene. **(b)** Genotype data for Gene sequences, orthologous/paralogous gene lists, and gene variants available for customized download via the GrameneMart

bedGraph, gbrowse, PSL, WIG, BigBed, BigWig, and TrackHub. Some data like GFF annotations may be directly uploaded from a local machine. Large data files like BED/BAM alignments or BigWig graphic display configurations need to be uploaded onto a local server that is accessible to the browser via a URL. Another way to share third-party data is via a DAS (Distributed Annotation System) registry, which would need to be set up by a software engineer.

Test data sets consisting of BAM alignments and CpG methylation for B73 and Mo17 maize lines used in the study by Regulski et al. [58] are available from the Gramene outreach pages to upload and visualize for this exercise. To upload the data simply click on the “Add/Manage your data” option on the left bar menu of any genome browser page (Fig. 10a). This action will take you to the upload page (Fig. 10b) where you need to specify the format of the file you intend to upload (formats and test sets are also described in the “Help on supported formats, display types, etc.” link therein). To visualize custom data in a new browser track, make sure that your track is turned on in the configuration menu and you are looking at a region that includes the new data you have just uploaded. The BAM alignments and CpG methylation ratios are shown in Fig. 10c.

---

## 4 Notes

1. Apache 2.x is not supported yet due to significant differences in the persistent Perl interpreter module (`mod_perl`).
2. For the complete protein domain structure of a gene, go to the Transcript page and select “Domains & features”. By clicking on a particular “Display all genes with this domain” link, you will get a list of all genes in the same species that contain any given InterPro domain. To download the list, click on the “Download” icon to the right of the Filter box.

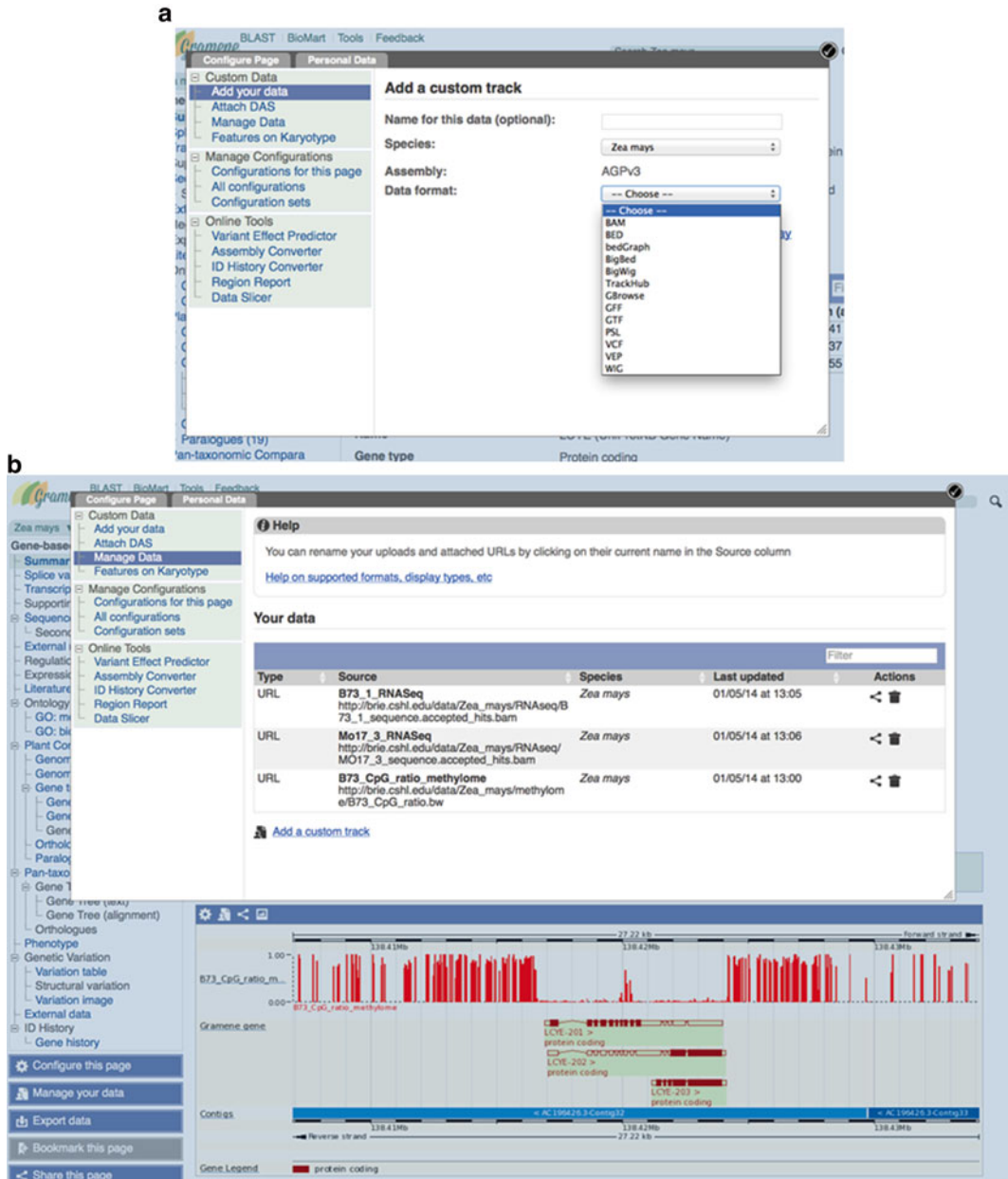
---

## Acknowledgements

The authors would like to thank all members of the Gramene Project, especially Bo Wang for going through the exercises and providing feedback for clarity in the protocols and Peter van Buren for system technology support. We are also grateful to Gramene’s users for valuable suggestions, and our collaborators for sharing genomic-scale data sets that make Gramene an outstanding community resource. The Genomes and Pathways modules in Gramene would not have been possible without the synergistic collaborations with the Ensembl Genomes project at the EMBL-European Bioinformatics Institute, and the Reactome project at the Ontario Institute for Cancer Research, respectively.

Gramene is supported by an NSF grant (IOS-1127112).





**Fig. 10** Uploading and managing user-provided data to display as a new genome browser track. **(a)** Pop-up window upon clicking on the “Add/Upload Data” link from the Gene page. User selects species and file format for the data to be uploaded. **(b)** Preloaded BAM alignments and CpG methylation ratios. New track shows CpG methylation data in the selected gene region

## References

1. Jia J, Zhao S, Kong X et al (2013) Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 496(7443):91–95
2. Amborella Genome Project (2013) The Amborella genome and the evolution of flowering plants. *Science* 342(6165): 1241089
3. Chamala S, Chanderbali AS, Der JP et al (2013) Assembly and validation of the genome

- of the nonmodel basal angiosperm *Amborella*. *Science* 342(6165):1516–1517
4. Hu TT, Pattyn P, Bakker EG et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43(5):476–481
  5. The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796–815
  6. The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282):763–768
  7. Wang X, Wang H, Wang J et al (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 43(10):1035–1039
  8. Merchant SS, Prochnik SE, Vallon O et al (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318(5848):245–250
  9. Matsuzaki M, Misumi O, Shin IT et al (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428(6983):653–657
  10. Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278):178–183
  11. International Barley Genome Sequencing Consortium, Mayer KF, Waugh R et al (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491(7426):711–716
  12. Young ND, Debelle F, Oldroyd GE et al (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480(7378):520–524
  13. D’Hont A, Denoeud F, Aury JM et al (2012) The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 488(7410):213–217
  14. Chen J, Huang Q, Gao D et al (2013) Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun* 4:1595
  15. Yu J, Hu S, Wang J et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296(5565):79–92
  16. Zhao W, Wang J, He X et al (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res* 32(Database issue):D377–D382
  17. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436(7052):793–800
  18. Kawahara Y, de la Bastide M, Hamilton JP et al (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* 6(1):4
  19. Rensing SA, Lang D, Zimmer AD et al (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319(5859):64–69
  20. Tuskan GA, Difazio S, Jansson S et al (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793):1596–1604
  21. The International Peach Genome Initiative, Verde I, Abbott AG 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(5):487–494
  22. Banks JA, Nishiyama T, Hasebe M et al (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332(6032):960–963
  23. Bennetzen JL, Schmutz J, Wang H et al (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol* 30(6):555–561
  24. Zhang G, Liu X, Quan Z et al (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30(6):549–554
  25. Tomato Genome C (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400):635–641
  26. Potato Genome Sequencing Consortium, Xu X, Pan S et al (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355):189–195
  27. Paterson AH, Bowers JE, Bruggmann R et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457(7229):551–556
  28. Brenchley R, Spannagl M, Pfeifer M et al (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* 491(7426):705–710
  29. Ling HQ, Zhao S, Liu D et al (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* 496(7443):87–90
  30. Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449(7161):463–467
  31. Myles S, Chia JM, Hurwitz B et al (2010) Rapid genomic characterization of the genus *vitis*. *PLoS One* 5(1), e8219
  32. Atwell S, Huang YS, Vilhjalmsón BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465(7298):627–631

33. Fox SE, Preece J, Kimbrel JA et al (2013) Sequencing and de novo transcriptome assembly of *Brachypodium sylvaticum* (Poaceae). *Appl Plant Sci* 1(3):1200011
34. McNally KL, Childs KL, Bohnert R et al (2009) Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc Natl Acad Sci U S A* 106(30):12273–12278
35. Zhao K, Wright M, Kimball J et al (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS One* 5(5): e10780
36. Yu J, Wang J, Lin W et al (2005) The genomes of *Oryza sativa*: a history of duplications. *PLoS Biol* 3(2), e38
37. Morris GP, Ramu P, Deshpande SP et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci U S A* 110(2):453–458
38. Zheng LY, Guo XS, He B et al (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biol* 12(11):R114
39. Gore MA, Chia JM, Elshire RJ et al (2009) A first-generation haplotype map of maize. *Science* 326(5956):1115–1117
40. Chia JM, Song C, Bradbury PJ et al (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* 44(7): 803–807
41. Flicek P, Amode MR, Barrell D et al (2014) Ensembl 2014. *Nucleic Acids Res* 42(Database issue):D749–D755
42. Kersey PJ, Allen JE, Christensen M et al (2014) Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Res* 42(Database issue):D546–D552
43. Ware D (2007) Gramene. *Methods Mol Biol* 406:315–329
44. Dharmawardhana P, Ren L, Amarasinghe V et al (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(1):15
45. Monaco MK, Sen TZ, Dharmawardhana PD et al (2013) Maize metabolic network construction and transcriptome analysis. *Plant Genome* 6(1):1–12
46. Youens-Clark K, Buckler E, Casstevens T et al (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Res* 39(Database issue):D1085–D1094
47. Karp PD, Paley SM, Krummenacker M et al (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 11(1):40–79
48. Caspi R, Altman T, Billington R et al (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 42(Database issue):D459–D471
49. Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol* 132(2):453–460
50. Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics* 23(11): 1418–1423
51. Zhang P, Dreher K, Karthikeyan A et al (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol* 153(4):1479–1491
52. Bombarely A, Menda N, Teclé IY et al (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res* 39(Database issue):D1149–D1155
53. Croft D, O’Kelly G, Wu G et al (2011) Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 39(Database issue):D691–D697
54. Spooner W, Youens-Clark K, Staines D et al (2012) GrameneMart: the BioMart data portal for the Gramene project. *Database (Oxford)* 2012:bar056
55. Monaco MK, Stein J, Naithani S et al (2014) Gramene 2013: comparative plant genomics resources. *Nucleic Acids Res* 42(Database issue):D1193–D1199
56. Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386(6624):485–488
57. Harjes CE, Rocheford TR, Bai L et al (2008) Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* 319(5861):330–333
58. Regulski M, Lu Z, Kendall J et al (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res* 23(10):1651–1662