

# Maize Metabolic Network Construction and Transcriptome Analysis

Marcela K. Monaco, Taner Z. Sen, Palitha D. Dharmawardhana, Liya Ren, Mary Schaeffer, Sushma Naithani, Vindhya Amarasinghe, Jim Thomason, Lisa Harper, Jack Gardiner, Ethalinda K.S. Cannon, Carolyn J. Lawrence, Doreen Ware, and Pankaj Jaiswal\*

## Abstract

A framework for understanding the synthesis and catalysis of metabolites and other biochemicals by proteins is crucial for unraveling the physiology of cells. To create such a framework for *Zea mays* L. subsp. *mays* (maize), we developed MaizeCyc, a metabolic network of enzyme catalysts, proteins, carbohydrates, lipids, amino acids, secondary plant products, and other metabolites by annotating the genes identified in the maize reference genome sequenced from the B73 variety. MaizeCyc version 2.0.2 is a collection of 391 maize pathways involving 8889 enzyme mapped to 2110 reactions and 1468 metabolites. We used MaizeCyc to describe the development and function of maize organs including leaf, root, anther, embryo, and endosperm by exploring the recently published microarray-based maize gene expression atlas. We found that 1062 differentially expressed metabolic genes mapped to 524 unique enzymatic reactions associated with 310 pathways. The MaizeCyc pathway database was created by running a library of evidences collected from the maize genome annotation, gene-based phylogeny trees, and comparison to known genes and pathways from rice (*Oryza sativa* L.) and *Arabidopsis thaliana* (L.) Heynh. against the Pathologic module of Pathway Tools. The network and the database that were also developed as a community resource are freely accessible online at <http://maizecyc.maizegdb.org> to facilitate analysis and promote studies on metabolic genes in maize.

**M**AIZE (*Zea mays* subsp. *mays*) is one of the most agriculturally and economically important crops worldwide (FAOSTAT, 2011). Its widespread use is not limited to food and feedstock, but various commodities such as paint, plastics, soap, tiles, and packaging material are also made from maize. More recently it has been recognized as an excellent source of lignocellulosic biofuel. In addition, maize is one of the founding model organisms for genetics research and, along with rice and *Arabidopsis thaliana*, currently is one of the leading models for plant functional genomics (Gaut et al., 2000; Rabinowicz and Bennetzen, 2006; Strable and Scanlon, 2009). The discovery of molecular interactions that lead to desirable traits in this crop is ongoing. The availability of the draft reference genome sequence for the maize inbred cultivar B73 (Schnable et al., 2009) opened avenues to explore the interaction of genes,

M.K. Monaco, L. Ren, J. Thomason, and D. Ware, Cold Spring Harbor Lab., 1 Bungtown Rd., Cold Spring Harbor, NY 11724; T.Z. Sen, M. Schaeffer, L. Harper, C.J. Lawrence, and D. Ware, USDA-ARS, Jamie L. Whitten Building, Room 302A, 1400 Independence Ave., S.W., Washington, DC 20250; T.Z. Sen, J. Gardiner, E.K.S. Cannon, and C.J. Lawrence, Dep. of Genetics, Development and Cell Biology, Iowa State Univ., Ames, IA 50011; P.D. Dharmawardhana, V. Amarasinghe, and P. Jaiswal, Dep. of Botany and Plant Pathology, 2082 Cordley Hall, Oregon State Univ., Corvallis, OR 97331; M. Schaeffer, Division of Plant Sciences, Dep. of Agronomy, Univ. of Missouri, Columbia, MO 65211; S. Naithani, Dep. of Horticulture, 4017 ALS Bldg., Oregon State Univ., Corvallis, OR 97331; L. Harper, Dep. of Molecular and Cell Biology, Univ. of California, Berkeley, CA 94720; J. Gardiner, School of Plant Sciences, Univ. of Arizona, Tucson, AZ 85721. The gene expression data analysis was performed on the published and publicly available NCBI GEO record GSE27004. Received 20 Sept. 2012. \*Corresponding author (jaiswalp@science.oregonstate.edu).

Published in The Plant Genome 6.  
doi: 10.3835/plantgenome2012.09.0025  
© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

**Abbreviations:** 16DAP, 16 d after pollination; CoA, coenzyme A; EC, Enzyme Commission; FGS, filtered gene set; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Maize Core DB, MaizeSequence.org's Ensembl Core Database; MaizeGDB, Maize Genetics and Genomics Database; ORF, open reading frame; PO, plant ontology.

gene products, and metabolites that regulate the development of cellular components, cells, tissues, organs, and physiological manifestations of the biochemical networks in response to various extrinsic and intrinsic signals. Understanding maize metabolism at a systems level requires a multifaceted approach to analyze gene functions with respect to subcellular localization and sites of mechanistic function of its products, namely transcripts and proteins to discover their overall role in biological processes. Notably, levels of gene expression change in response to growth, development, and various biotic and abiotic signals from the environment where a plant grows. Similarly, the localization of gene products (immature and mature forms) can be intra- or extracellular, be confined to one or many tissues and/or distributed throughout an organ, and vary across growth and developmental stages. Although our representation of the maize metabolic network mainly focuses on catalytic events performed by a small number of genes encoding enzymes and transporters that are responsible for phenotype and function, it is important to bear in mind that a proportionally large number of gene products are involved in physical interactions with DNA, ribonucleic acid (RNA), proteins, and metabolites to carry out regulatory, signaling, and transport functions. As more genomes are sequenced for an increasing number of species, complementary metabolomics- and proteomics-based genome-scale metabolic reconstructions must be developed to discover spatial, temporal, and organism-specific differences. Several metabolic network reconstructions are currently available (Stelling et al., 2002; Papp et al., 2004; Vastrik et al., 2007; Tsesmetzis et al., 2008; Ferrer, 2009; Latendresse et al., 2011). These include the Plant Metabolic Network (PlantCyc) (Zhang et al., 2010) and species-specific metabolic networks for both dicots (Mueller et al., 2003, 2005; Zhang et al., 2005; Urbanczyk-Wochniak and Sumner, 2007) and monocots (Jaiswal et al., 2006; <http://pathway.gamene.org/gramene/ricecyc.shtml> [accessed 15 July 2011]). Much of the plant pathway information is accessible in reference libraries provided by the Encyclopedia of Metabolic Pathways (MetaCyc) (Caspi et al., 2012), PlantCyc (Zhang et al., 2010), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2011). The field of study advances rapidly; new reports describing genome-scale metabolic reconstructions focused on flux balance studies for *Arabidopsis thaliana* (L.) Heynh. (Poolman et al., 2009) and *Zea mays* (Dal'Molin et al., 2010) are some examples. Both models rely heavily on the curated resources *Arabidopsis* Metabolic Network (AraCyc) and KEGG with no additional manual inputs. The C4GEM model (Dal'Molin et al., 2010) was used successfully to highlight significant metabolic differences between photosynthetic and nonphotosynthetic reactions and focused on plants such as maize, sugarcane (*Saccharum officinarum* L.), and sorghum [*Sorghum bicolor* (L.) Moench] and assessed flux distributions in leaf mesophyll and bundle sheath cells during C<sub>4</sub> photosynthesis (Hatch, 1978; Rawsthorne, 1992; Edwards et al., 2001; Sage and Sage, 2009). Recently, another genome-scale computational model of a maize metabolic network that adds many reactions to

Dal'Molin et al.'s C4GEM model was reconstructed with about 1500 genes mapped to 1985 reactions and containing 1825 metabolites (Saha et al., 2011). However, evidence from the literature was found for only 42% of the 1985 reactions. This model was used to perform flux balance analysis for three physiological states (photosynthesis, photorespiration, and respiration) and compared predictions against experimental observations for two naturally occurring maize mutants, *bm1* (*brown midrib1*) (Vignols et al., 1995; Vermerris et al., 2002) and *bm3* (*brown midrib3*) (Vignols et al., 1995) with defined defects in cell wall lignin biosynthesis. Because these reports on new models are available only in published article format and/or in supplementary data files, they are difficult to interact with and necessarily lack visualization mechanisms at the systems level. Thus these models are somewhat inaccessible to the plant research community at large. Other recent network reconstruction examples include the one by Grafahrend-Belau et al. (2009) who collected pathway annotations from KEGG (Kanehisa et al., 2011) and investigated primary metabolism in barley (*Hordeum vulgare* L.) seeds associated with grain yield and metabolic fluxes under oxygen stress. A similar network study in rapeseed (*Brassica napus* L.) (Pilalis et al., 2011) focused on seed development and metabolism based on annotations from AraCyc.

The Gramene (Youens-Clark et al., 2011) and Maize Genetics and Genomics Database (MaizeGDB) (Schaeffer et al., 2011a) teams collaborated to assemble an integrated data resource called MaizeCyc. MaizeCyc is an online metabolic pathways network database that enables researchers to visualize and study maize metabolism. Conceptually, MaizeCyc is a representation of the interaction between metabolites as input and output molecules of (i) biochemical reactions in which enzymes serve as catalysts and (ii) transport reactions in which metabolites are transported from one compartment to another within the cell or between the cells via transporter protein carriers and channels. We created MaizeCyc version 2.0.2 based on the protein coding genes identified in the B73 maize variety reference genome sequence assembly named B73 RefGen\_v2 (Schnable et al., 2009). This was done by using an integrated approach that involved electronic and systematic annotation of metabolic pathways, mapping genes to enzymes with known function, data mining from published literature, and manual curation to assign function to genes and enzymes where functional annotation was lacking. MaizeCyc enables researchers, for example, to study maize metabolism and its interaction with the environment by exploring pathway enrichment from gene expression experiments under various biotic and abiotic treatments and growth conditions within the context of plant development. To illustrate this example, we report the analysis of microarray-based genomewide expression data from a maize atlas (Sekhon et al., 2011) to create exemplar visualizations of developmentally regulated global expression of maize genes.

## MATERIALS AND METHODS

### MaizeCyc Development

#### Genome and Protein Sequence Data

The MaizeCyc database was built for the B73 reference genome version 2 (Schnable et al., 2009) using the high-quality filtered gene set (FGS) representing 39,656 gene models downloaded from the MaizeSequence.org file transfer protocol (FTP) site (<http://ftp.maizesequence.org/>).

#### MaizeCyc Gene Product Annotation

To create MaizeCyc, we used the Pathway Tools software (Karp et al., 2002). Annotation was formatted according to the PathoLogic input file format. As described in Supplemental Table S1, we used Gramene gene identifiers (also known as Ensembl IDs) in the MaizeSequence.org's Ensembl Core Database (hereafter referred as Maize Core DB) to populate the "ID" attribute. The NAME attributes were assigned by using multiple approaches, (i) common gene names from the list of named maize genes (MaizeSequence.org, 2011) provided by Maize Core DB, (ii) common gene names for the gene models assigned by CoGePedia (Schnable and Freeling, 2011) and stored in the MaizeGDB's central database, (iii) additional names and gene symbols for a given gene model were stored as SYNONYM attribute, and (vi) for the gene models with no known name and/or symbol, the canonical transcript "ID" attribute of the transcript with the longest open reading frame (ORF) in the Maize Core DB was assigned as the NAME attribute. Accordingly, the STARTBASE and ENDBASE attributes were defined by the genomic span (chromosomal coordinates) of the canonical transcript. The gene models were then aligned to UniProt peptides and the corresponding best hits were selected according to the Exonerate score generated via the Ensembl XRef pipeline (Slater and Birney, 2005). FUNCTION values were primarily obtained from mapped UniProt entries via gene descriptors in Maize Core DB, gene product descriptions in MaizeGDB, and direct inference from gene ontology (GO) attributes (see below). Although multiple values for FUNCTION are plausible, we undertook a manual curatorial approach to distinguish truly distinct functions (e.g., bifunctional proteins) and alternative descriptions of the same FUNCTION (which we assigned as FUNCTION-SYNONYM). According to Maize Core DB's selection criteria, all models in the B73 RefGen\_v2 filtered set were protein-coding genes and therefore PRODUCT-TYPE code "P" (for protein or hypothetical ORF) was assigned. Enzyme Commission (EC) numbers were assigned based on the annotations extracted from (i) mapped UniProt entries, (ii) MaizeGDB curated pathways, and (iii) orthologous gene annotation projections from RiceCyc, AraCyc, and PlantCyc but supported by the phylogeny-based gene orthology tree clustering methods (Vilella et al., 2009) run on plant genomes by the Gramene and Plant Ensembl projects (Kersey et al., 2009; Youens-Clark et al., 2011). In addition, EC numbers were cross-checked with the latest

ENZYME release from the Enzyme Nomenclature Database at ExPASy (Schneider et al., 2004), and if available, corresponding EC-formatted MetaCyc (Caspi et al., 2012) cross-references were added. Gene ontology annotations were mainly drawn from the Maize Core DB, which are based on InterPro scans as part of the Ensembl Protein annotation pipeline, UniProt annotations, and Gramene Compara orthologous tree-based projections from *Arabidopsis thaliana* and rice (UniProt-GOA, 2011). Gene ontology terms were also inferred from EC annotations by using ec2go mappings provided by the Gene Ontology Consortium (Harris et al., 2004). In addition, GO terms for cellular compartment were manually assigned to the genes with evidence from proteomics-based metabolic reconstruction (Friso et al., 2010). DBLINK references were drawn to connect pathway annotations to the following external data sources: UniProt, Entrez, RefSeq, UniGene, MetaCyc, Gramene, MaizeGDB, MaizeGDB\_Locus, and MaizeSequence.org.

#### Mapping the Pathways and Reactions

The "Pathologic" option available from the Pathway Tools software (Karp et al., 2002) was executed with taxonomic filtering. This tool reads the annotations and finds matches with an existing library of EC numbers, names, reactions, and synonyms for gene, proteins, and reactions in its reference MetaCyc (version 15.5) database. MetaCyc was used as a default reference databases for finding the best matches for reactions and pathways.

#### Quality Control Filters

Quality control (QC) tools included in the Pathway Tools suite (Karp et al., 2002) were executed to check for consistency and annotations. (i) We removed annotations that were based on rice genes containing transposons. These genes carry portions of the well-known enzymes with EC assignments but have high likelihood of being nonfunctional. Eighty-eight such genes were excluded after confirming their protein description in UniProt and GO assignments. (ii) UniProt, MaizeGDB, and other manually curated databases (RiceCyc version 3.1, AraCyc version 8, and PlantCyc version 5) were referred to for assignments of EC numbers to enzymes. We also checked for the most recent updates to EC numbers corresponding to enzyme descriptions provided by the ExPASy Enzyme Database at <http://enzyme.expasy.org/> (10 July 2011). (iii) Enzyme names were manually edited to match with associated EC and GO terms, as slightly different names of the same protein were found in different sources. The variant enzyme names were added as FUNCTION-SYNONYM attributes for EC and/or GO match with the best match added as the main label. (iv) Pathways specific to bacteria, fungi, and animals were removed. (v) We manually curated name-based assignments. Conflicted EC annotations were not included in the database, but errors were flagged to alert curators to make recommendations. We curated 200 of such cases based on BLASTP results against UniProt. Additional



checks included visual confirmation of domains (InterPro) present on the suggested list of enzymes. Finally, we performed updates and rescoring of pathways. Pathways and reactions were routinely rescored and cross-checked against the manually curated gold standard reference libraries provided by MetaCyc and PlantCyc. This was performed after each cycle of manual checks involving acceptance, edits, or rejections.

## Gene Expression Data Analysis

### Microarray Data Set

The 80,301 NimbleGen microarray probe sets (Sekhon et al., 2011) were mapped in this report to the B73 Ref-Gen\_v2 Working Gene Set (which includes the FGS used by MaizeCyc) by aligning each individual probe to the pseudomolecules and requiring a 100% coverage match to a gene model. Up to two base mismatches were permitted. Probes that mapped to more than two places were dropped and the remaining probes were assigned probe sets by mapping them to consensus gene models (not the alternative splice forms). The probe set mappings were shared with the PLEXdb database (Dash et al., 2011) where the data were renormalized (experiment identifier [ID] ZM37).

### Gene Expression Data Analysis

The maize gene expression atlas data set was developed for 60 diverse tissues previously by Sekhon et al. (2011) and annotated (Schaeffer et al., 2011b) with plant ontology (PO) terms (Jaiswal et al., 2005; Cooper et al., 2012; Walls et al., 2012) by curating the appropriate plant anatomical entity (Ilic et al., 2007; Cooper et al., 2012) and the respective plant growth and development stage (Pujar et al., 2006). In this report we focused the gene expression analysis on five major organ types: (i) V1\_pooled leaves (pooled leaves from the plants vegetative stage V1) (L), (ii) V1\_primary root (primary roots from plants at the vegetative stage V1) (R), (iii) R1\_anthers (anthers from florets and plants at the reproductive stage R1) (A), (vi) 16DAP\_endosperm (endosperms from developing seeds at 16 d after pollination [16DAP]) (D), and (v) 16DAP\_embryo (embryo from developing seeds at 16DAP) (E) (Supplemental Table S2). These tissue types were reported to have high replicate correlations (correlation coefficients > 0.95 and  $P < 0.001$ , as reported in the supplementary table S2 of Sekhon et al. [2011]). The renormalized gene expression data from the selected five source tissue samples downloaded from PLEXdb. The data set was analyzed to remove low and/or nonexpressed genes with an expression threshold cutoff of five times the highest signal intensity of the negative probes (random sequence signal range of 25–65). The resulting gene list was further filtered to identify genes that show expression levels fourfold or more in only one of the five tissue types that we analyzed. The resulting 10,057 genes showing tissue specific upregulation were used for further analysis (Supplemental Table S3). Hierarchical clustering of mean centered expression patterns based on Pearson correlation was performed using GeneSpring

7.4 (Agilent Technologies, 2008). The tissue-specific gene sets extracted from the expression pattern clusters were mapped to MaizeCyc metabolic pathways using the OMICs viewer tool provided by the Pathway Tools (Karp et al., 2002) and software scripts developed in-house to identify pathways, genes, and reactions dominant in specific tissue types (Supplemental Tables S4, S5, S6, and S7). The gene loci based annotation using the PO was submitted to the PO database (Schaeffer et al., 2011b).

## Data Downloads

Users can retrieve the MaizeCyc data set from the “Download” link provided in the table on the MaizeCyc page at <http://maizecyc.maizegdb.org> (Fig. 1). Installation requires users to obtain licensed Pathway Tools software (Karp et al., 2002) available from SRI International (<http://www.ecocyc.org/download.shtml>). For advanced users working on network modeling, we provide pathway data dumps in the standardized BioPax levels 2 and 3 (Demir et al., 2010) and systems biology markup language (SBML) (Hucka et al., 2003) formats. These files are compatible for viewing networks using software such as Cytoscape (Killcoyne et al., 2009). The expression data set used here is available in Supplemental Table S3 and also over the web at [http://ftp.maizegdb.org/MaizeGDB/FTP/MaizeCyc\\_manuscript\\_files/](http://ftp.maizegdb.org/MaizeGDB/FTP/MaizeCyc_manuscript_files/).

## RESULTS

### Manual Curation of the MaizeCyc Database

As described in the Methods section, after several rounds of computational analysis, quality control steps, and manual curation, we successfully projected a maize metabolic network. Manually curated pathways included carotenoid biosynthesis (from lycopene to carotene and xanthophylls) and flavonoid and flavonol biosynthesis leading to anthocyanin biosynthesis and accumulation. Carotenoid biosynthesis is a superpathway that includes the  $\beta$ -,  $\delta$ -, and  $\epsilon$ -carotene biosynthesis, lutein biosynthesis, and zeaxanthin biosynthesis pathways. Similarly, flavonoid and flavonol biosynthesis encompasses leucopelargonidin and leucocyanidin biosynthesis, luteolin biosynthesis, flavonol biosynthesis, and flavonoid biosynthesis. Manual curation also involved confirming computational mappings and assigning genes to reactions from pathways that were missed by automated mapping. MaizeCyc consists of 391 pathways with 8889 enzymes (about 22% of the FGS) and 291 transporter proteins, mapping to 2110 enzymatic and 68 transport reactions, respectively, in addition to 1468 compounds. In building MaizeCyc, we found that not every reaction of a given pathway is supported by a mapped protein responsible for performing a given enzymatic activity. The possible reasons are (i) the enzyme in question had not been previously identified from a plant and/or other source and/or (ii) although it may be a known enzyme, there was no entry in the reference library maintained by Pathway Tools (Karp et al., 2002) and that annotation is awaiting secondary curation.

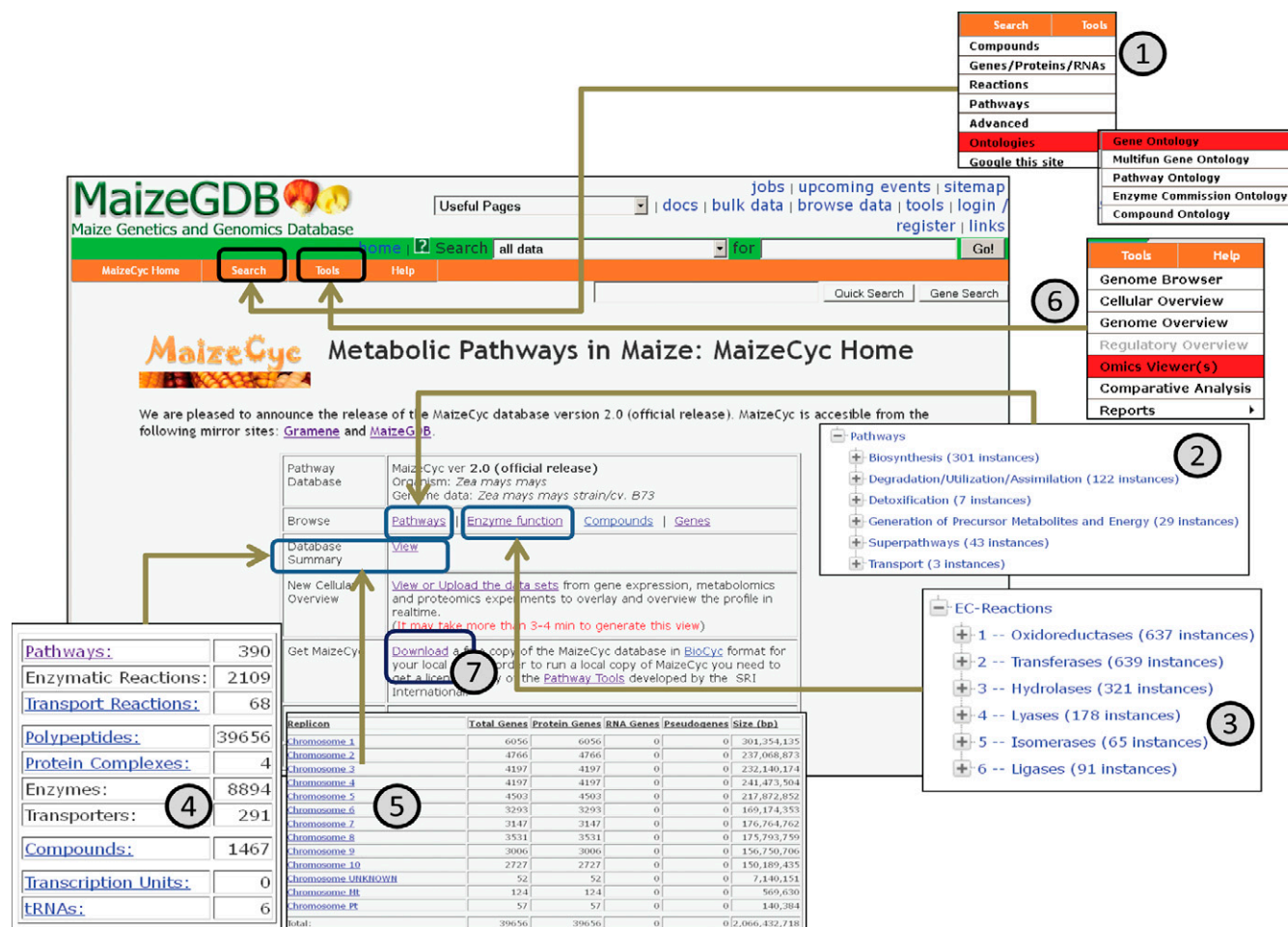


Figure 1. The MaizeCyc database web portal is accessible from <http://maizecyc.maizegdb.org>. (1) Search menu, (2) Pathway browsing, (3) Enzyme function browsing, (4) Summary of database statistics, (5) Database statistics broken down for each chromosome, (6) Tools, and (7) Download link. For example, when the pathways link is clicked on the query results page, the user is directed to a page where pathways are displayed under functional classification provided in an expandable hierarchical view (Step 2). The database summary includes the total number of pathways, reactions, proteins, and pathways (Step 4) as well as more granular information including gene distributions (protein coding, ribonucleic acid [RNA], and pseudogenes) in a tabular format for each chromosome (Step 5).

## MaizeCyc: A Resource for Biologists

Users can access instances of MaizeCyc via both the MaizeGDB (<http://maizecyc.maizegdb.org>; Fig. 1 and 2) and Gramene (<http://pathway.gramene.org>) websites. We provide a short MaizeCyc tutorial (Supplemental File S1) aimed at new users on how to browse, search, and use tools such as the OMICS viewer (Karp et al., 2002) (Supplemental Fig. S1) to overlay gene, protein, and metabolic expression data sets. Excellent tutorial webinars are also available from the BioCyc website (<http://biocyc.org/webinar.shtml>).

## Differential Expression of Metabolic Pathway Genes

To identify differentially expressed genes, reactions, and pathways and to evaluate the utility of the MaizeCyc resource, we analyzed the recently published microarray-based maize gene expression atlas (Sekhon et al., 2011) data set focusing on metabolic pathways. The analysis was performed on five selected tissue samples: (a) pooled leaves from the vegetative stage V1 (where only one leaf is fully expanded), (b) primary root at the V1, (c) anthers

at the reproductive stage R1 (silks emerge from the husk), (d) embryo from a developing seed at 16DAP, and (e) endosperm, also at 16DAP. Of the 10,057 genes identified as upregulated in these five organs, 7557 genes were listed in the FGS used to reconstruct MaizeCyc. Of these, only 1957 genes mapped to known reactions in MaizeCyc (Fig. 3; Supplemental Tables S3 and S4). In the database, not all reactions are part of a pathway; therefore, out of a total 1957 differentially expressed MaizeCyc gene entries we found 1062 genes mapped to 513 unique reactions associated with 308 pathways (Supplemental Table S5). Among the 1062 pathway associated genes, the greatest number, 338 genes, were upregulated in leaf sample and the least number, 90 genes, were upregulated in endosperm (Table 1; Supplemental Table S6).

## Insight That Can Be Gained for Experimental Research Using MaizeCyc

The summary and breakdown of these upregulated genes for the tissues analyzed here can be found in Table 1 and Supplemental Table S6. Of the 310 unique and upregulated

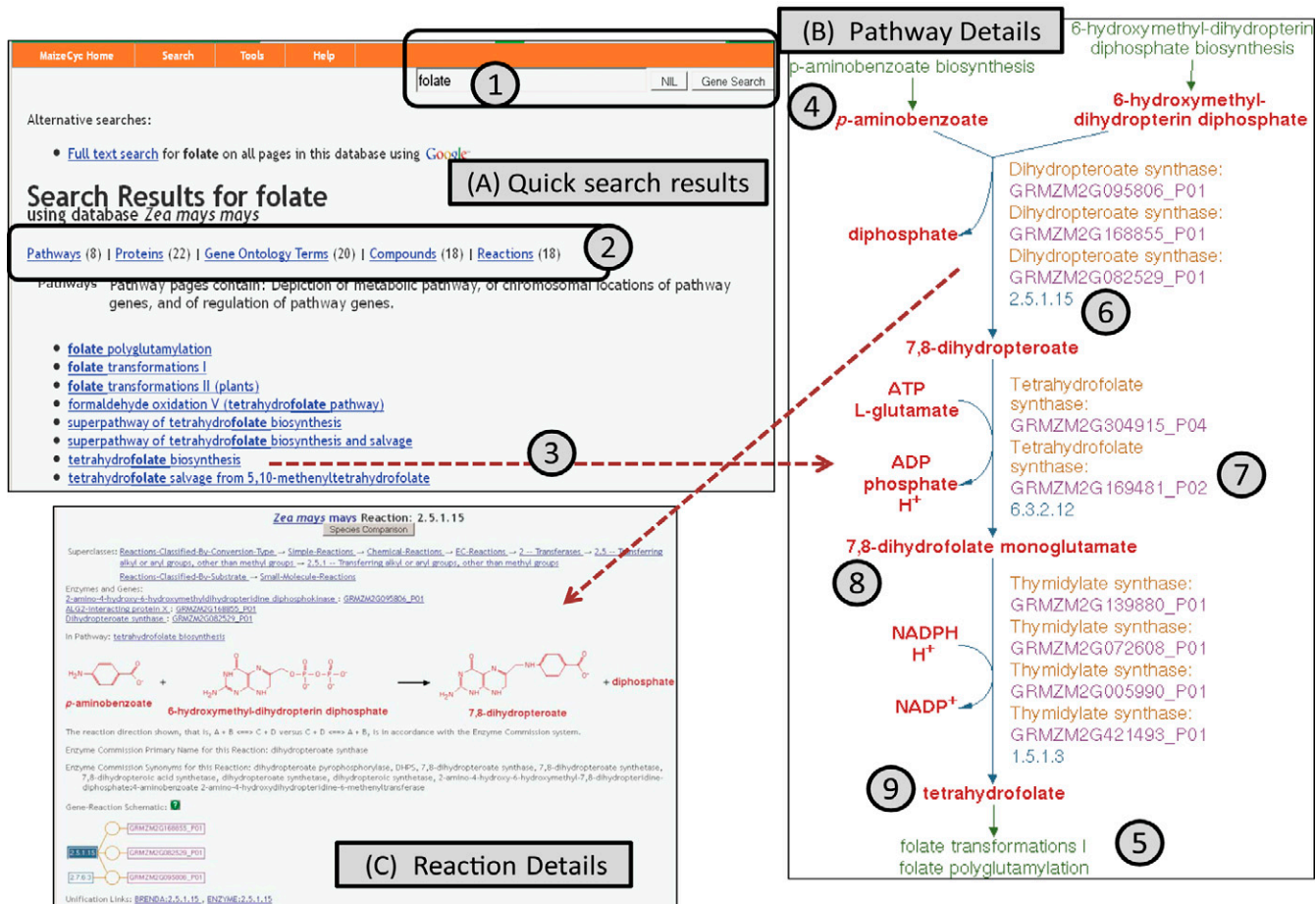


Figure 2. Search results and detailed views of pathways and reactions. Here the keyword “folate” is used as an example. (A) Search and search results view: (1) the search term is entered in the search box and (2) results are categorized and (3) linked to pathway detail page. (B) Pathway details: the users can obtain information about linked pathways (4 and 5), (6) Enzyme Commission numbers, and (7) gene identifier (ID) of the enzymes involved in reactions and metabolites as (8 and 9) reactants and products. (C) Reaction detail view: Clicking on the compound name and/or structure leads to compound detail page (not shown). Every page provides a list of associated literature citations and cross references to similar information in other resources such as Kyoto Encyclopedia of Genes and Genomes (KEGG), BRENDA (<http://www.brenda-enzymes.info>), etc.

pathways, 32 were upregulated in all five tissue samples, and some were uniquely upregulated per tissue: 22 pathways in anther, 11 in embryo, 15 in endosperm, 42 in leaf, and 22 in root (Fig. 3b; Supplemental Table S6). Among the commonly upregulated pathways, biosynthetic pathways for cellulose, flavonol, flavonoid, suberin (Supplemental Fig. S2),  $\beta$ -caryophyllene, and the Calvin-Benson-Bassham cycle were upregulated in all organs except the endosperm. The biosynthesis pathways of chlorophyllide *a* (Fig. 4) and  $\beta$ -carotene, zeaxanthin, and xanthophylls (Supplemental Fig. S3) were upregulated in leaf. The other carotenoid-like  $\delta$ - and  $\epsilon$ -carotene biosynthesis genes were upregulated only in anther samples. The geranyl-geranyl-diphosphate biosynthesis and the transycopene biosynthesis pathways that provide precursor metabolites for anthocyanin and carotenoids biosynthesis, respectively, were upregulated in leaf samples whereas the anthocyanin biosynthesis pathway (Supplemental Fig. S4) was upregulated in anther, leaf, and primary root samples. Consistent with the findings reported by Sekhon et al. (2011), the lignin (Fig. 4) and suberin (Supplemental Fig.

S2) biosynthesis pathways were found preferentially upregulated in root samples. This is expected given that the casparian strip in the root endodermis forms an extracellular barrier to apoplastic transport of water and solute loading from the root cortex to the xylem in plant roots and is highly suberized (Zeier et al., 1999; Baxter et al., 2009). Besides root-specific overexpression of suberin pathway genes, we also observed upregulation in nonroot samples for genes from the phenylpropanoid-related suberin pathway encoding *trans*-feruloyl-coenzyme A (CoA) synthase (EC 6.2.1.34), caffeoyl-CoA-*O*-methyltransferase (EC 2.1.1.104), 4-coumarate-CoA ligase (EC 6.2.1.12), and phenylalanine ammonia lyase (EC 4.3.1.24) (Supplemental Fig. S2). Xylan and xyloglucan biosynthesis genes were expressed mainly in the embryo while the expression of genes mapping to indole3-acetic acid conjugate biosynthesis and fatty acid activation appeared to be limited to embryo and endosperm tissues (Supplemental Table S7). A detailed discussion of differentially expressed genes associated with the processes of photosynthesis and lignification in maize can be found in Supplemental File S2.



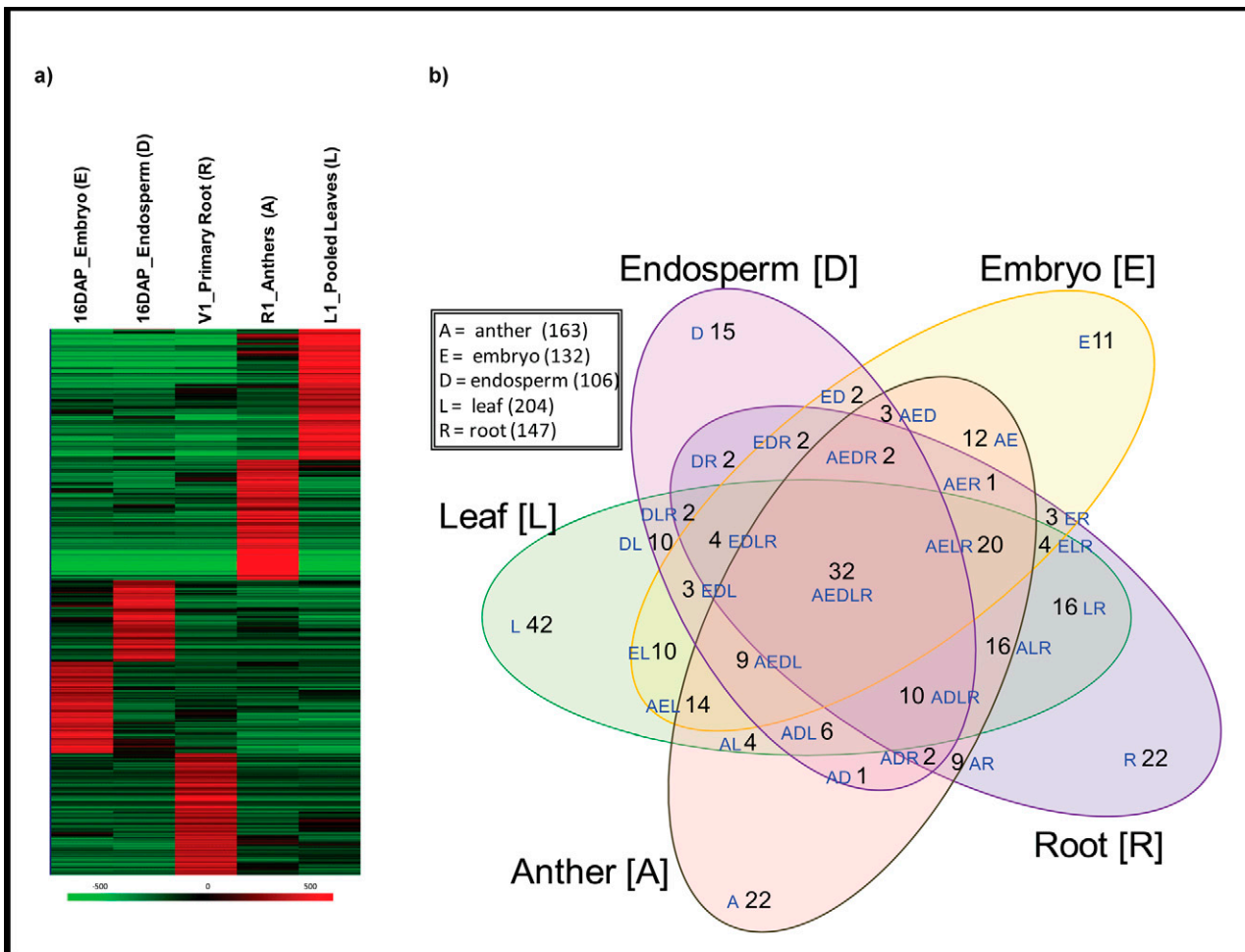


Figure 3. (A) Hierarchical clustering of tissue-specific gene expression profiles. The identified 10,057 genes showed 4.0-fold or more upregulation in one of the five tissue types. Mean centered expression levels are represented with green color denoting downregulation, and red color indicates upregulation. (B) Venn diagram of tissue-specific enrichment of metabolic pathways associated to the upregulated genes identified in Fig. 1 and Supplemental Tables S5 and S6. As described by the authors (Sekhon et al., 2011), the ribonucleic acid (RNA) samples are from E, which represents 16DAP\_embryo (embryo from developing seeds at 16 d after pollination [16DAP]), D, which represents 16DAP\_endosperm (endosperms from developing seeds at 16DAP), R, which represents V1\_primary root (primary roots from plants at the vegetative stage V1), A, which represents R1\_anther (anthers from florets and plants at the reproductive stage R1), and L, which represents V1\_pooled leaves (pooled leaves from the plants vegetative stage V1). The plant ontology annotations of tissue samples are provided in Supplemental Table S3.

**Table 1. Expression data analysis of differentially expressed genes associated to pathways and reactions.**

	Embryo	Endosperm	Root	Anther	Leaf
(A) Number of genes upregulated in organ-specific cluster	1680	1503	2247	2211	2416
(B) Number of upregulated genes mapped to pathways	129	90	245	257	341
(C) Number of reactions mapped to upregulated genes	149	112	152	180	285
(D) Number of pathways mapped to upregulated genes	128	106	146	159	199
(E) Average no. of reactions with associations to upregulated genes in pathways where $E = (C/D)$	1.16	1.06	1.04	1.13	1.42
(F) Average no. of genes upregulated per reaction where $F = (B/C)$	0.87	0.80	1.61	1.43	1.20

## DISCUSSION

Structure-function studies of enzymes and metabolic pathways have been extensively used to extract functional annotations of metabolic pathways and enzymes derived from the sequenced genomes. We created the core of MaizeCyc using electronic annotations of enzymes and metabolic pathways and manually curated annotations based on experimental evidence and

assignments reported in published literature. The KEGG database of pathways is one of the most sought-after databases for performing metabolic data analysis, which includes maize genes mapped to pathways and reactions. However, the pathway views in KEGG are based on reference pathways that are derived from many organisms and therefore largely represent a species-neutral view. In maize there are species-specific versions of some

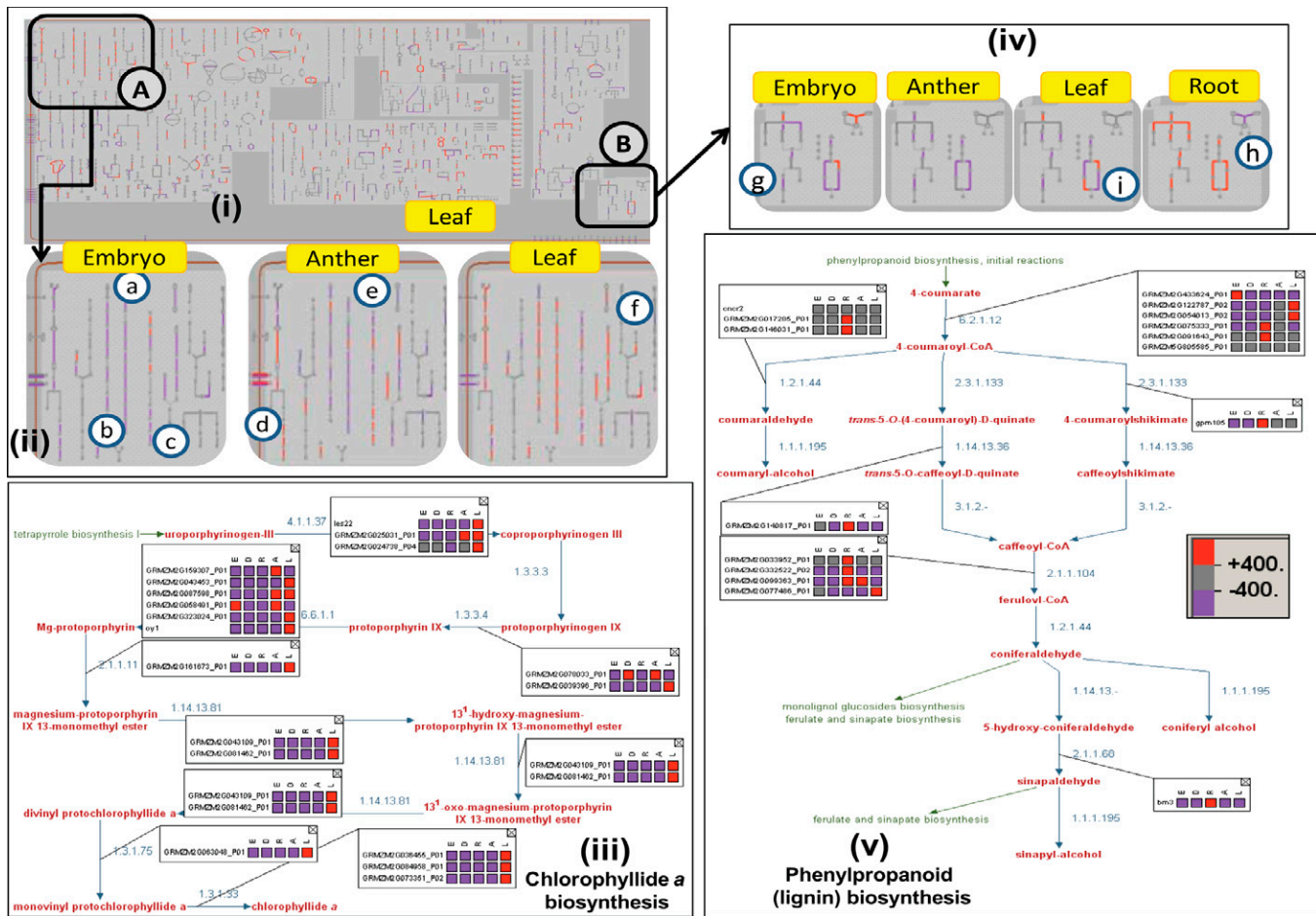


Figure 4. The OMICs viewer (Karp et al., 2002) analysis of the overexpressed genes from five different tissue samples. (i) Cellular overview of the pathways, reactions, and transporters painted with the gene expression data set from leaf as an example. Purple color represents downregulation, and red color represents upregulation. The view was generated on a local desktop version of the database using Pathway Tools software (Karp et al., 2002) and the color scheme was selected for three-color display with threshold of 400. (ii) Zoom-in view of the painted section A of the cellular overview from embryo, anther, and leaf tissue. In these views, the highlighted pathways include (a) chlorophyllide *a* biosynthesis, (b) tetrapyrrol and heme biosynthesis, (c) flavin biosynthesis, (d) geranyl-geranyl-diphosphate biosynthesis, (e) mevalonate biosynthesis, and (f) tetrahydrofolate biosynthesis. (iii) A pop-up view of the chlorophyllide *a* biosynthesis pathway with the painted gene expression identified for each of the expressed homologs. In the pathway, each reaction is identified by its Enzyme Commission (EC) number (in blue) connecting the two compounds (in red). The arrows point in the direction of the reaction. In the expression data blocks (available only in the locally installed version of Pathway Tools), rows correspond to gene "ID" attribute or symbol, and column headers represent expression data from E, which represents embryo, D, which represents endosperm, R, which represents root, A, which represents anther, and L, which represents leaf. Majority of the genes show leaf specific expression. (iv) A zoom-in view of the section B of the cellular overview from embryo, root, anther, and leaf expression data samples. The painted pathways include biosynthetic pathways of (g) phenylpropanoid or lignin, (h) xylan, and (i) suberin. (v) A pop-up view with reaction, compound, and enzyme (EC number) details of the lignin biosynthesis pathway painted with expression data from the five different ribonucleic acid (RNA) samples. Majority of genes show root-specific expression. The expression data is provided in the Supplemental Table S3.

pathways that deviate from the reference KEGG pathway (Zelitch, 1973). In addition, KEGG pathways are not associated with literature citations and it is difficult to check the accuracy of their annotation based on experimental evidence. For plants, MetaCyc and PlantCyc are more suitable reference databases, because both are enriched with manually curated plant-specific primary metabolic pathways that are either universal to plants or unique to a species. These plant-specific data sets are provided with the caveat that they may contain only a partial set of all known secondary metabolic pathways

due to curation lag and some limited contribution from community curation. Many databases such as KEGG set their own priorities on curation of pathways and do not allow direct curation by the community. In contrast, Bio-Cyc databases permit data editing by community authors (in this case, MaizeGDB and Gramene). Authors can modify the reference pathway and/or create new pathways and a subset of reactions to better represent specific aspects of the plant's biology. See, for example, the well-known  $\beta$ -carotene (provitamin A) biosynthesis (Wurtzel et al., 2012) (Supplemental Fig. S3) and  $C_4$  photosynthesis



(Slack et al., 1969; Sheen and Bogorad, 1987; Nomura et al., 2000) pathways.

The current version (2.0.2) of MaizeCyc was constructed based on the high-confidence protein coding genes identified in the maize reference genome sequenced from the inbred line B73. Our integrated approach on assigning functional annotations to genes include computational analysis of functional assignments supported by phylogenetic and syntenic relationships to allow integration of the known gene functions and pathways from maize, *Arabidopsis thaliana*, and rice. An advantage of developing such a metabolic pathway database is that it allows cross-species comparisons of networks and gene assignments to find functional orthologs. Such comparisons are available from the MaizeCyc mirror at the Gramene database.

Here, we also present a comprehensive analysis of the tissue-specific expression of genes that are represented in the metabolic network. Gene expression analysis of leaf, primary root, anther, endosperm, and embryo tissues revealed that about 20% of the 10,057 differentially expressed genes (Fig. 3a) map to 310 (Fig. 3b) of the total 391 metabolic pathways. We were able to identify differentially regulated genes mapping uniquely to 22 unique metabolic pathways in anther, 12 in embryo, 14 in endosperm, 40 in leaf, and 23 in root (Fig. 3b). As many of the pathways are found in multiple tissues, it is highly likely that these genes represent homologs specific to these tissues.

While we continue these efforts to improve MaizeCyc, improvements in genome annotation and growing evidence provided by experimental analyses including high-throughput phenotyping, gene expression (transcriptomics and proteomics), and metabolomics can be incorporated as they become available. Such data sets will help validate what was annotated computationally in the current version and will help add new information. We welcome suggestions from the community concerning deficiencies and inconsistencies in the knowledge areas represented by the MaizeCyc database, and we will be happy to work with researchers to improve Pathway Tools (Karp et al., 2002) representations of maize metabolic pathways. Subsequent releases of MaizeCyc will likely include enrichment of the networks by including, for example, (i) annotations captured in the published metabolic networks for maize (Dal’Molin et al., 2010; Saha et al., 2011), (ii) subcellular locations identified by computational methods (Westerlund et al., 2003; Small et al., 2004; Emanuelsson et al., 2007), (iii) findings from proteomics experiments (Zybailov et al., 2008; Majeran et al., 2011), (iv) references to the probe set identifiers from various expression platforms developed for maize, and (v) references to the gene coordinates and identifiers (IDs) from the new assemblies of the B73 maize genome and genomes from other inbred lines and diverse materials (e.g., Mo17, Palomero Toluqueño, and others).

To our knowledge, the MaizeCyc metabolic pathway resource we have developed is one of the first

attempts to establish a comprehensive approach for reconstructing metabolic pathways in a manner that both complements and contributes to the maize genome’s functional annotation. MaizeCyc analysis provides specific references to the candidate genes and their tight association to metabolic function. It is available for maize researchers to browse, search, and use as a tool to guide their research. In addition, MaizeCyc provides a new option or context within which researchers can analyze metabolic pathway representations in large-scale transcriptome, metabolic, and proteomics studies.

### Supplemental Information Available

Supplemental material is included with this manuscript.

Supplemental Figure S1. The Omics Viewer description. (1) The cellular overview (available in the online and local desktop version of the Pathway Tools software [Karp et al., 2002]) (2) and genome overview chart of the maize gene expression atlas data (Sekhon et al., 2011) for shoot apical meristem and stem V4 expression. (3) The ratio of expressed genes in (a) shoot apical meristem and stem V4 vs. leaf base of expanding leaf V5 and (b) embryo 24 days after pollination vs. kernel 24 days after pollination. Suberin biosynthesis pathway (red box) and C4 photosynthetic carbon assimilation cycle pathways (blue box) are highlighted to show tissue-specific expression differences. (4) From the online version of the tool, a zoomed view of (a) suberin biosynthesis pathways for shoot apical meristem and stem V4 vs. leaf base of expanding leaf V5 and (b) embryo 24 days after pollination vs. kernel 24 days after pollination.

Supplemental Figure S2. A pop-up view on the desktop version of the tool, with reaction, compound, and enzyme (Enzyme Commission [EC] number) details of the suberin biosynthesis pathway painted with expression data from the five different ribonucleic acid (RNA) samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally installed version of Pathway Tools [Karp et al., 2002]), rows correspond to gene “ID” attribute or symbol, and column headers represent expression data from E:embryo, D:endosperm, R:root, A:anther, and L:leaf.

Supplemental Figure S3. A pop-up view of the desktop version of the pathway tool, with reaction, compound, and enzyme (Enzyme Commission [EC] number) details of the carotenoid biosynthesis pathway painted with expression data from the five different ribonucleic acid (RNA) samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally installed version of Pathway Tools [Karp et al., 2002]), rows correspond to gene “ID” attribute or symbol, and column headers represent expression data from E:embryo, D:endosperm, R:root, A:anther, and L:leaf.

Supplemental Figure S4. A pop-up view of the desktop version of the pathway tool, with reaction, compound, and enzyme (Enzyme Commission [EC] number) details of the anthocyanin biosynthesis pathway painted with expression data from the five different ribonucleic acid (RNA)

samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally-installed version of Pathway Tools [Karp et al., 2002]), rows correspond to gene “ID” attribute or symbol, and column headers represent expression data from E:embryo, D:endosperm, R:root, A:anther, and L:leaf.

Supplemental File S1. A MaizeCyc tutorial, featuring a section describing how to perform gene expression data analyses with the OMICs viewer.

Supplemental File S2. Detailed discussion of differentially expressed genes associated with photosynthesis and lignification pathways in maize.

Supplemental Table S1. The attributes of the functional annotation file in the PathoLogic format and source database and database element used to create these attributes. Ensembl Core application programming interfaces (APIs) were used to query and retrieve records from Gramene and may be requested by research groups to create their own custom pipelines. For MaizeCyc version 2.0.2, Gramene version 32 and Ensembl Core D for maize was used.

Supplemental Table S2. List of source tissue samples and the mapping to plant ontology terms. We are listing only the tissue samples used in the gene expression analysis. For the complete list please see (Schaeffer et al., 2011b).

Supplemental Table S3. Mean centered expression data values for 10,057 genes from the five tissue samples.

Supplemental Table S4. List of all the tissue specific genes (10,057 genes) from maize based on the fourfold cutoff.

Supplemental Table S5. Complete list of expressed gene IDs mapped to MaizeCyc reaction identifier (ID) and the MaizeCyc pathway name.

Supplemental Table S6. List of pathways and corresponding tissue- or organ-specific cluster. Counts used for creating the Venn Diagram (Figure 3B).

Supplemental Table S7. Complete list of pathways and the associated organs. Includes information on (1) pathway name, (2) number of expressed genes mapped to the pathway, (3) number of reactions to which expressed genes map in the pathway, (4) total number of reactions in the pathway, and (5) total number of genes mapped in the pathway (6) organ-specific cluster.

## Acknowledgments

We are grateful to PLEXdb and John Van Hemert for normalizing the remapped expression data for the maize gene expression atlas; Peter Karp, Markus Krummenacker, and Suzanne Paley from SRI for excellent Pathway Tools support; Peter van Buren and Ken Youens-Clark from Cold Spring Harbor Laboratory for help on setting up the pathway database webservice installation and for successfully integrating the MaizeCyc mirror database in the Gramene project database; Shiran Pasternak and Joshua Stein from the Maize Genome Sequencing Project and Cold Spring Harbor Laboratory for providing the FGS and helping identify false positive annotations; Carson Andorf, Darwin Campbell, and Bremen Braun at MaizeGDB for supporting the database infrastructure and integration into the MaizeGDB resource; Justin Elser and Laurel Cooper from the Plant Ontology project at Oregon State University and Ramona Walls from the New York Botanical Garden and Plant Ontology project ([www.plantontology.org](http://www.plantontology.org)) for successfully integrating the maize atlas PO annotation in the PO data archives and database; and Kate Dreher at the Plant Metabolic Network project for training MaizeGDB staff and

sharing annotations. This work was supported by the NSF Plant Genome Research Resource grant award IOS:0703908 (Gramene: A Platform for Comparative Plant Genomics) and the USDA-ARS and National Corn Growers Association (The Maize Genetics and Genomics Database).

## References

- Agilent Technologies. 2008. GeneSpring 7.4 Agilent Technologies Inc., Santa Clara CA.
- Baxter, I., P.S. Hosmani, A. Rus, B. Lahner, J.O. Borevitz, B. Muthukumar, et al. 2009. Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. *PLoS Genet* 5(5):e1000492.
- Caspi, R., T. Altman, K. Dreher, C.A. Fulcher, P. Subhraveti, I.M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L.A. Mueller, et al. 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40:D742–D753. doi:10.1093/nar/gkr1014
- Cooper, L., R.L. Walls, J. Elser, M.A. Gandolfo, D.W. Stevenson, B. Smith, J. Preece, B. Athreya, C.J. Mungall, S. Rensing, et al. 2012. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* doi:10.1093/pcp/pcs163
- Dal’Molin, C.G., L.E. Quek, R.W. Palfreyman, S.M. Brumbley, and L.K. Nielsen. 2010. C4GEM, a genome-scale metabolic model to study C4 plant metabolism. *Plant Physiol.* 154:1871–1885. doi:10.1104/pp.110.166488
- Dash, S., J. Van Hemert, L. Hong, R.P. Wise, and J.A. Dickerson. 2011. PLEXdb: Gene expression resources for plants and plant pathogens. *Nucleic Acids Res.* 40:D1194–D1201. doi:10.1093/nar/gkr938
- Demir, E., M.P. Cary, S. Paley, K. Fukuda, C. Lemer, I. Vastrik, G. Wu, P. D’Eustachio, C. Schaefer, J. Luciano, et al. 2010. The BioPAX community standard for pathway data sharing. *Nat. Biotechnol.* 28:935–942. doi:10.1038/nbt.1666
- Edwards, G.E., R.T. Furbank, M.D. Hatch, and C.B. Osmond. 2001. What does it take to be C4? Lessons from the evolution of C4 photosynthesis. *Plant Physiol.* 125:46–49. doi:10.1104/pp.125.1.46
- Emanuelsson, O., S. Brunak, G. von Heijne, and H. Nielsen. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2:953–971. doi:10.1038/nprot.2007.131
- FAOSTAT. 2011. Food and agricultural commodities production. FAO, Rome, Italy. <http://faostat.fao.org/site/339/default.aspx> (accessed 5 Nov. 2011).
- Ferrer, M. 2009. The microbial reactome. *Microb. Biotechnol.* 2:133–135. doi:10.1111/j.1751-7915.2009.00090\_5.x
- Friso, G., W. Majeran, M. Huang, Q. Sun, and K.J. van Wijk. 2010. Reconstruction of metabolic pathways, protein expression, and homeostasis machineries across maize bundle sheath and mesophyll chloroplasts: Large-scale quantitative proteomics using the first maize genome assembly. *Plant Physiol.* 152:1219–1250. doi:10.1104/pp.109.152694
- Gaut, B.S., M. Le Thierry d’Ennequin, A.S. Peek, and M.C. Sawkins. 2000. Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* 97:7008–7015. doi:10.1073/pnas.97.13.7008
- Grafahrend-Belau, E., F. Schreiber, D. Koschutski, and B.H. Junker. 2009. Flux balance analysis of barley seeds: A computational approach to study systemic properties of central metabolism. *Plant Physiol.* 149:585–598. doi:10.1104/pp.108.129635
- Harris, M.A., J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, et al. 2004. The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32:D258–D261. doi:10.1093/nar/gkh066
- Hatch, M.D. 1978. Regulation of enzymes in C4 photosynthesis. *Curr. Top. Cell. Regul.* 14:1–27.
- Hucka, M., A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, A.P. Arkin, B.J. Bornstein, D. Bray, A. Cornish-Bowden, et al. 2003. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531. doi:10.1093/bioinformatics/btg015
- Ilic, K., E.A. Kellogg, P. Jaiswal, F. Zapata, P.F. Stevens, L.P. Vincent, S. Avraham, L. Reiser, A. Pujar, M.M. Sachs, et al. 2007. The plant structure ontology, a unified vocabulary of anatomy and

- morphology of a flowering plant. *Plant Physiol.* 143:587–599. doi:10.1104/pp.106.092825
- Jaiswal, P., S. Avraham, K. Ilic, E.A. Kellogg, S. McCouch, A. Pujar, L. Reiser, S.Y. Rhee, M.M. Sachs, M. Schaeffer, et al. 2005. Plant ontology (PO): A controlled vocabulary of plant structures and growth stages. *Comp. Funct. Genomics* 6:388–397. doi:10.1002/cfg.496
- Jaiswal, P., J. Ni, I. Yap, D. Ware, W. Spooner, K. Youens-Clark, L. Ren, C. Liang, W. Zhao, K. Ratnapu, et al. 2006. Gramene: A bird's eye view of cereal genomes. *Nucleic Acids Res.* 34:D717–D723. doi:10.1093/nar/gkj154
- Kanehisa, M., S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. 2011. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40:D109–D114. doi:10.1093/nar/gkr988
- Karp, P.D., S. Paley, and P. Romero. 2002. The Pathway Tools software. *Bioinformatics* 18(Suppl. 1):S225–S232. doi:10.1093/bioinformatics/18.suppl\_1.S225
- Kersey, P.J., D. Lawson, E. Birney, P.S. Derwent, M. Haimel, J. Herrero, S. Keenan, A. Kerhornou, G. Koscielnny, A. Kahari, et al. 2009. Ensembl genomes: Extending Ensembl across the taxonomic space. *Nucleic Acids Res.* 38:D563–D569. doi:10.1093/nar/gkp871
- Killcoyne, S., G.W. Carter, J. Smith, and J. Boyle. 2009. Cytoscape: A community-based framework for network modeling. *Methods Mol. Biol.* 563:219–239. doi:10.1007/978-1-60761-175-2\_12
- Latendresse, M., S. Paley, and P.D. Karp. 2011. Browsing metabolic and regulatory networks with BioCyc. *Methods Mol. Biol.* 804:197–216. doi:10.1007/978-1-61779-361-5\_11
- MaizeSequence.org. 2011. Maize gene names. City: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. <http://maizesequence.org/info/docs/namedgenes.html>. (accessed 10 July 2011)
- Majeran, W., G. Friso, Y. Asakura, X. Qu, M. Huang, L. Ponnala, K.P. Watkins, A. Barkan, and K.J. van Wijk. 2011. Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: A new conceptual framework for nucleoid functions. *Plant Physiol.* 158:156–189. doi:10.1104/pp.111.188474
- Mueller, L.A., T.H. Solow, N. Taylor, B. Skwarecki, R. Buels, J. Binns, C. Lin, M.H. Wright, R. Ahrens, Y. Wang, et al. 2005. The SOL genomics network: A comparative resource for Solanaceae biology and beyond. *Plant Physiol.* 138:1310–1317. doi:10.1104/pp.105.060707
- Mueller, L.A., P. Zhang, and S.Y. Rhee. 2003. AraCyc: A biochemical pathway database for Arabidopsis. *Plant Physiol.* 132:453–460. doi:10.1104/pp.102.017236
- Nomura, M., N. Sentoku, A. Nishimura, J.H. Lin, C. Honda, M. Taniguchi, Y. Ishida, S. Ohta, T. Komari, M. Miyao-Tokutomi, et al. 2000. The evolution of C4 plants: Acquisition of cis-regulatory sequences in the promoter of C4-type pyruvate, orthophosphate dikinase gene. *Plant J.* 22:211–221. doi:10.1046/j.1365-313x.2000.00726.x
- Papp, B., C. Pal, and L.D. Hurst. 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429:661–664. doi:10.1038/nature02636
- Pilalis, E., A. Chatziioannou, B. Thomasset, and F. Kolisis. 2011. An in silico compartmentalized metabolic model of *Brassica napus* enables the systemic study of regulatory aspects of plant central metabolism. *Biotechnol. Bioeng.* 108:1673–1682. doi:10.1002/bit.23107
- Poolman, M.G., L. Miguet, L.J. Sweetlove, and D.A. Fell. 2009. A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiol.* 151:1570–1581. doi:10.1104/pp.109.141267
- Pujar, A., P. Jaiswal, E.A. Kellogg, K. Ilic, L. Vincent, S. Avraham, P. Stevens, F. Zapata, L. Reiser, S.Y. Rhee, et al. 2006. Whole-plant growth stage ontology for angiosperms and its application in plant biology. *Plant Physiol.* 142:414–428. doi:10.1104/pp.106.085720
- Rabinowicz, P.D., and J.L. Bennetzen. 2006. The maize genome as a model for efficient sequence analysis of large plant genomes. *Curr. Opin. Plant Biol.* 9:149–156. doi:10.1016/j.pbi.2006.01.015
- Rawsthorne, S. 1992. Towards an understanding of C3–C4 photosynthesis. *Essays Biochem.* 27:135–146.
- Sage, T.L., and R.F. Sage. 2009. The functional anatomy of rice leaves: Implications for refixation of photorespiratory CO<sub>2</sub> and efforts to engineer C4 photosynthesis into rice. *Plant Cell Physiol.* 50:756–772. doi:10.1093/pcp/pcp033
- Saha, R., P.F. Suthers, and C.D. Maranas. 2011. Zea mays iRS1563: A comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE* 6:e21784. doi:10.1371/journal.pone.0021784
- Schaeffer, M.L., L.C. Harper, J.M. Gardiner, C.M. Andorf, D.A. Campbell, E.K. Cannon, T.Z. Sen, and C.J. Lawrence. 2011a. MaizeGDB: Curation and outreach go hand-in-hand. *Database* 2011:Bar022.
- Schaeffer, M., R. Walls, and L. Cooper. 2011b. Initial meeting – Aug 2011. Plant Ontology Consortium Project, Corvallis, OR. [http://wiki.plantontology.org/index.php/Submission\\_of\\_Association\\_files\\_for\\_the\\_Kaeppeler\\_gene\\_expression\\_data\\_from\\_MaizeGDB-Aug\\_2011](http://wiki.plantontology.org/index.php/Submission_of_Association_files_for_the_Kaeppeler_gene_expression_data_from_MaizeGDB-Aug_2011) (accessed 11 Apr. 2012).
- Schnable, J.C., and M. Freeling. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS ONE* 6:E17855. doi:10.1371/journal.pone.0017855
- Schnable, P.S., D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115. doi:10.1126/science.1178534
- Schneider, M., M. Tognolli, and A. Bairoch. 2004. The Swiss-Prot protein knowledgebase and ExPASy: Providing the plant community with high quality proteomic data and tools. *Plant Physiol. Biochem.* 42:1013–1021. doi:10.1016/j.plaphy.2004.10.009
- Sekhon, R.S., H. Lin, K.L. Childs, C.N. Hansey, C.R. Buell, N. de Leon, and S.M. Kaeppeler. 2011. Genome-wide atlas of transcription during maize development. *Plant J.* 66:553–563. doi:10.1111/j.1365-313X.2011.04527.x
- Sheen, J.Y., and L. Bogorad. 1987. Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. *J. Biol. Chem.* 262:11726–11730.
- Slack, C.R., M.D. Hatch, and D.J. Goodchild. 1969. Distribution of enzymes in mesophyll and parenchyma-sheath chloroplasts of maize leaves in relation to the C4-dicarboxylic acid pathway of photosynthesis. *Biochem. J.* 114:489–498.
- Slater, G.S., and E. Birney. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf.* 6:31. doi:10.1186/1471-2105-6-31
- Small, I., N. Peeters, F. Legeai, and C. Lurin. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4:1581–1590. doi:10.1002/pmic.200300776
- Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420:190–193. doi:10.1038/nature01166
- Strable, J., and M.J. Scanlon. 2009. Maize (*Zea mays*): A model organism for basic and applied research in plant biology. *Cold Spring Harb. Protoc.* 2009(10):pdb.emo132. doi:10.1101/pdb.emo132
- Tsesmetzis, N., M. Couchman, J. Higgins, A. Smith, J.H. Doonan, G.J. Seifert, E.E. Schmidt, I. Vastrik, E. Birney, G. Wu, et al. 2008. Arabidopsis reactome: A foundation knowledgebase for plant systems biology. *Plant Cell* 20:1426–1436. doi:10.1105/tpc.108.057976
- UniProt-GOA. 2011. Ensembl Compara. UniProt-GOA. European Bioinformatics Institute, Hinxton, Cambridge, UK. [http://www.ebi.ac.uk/GOA/compara\\_go\\_annotations.html](http://www.ebi.ac.uk/GOA/compara_go_annotations.html) (accessed 15 Mar. 2011).
- Urbanczyk-Wochniak, E., and L.W. Sumner. 2007. MedicCyc: A biochemical pathway database for *Medicago truncatula*. *Bioinformatics* 23:1418–1423. doi:10.1093/bioinformatics/btm040
- Vastrik, I., P. D'Eustachio, E. Schmidt, G. Gopinath, D. Croft, B. de Bono, M. Gillespie, B. Jassal, S. Lewis, L. Matthews, et al. 2007. Reactome: A knowledge base of biologic pathways and processes. *Genome Biol.* 8:R39. doi:10.1186/gb-2007-8-3-r39
- Vermerris, W., K.J. Thompson, and L.M. McIntyre. 2002. The maize Brown midrib1 locus affects cell wall composition and plant development in a dose-dependent manner. *Heredity* 88:450–457. doi:10.1038/sj.hdy.6800078
- Vignols, F., J. Rigau, M.A. Torres, M. Capellades, and P. Puigdomenech. 1995. The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. *Plant Cell* 7:407–416.
- Vilella, A.J., J. Severin, A. Ureta-Vidal, L. Heng, R. Durbin, and E. Birney. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19:327–335. doi:10.1101/gr.073585.107



- Walls, R.L., B. Athreya, L. Cooper, J. Elser, M.A. Gandolfo, P. Jaiswal, C.J. Mungall, J. Preece, S. Rensing, B. Smith, et al. 2012. Ontologies as integrative tools for plant science. *Am. J. Bot.* 99:1263–1275. doi:10.3732/ajb.1200222
- Westerlund, I., G. Von Heijne, and O. Emanuelsson. 2003. LumenP – A neural network predictor for protein localization in the thylakoid lumen. *Protein Sci.* 12:2360–2366. doi:10.1110/ps.0306003
- Wurtzel, E.T., A. Cuttriss, and R. Vallabhaneni. 2012. Maize provitamin a carotenoids, current resources, and future metabolic engineering challenges. *Front. Plant Sci.* 3:29. doi:10.3389/fpls.2012.00029
- Youens-Clark, K., E. Buckler, T. Casstevens, C. Chen, G. Declerck, P. Derwent, P. Dharmawardhana, P. Jaiswal, P. Kersey, A.S. Karthikeyan, et al. 2011. Gramene database in 2010: Updates and extensions. *Nucleic Acids Res.* 39:D1085–D1094. doi:10.1093/nar/gkq1148
- Zeier, J., K. Ruel, U. Rysler, and L. Schreiber. 1999. Chemical analysis and immunolocalisation of lignin and suberin in endodermal and hypodermal/rhizodermal cell walls of developing maize (*Zea mays* L.) primary roots. *Planta* 209:1–12.
- Zelitch, I. 1973. Alternate pathways of glycolate synthesis in tobacco and maize leaves in relation to rates of photorespiration. *Plant Physiol.* 51:299–305. doi:10.1104/pp.51.2.299
- Zhang, P., K. Dreher, A. Karthikeyan, A. Chi, A. Pujar, R. Caspi, P. Karp, V. Kirkup, M. Latendresse, C. Lee, 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.* 153:1479–1491. doi:10.1104/pp.110.157396
- Zhang, P., H. Foerster, C.P. Tissier, L. Mueller, S. Paley, P.D. Karp, and S.Y. Rhee. 2005. MetaCyc and AraCyc. Metabolic pathway databases for plant research. *Plant Physiol.* 138:27–37. doi:10.1104/pp.105.060376
- Zybailov, B., H. Rutschow, G. Friso, A. Rudella, O. Emanuelsson, Q. Sun, and K.J. van Wijk. 2008. Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS ONE* 3:E1994. doi:10.1371/journal.pone.0001994