The Plant Genome: Posted 23 Jan. 2013; doi: 10.3835/plantgenome2012.09.0025

Maize Metabolic Network Construction and Transcriptome Analysis

Marcela K Monaco[1], Taner Z. Sen[2,3], Palitha D. Dharmawardhana[4], Liya Ren[1], Mary Schaeffer[2,5], Sushma Naithani[6], Vindhya Amarasinghe[4], Jim Thomason[1], Lisa Harper[2,7], Jack Gardiner[3,8], Ethalinda K.S. Cannon[3], Carolyn J. Lawrence[2,3], Doreen Ware[1,2], and Pankaj Jaiswal[4]*

[1] Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724

[2] United States Department of Agriculture - Agricultural Research Service

[3] Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA 50011; [4] Department of Botany and Plant Pathology, 3082 Cordley Hall, Oregon State University, Corvallis, OR 97331; [5] Division of Plant Sciences, Department of Agronomy, University of Missouri, Columbia, MO 65211; [6] Department of Horticulture, 4017 ALS Bldg., Oregon State University, Corvallis, OR 97331; [7] Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [8] School of Plant Sciences, University of Arizona, Tucson, AZ 85721

The gene expression data analysis was performed on the published and publicly available NCBI GEO record GSE27004

*Corresponding author's e-mail: jaiswalp@science.oregonstate.edu

Received:

Journal: The Plant Genome

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* ssp. *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 v2.0.2 is a collection of 391 maize pathways involving 8,889 enzyme mapped to 2,110 reactions and 1,468 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 1,062 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 and Arabidopsis 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.

INTRODUCTION

Maize (*Zea mays* ssp. *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, 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 variety/cultivar B73 (Schnable et al., 2009) opened avenues to explore the interaction of genes, 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 function(s) with respect to subcellular localization and site(s) of mechanistic function of its products, namely transcript(s) and protein(s) 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 extra-cellular, 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 carried out 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, 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; Mueller, 2005; Mueller et al., 2005; Zhang et al., 2005; Urbanczyk-Wochniak and Sumner, 2007) and monocots (Jaiswal et al., 2006; dharmawardhana et al., 2012). 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* (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 non-photosynthetic reactions, and focused on plants such as maize, sugarcane, and sorghum, and assessed flux distributions in leaf mesophyll and bundle sheath cells during C_4 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 1,500 genes mapped to 1,985 reactions and containing 1,825 metabolites (Saha et al., 2011). However, evidence from the literature was found for only 42% of the 1,985 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 seeds associated with grain yield and metabolic fluxes under oxygen stress. A similar network study in rapeseed *Brassica napus* (Pilalis et al., 2011) focused on seed development and metabolism based on annotations from AraCyc.

Gramene (Youens-Clark et al., 2011) and MaizeGDB (Schaeffer et al., 2011) 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 where enzymes serve as catalysts, and (ii) transport reactions where metabolites are transported from one compartment to another within the cell or between the cells via transporter protein(s) carriers, and channels. We created MaizeCyc v2.0.2 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. In order to illustrate this example, we report the analysis of microarray-based genome-wide 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 v2 (Schnable et al., 2009) using the high-quality filtered gene set (FGS) representing 39,656 gene models downloaded from the MaizeSequence.org FTP site (http://ftp.maizesequence.org/).

MaizeCyc Gene Product Annotation

In order to create MaizeCyc, we used the Pathways Tools software (Karp et al., 2002). Annotation was formatted according to the PathoLogic input file format. As described in Supplementary Table ST1, 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, (1) common gene names from the list of named maize genes (http://maizesequence.org/info/docs/namedgenes.html) provided by Maize Core DB, (2) common gene names for the gene models assigned by CoGePedia (Schnable and Freeling, 2011) and stored in the MaizeGDB's central database, (3) additional names and gene symbols for a given gene model were stored as SYNONYM attribute and (4) for the gene models with no known name/symbol, the canonical transcript ID 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 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 thus PRODUCT-TYPE code "P" (for protein or hypothetical ORF) was assigned. EC numbers were assigned based on the annotations extracted from (1) mapped UniProt entries, (2) MaizeGDB curated pathways, and (3) 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. GO 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* and rice (UniProt-GOA, 2011). GO 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 (v15.5) database. MetaCyc was used as a default reference databases for finding the best matches for reactions and pathways.

Quality Control (QC) Filters

QC tools included in the Pathway Tools suite were executed to check for consistency and annotations. (1) 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 non-functional. 88 such genes were excluded after confirming their protein description in UniProt and GO assignments. (2) UniProt, MaizeGDB, and other manually curated databases (RiceCyc ver3.1, AraCyc ver 8, and PlantCyc ver 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 at http://enzyme.expasy.org/ (July 2011). (3). 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/GO match with the best match added as the main label. 4) Pathways specific to bacteria, fungi and animals were removed. (5) 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 carried out after each cycle of manual checks involving acceptance/edits/rejections.

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 (Figure 1). Installation requires users to obtain licensed Pathway Tools software available from the 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 SBML (Hucka et al., 2003) formats. These files are compatible

for viewing networks using software like Cytoscape (Killcoyne et al., 2009). The expression dataset used here is available in Supplementary Table ST3 and also over the web at http://ftp.maizegdb.org/MaizeGDB/FTP/MaizeCyc_manuscript_files/

Gene Expression Data Analysis

Microarray Dataset

The 80,301 NimbleGen microarray probesets (Sekhon et al., 2011) were mapped in this report to the B73 RefGen_v2 Working Gene Set (which includes the filtered gene set used by MaizeCyc) by aligning each individual probe to the pseudomolecules and requiring a 100% coverage match to a gene model. Up to 2 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 probeset mappings were shared with the PLEXdb database (Dash et al., 2011) where the data were re-normalized (experiment id ZM37).

Gene Expression Data Analysis

The maize gene expression atlas dataset was developed for 60 diverse tissues previously by Sekhon et al. (2011) and annotated (Schaeffer et al., 2011) 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 5 major organ types (1) V1_pooled

leaves (L), (2) V1_primary root (R), (3) R1_anthers (A), (4) 16DAP endosperm (D) and (5) 16DAP_embryo (E) (Supplementary Table ST2). These tissue types were reported to have high replicate correlations (correlation coefficients >0.95, P<0.001, as reported in the Supplementary table S2 of the Sekhon et al. (2011)). The re-normalized gene expression data from the selected five source tissue samples downloaded from PLEXdb. The dataset was analyzed to remove low/non-expressed genes with an expression threshold cutoff of 5 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 4 fold or more in only one of the 5 tissue types that we analyzed. The resulting 10,057 genes showing tissue specific upregulation were used for further analysis (supplementary table ST3). Hierarchical clustering of mean centered expression patterns based on Pearson correlation was performed using GeneSpring 7.4 (Agilent Technologies, USA). 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 and software scripts developed in-house to identify pathways/genes/reactions dominant in specific tissue types (Supplementary tables ST4, ST5, ST6, ST7). The gene loci based annotation using the PO was submitted to the PO database (Schaeffer et al., 2011).

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 super-pathway 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 8,889 enzymes (about 22% of the filtered gene set), and 291 transporter proteins, mapping to 2,110 enzymatic and 68 transport reactions respectively, in addition to 1,468 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 (1) the enzyme in question had not been previously identified from a plant/other source, and/or (2) though it may be a known enzyme, there was no entry in the reference library maintained by Pathway Tools and that annotation is awaiting secondary curation.

MaizeCyc: A Resource for Biologists

Users can access instances of MaizeCyc via both the MaizeGDB (http://maizecyc.maizegdb.org and Figures 1 and 2) and Gramene (http://pathways.gramene.org) websites. We provide a short MaizeCyc tutorial (Supplementary File SF1) aimed at new users on how to browse, search and use tools

such as the OMICS viewer (Supplementary Figure S1) to overlay gene, protein, and metabolic expression datasets. Excellent tutorial webinars are also available from the BioCyc website (http://biocyc.org/webinar.shtml).

Differential Expression of Metabolic Pathway Genes

In order 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) dataset focusing on metabolic pathways. The analysis was performed on five selected tissue samples: (a) pooled leaves from the Vegetative 1 (V1) stage (where only one leaf is fully expanded), (b) primary root at the V1 stage, (c) anthers at the Reproductive stage R1 (silks emerge from the husk), (d) embryo from a developing seed at 16 days after pollination (16DAP), and (e) endosperm, also at 16DAP. Of the 10,057 genes identified as upregulated in these five organs, 7,557 genes were listed in the FGS used to reconstruct MaizeCyc. Of these, only 1,957 genes mapped to known reactions in MaizeCyc. (Figure 3, Supplementary Tables ST3 and ST4). In the database, not all reactions are part of a pathway: therefore, out of a total 1,957 differentially expressed MaizeCyc gene entries we found 1,062 genes mapped to 513 unique reactions associated with 308 pathways (Supplementary Table ST5). Among the 1,062 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 and Supplementary table ST6).

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 Supplementary Table ST6. Of the 310 unique and upregulated 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 (Figure 3b, and Supplementary Table ST6). Among the commonly upregulated pathways, biosynthetic pathways for cellulose, flavonol, flavonoid, suberin (Supplementary Figure S2), beta-caryophyllene, and the Calvin-Benson-Bassham cycle were upregulated in all organs except the endosperm. The biosynthesis pathways of chlorophyllide *a* (figure-4) and beta-carotene, zeaxanthin, xanthophylls (Supplementary Figure S3) were upregulated in leaf. The other catorenoids-like delta- and epsilon-carotene biosynthesis genes were upregulated only in anther samples. The geranyl-geranyl-diphosphate biosynthesis and the trans-lycopene biosynthesis pathways that provide precursor metabolites for anthocyanin and carotenoids biosynthesis, respectively, were upregulated in leaf samples, whereas the anthocyanin biosynthesis pathway (Supplementary Figure S4) was upregulated in anther, leaf and primary root samples. Consistent with the findings reported by Sekhon et al (2011), the lignin (Figure 4) and suberin (Supplementary Figure S2) biosynthesis pathway(s) 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 over-expression of suberin pathway genes, we also observed upregulation in

non-root samples for genes from the phenylpropanoid related suberin pathway encoding *trans*-feruloyl-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) (Supplementary Figure S2). Xylan and xyloglucan biosynthesis genes were expressed mainly in the embryo, while the expression of genes mapping to indole3-acetic acid (IAA) conjugate biosynthesis, and fatty acid activation, appeared to be limited to embryo and endosperm tissues (Supplementary Table ST7). A detailed discussion of differentially expressed genes associated with the processes of photosynthesis and lignification in maize can be found in Supplementary File SF2.

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 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 datasets 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, BioCyc 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) (Supplementary Figure S3) and C_4 photosynthesis (Slack et al., 1969; Sheen and Bogorad, 1987; Nomura et al., 2000) pathways.

The current version (v2.0.2) 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, and rice. An advantage of developing such 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 (Figure 3a) map to 310 (Figure 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 (Figure 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 datasets 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 representations of maize metabolic pathways. Subsequent releases of MaizeCyc will likely include enrichment of the networks by including, e.g., (1) annotations captured in

the published metabolic networks for maize (Dal'Molin et al., 2010; Saha et al., 2011), (2) subcellular locations identified by computational methods (Westerlund et al., 2003; Small et al., 2004; Emanuelsson et al., 2007), (3) findings from proteomics experiments (Zybailov et al., 2008; Majeran et al., 2011), (4) references to the probeset identifiers from various expression platforms developed for maize, (5) references to the gene coordinates and 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.

ACKNOWLEDGEMENTS

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 webserver 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), 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:

Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, et al. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 40: D742-753

Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, Preece J, Athreya B, Mungall CJ, Rensing S, et al. (2012) The Plant Ontology As A Tool For Comparative Plant Anatomy And Genomic Analyses. Plant and Cell Physiology

Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) C4GEM, a genome-scale metabolic model to study C4 plant metabolism. Plant Physiol 154: 1871-1885

Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA (2011) PLEXdb: gene expression resources for plants and plant pathogens. Nucleic Acids Res 40: D1194-1201

Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'Eustachio P, Schaefer C, Luciano J, et al. (2010) The BioPAX community standard for pathway data sharing. Nat Biotechnol 28: 935-942

dharmawardhana P, Ren L, Amarasinghe V, Monaco M, Thomason J, Ravenscroft D, McCouch SR, Ware D, Jaiswal P (2012) Genome Scale metabolic network for rice and analysis of tryptophan and derivative biosynthesis regulation during biotic stress. Rice (submitted)

Edwards GE, Furbank RT, Hatch MD, Osmond CB (2001) What does it take to be C4? Lessons from the evolution of C4 photosynthesis. Plant Physiol 125: 46-49

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2: 953-971

FAOSTAT (2011) Food and Agricultural commodities production. FAO-FAOSTAT. http://faostat.fao.org/site/339/default.aspx

Ferrer M (2009) The microbial reactome. Microb Biotechnol 2: 133-135

Friso G, Majeran W, Huang M, Sun Q, van Wijk KJ (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

Gaut BS, Le Thierry d'Ennequin M, Peek AS, Sawkins MC (2000) Maize as a model for the evolution of plant nuclear genomes. Proc Natl Acad Sci U S A 97: 7008-7015

Grafahrend-Belau E, Schreiber F, Koschutzki D, Junker BH (2009) Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism. Plant Physiol 149: 585-598

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al. (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32: D258-261

Hatch MD (1978) Regulation of enzymes in C4 photosynthesis. Curr Top Cell Regul 14: 1-27

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19: 524-531

Ilic K, Kellogg EA, Jaiswal P, Zapata F, Stevens PF, Vincent LP, Avraham S, Reiser L, Pujar A, Sachs MM, et al. (2007) The plant structure ontology, a unified vocabulary of anatomy and morphology of a flowering plant. Plant Physiol. 143: 587-599

Jaiswal P, Avraham S, Ilic K, Kellogg EA, McCouch S, Pujar A, Reiser L, Rhee SY, Sachs MM, Schaeffer M, et al. (2005) Plant Ontology (PO): A Controlled Vocabulary of Plant Structures and Growth Stages. Comp Funct Genomics 6: 388-397

Jaiswal P, Ni J, Yap I, Ware D, Spooner W, Youens-Clark K, Ren L, Liang C, Zhao W, Ratnapu K, et al. (2006) Gramene: a bird's eye view of cereal genomes. Nucleic Acids Res 34: D717-723

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2011) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40: D109-114

Karp PD, Paley S, Romero P (2002) The Pathway Tools software. Bioinformatics 18 Suppl 1: S225-232

Kersey PJ, Lawson D, Birney E, Derwent PS, Haimel M, Herrero J, Keenan S, Kerhornou A, Koscielny G, Kahari A, et al. (2009) Ensembl Genomes: extending Ensembl across the taxonomic space. Nucleic Acids Res 38: D563-569

Killcoyne S, Carter GW, Smith J, Boyle J (2009) Cytoscape: a community-based framework for network modeling. Methods Mol Biol 563: 219-239

Latendresse M, Paley S, Karp PD (2011) Browsing metabolic and regulatory networks with BioCyc. Methods Mol Biol 804: 197-216

Majeran W, Friso G, Asakura Y, Qu X, Huang M, Ponnala L, Watkins KP, Barkan A, van Wijk KJ (2011) Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. Plant Physiol 158: 156-189

Mueller L (2005) SOL Project Sequencing and Bioinformatics Standards and Guidelines. http://www.sgn.cornell.edu/documents/solanaceae-project/docs/tomato-standards.pdf

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, et al. (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138: 1310-1317

Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol. 132: 453-460.

Nomura M, Sentoku N, Nishimura A, Lin JH, Honda C, Taniguchi M, Ishida Y, Ohta S, Komari T, Miyao-Tokutomi M, 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

Papp B, Pal C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429: 661-664

Pilalis E, Chatziioannou A, Thomasset B, Kolisis F (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

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol 151: 1570-1581

Pujar A, Jaiswal P, Kellogg EA, Ilic K, Vincent L, Avraham S, Stevens P, Zapata F, Reiser L, Rhee SY, et al. (2006) Whole-plant growth stage ontology for angiosperms and its application in plant biology. Plant Physiol 142: 414-428

Rabinowicz PD, Bennetzen JL (2006) The maize genome as a model for efficient sequence analysis of large plant genomes. Curr Opin Plant Biol 9: 149-156

Rawsthorne S (1992) Towards an understanding of C3-C4 photosynthesis. Essays Biochem 27: 135-146

Sage TL, Sage RF (2009) The functional anatomy of rice leaves: implications for refixation of photorespiratory CO2 and efforts to engineer C4 photosynthesis into rice. Plant Cell Physiol 50: 756-772

Saha R, Suthers PF, Maranas CD (2011) Zea mays iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS ONE 6: e21784

Schaeffer M, Walls R, Cooper L (2011) Submission of Association files for the Kaeppler gene expression data from MaizeGDB- Aug 2011. Plant Ontology Consortium Project, http://wiki.plantontology.org/index.php/Submission_of_Association_files_for_the_Kaeppler_gene_expression_data_from_MaizeGDB-_Aug_2011

Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EK, Sen TZ, Lawrence CJ (2011) MaizeGDB: curation and outreach go hand-in-hand. Database (Oxford) 2011: bar022

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6: e17855

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112-1115

Schneider M, Tognolli M, Bairoch A (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

Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM (2011) Genome-wide atlas of transcription during maize development. Plant J 66: 553-563

Sheen JY, Bogorad L (1987) Differential expression of C4 pathway genes in mesophyll and bundle sheath cells of greening maize leaves. J Biol Chem 262: 11726-11730

Slack CR, Hatch MD, Goodchild DJ (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 GS, Birney E (2005) Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics 6: 31

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4: 1581-1590.

Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. Nature 420: 190-193

Strable J, Scanlon MJ (2009) Maize (Zea mays): a model organism for basic and applied research in plant biology. Cold Spring Harb Protoc 2009: pdb emo132

Tsesmetzis N, Couchman M, Higgins J, Smith A, Doonan JH, Seifert GJ, Schmidt EE, Vastrik I, Birney E, Wu G, et al. (2008) Arabidopsis reactome: a foundation knowledgebase for plant systems biology. Plant Cell 20: 1426-1436

UniProt-GOA (2011) UniProt-GOA - Compara Orthology Electronic Annotation Method. UniProt-GOA. http://www.ebi.ac.uk/GOA/compara_go_annotations.html

Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database for Medicago truncatula. Bioinformatics 23: 1418-1423

Vastrik I, D'Eustachio P, Schmidt E, Gopinath G, Croft D, de Bono B, Gillespie M, Jassal B, Lewis S, Matthews L, et al. (2007) Reactome: a knowledge base of biologic pathways and processes. Genome Biol 8: R39

Vermerris W, Thompson KJ, McIntyre LM (2002) The maize Brown midrib1 locus affects cell wall composition and plant development in a dose-dependent manner. Heredity (Edinb) 88: 450-457

Vignols F, Rigau J, Torres MA, Capellades M, Puigdomenech P (1995) The brown midrib3 (bm3) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase. Plant Cell 7: 407-416

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res 19: 327-335

Walls RL, Athreya B, Cooper L, Elser J, Gandolfo MA, Jaiswal P, Mungall CJ, Preece J, Rensing S, Smith B, et al. (2012) Ontologies as integrative tools for plant science. Am J Bot 99: 1263-1275

Westerlund I, Von Heijne G, Emanuelsson O (2003) LumenP--a neural network predictor for protein localization in the thylakoid lumen. Protein Sci 12: 2360-2366

Wurtzel ET, Cuttriss A, Vallabhaneni R (2012) Maize provitamin a carotenoids, current resources, and future metabolic engineering challenges. Front Plant Sci 3: 29

Youens-Clark K, Buckler E, Casstevens T, Chen C, Declerck G, Derwent P, Dharmawardhana P, Jaiswal P, Kersey P, Karthikeyan AS, et al. (2011) Gramene database in 2010: updates and extensions. Nucleic Acids Res 39: D1085-1094

Zelitch I (1973) Alternate pathways of glycolate synthesis in tobacco and maize leaves in relation to rates of photorespiration. Plant Physiol 51: 299-305

Zhang P, Dreher K, Karthikeyan A, Chi A, Pujar A, Caspi R, Karp P, Kirkup V, Latendresse M, Lee C, 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

Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY (2005) MetaCyc and AraCyc. Metabolic pathway databases for plant research. Plant Physiol. 138: 27-37.

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE 3: e1994

Figures:

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) 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, RNA, and pseudogenes) in a tabular format for each chromosome (Step 5).

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, (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 (EC) numbers and (7) gene id of the enzymes involved in reactions, and metabolites as (8-9) reactants and products. (C) Reaction detail view. Click on the compound name/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 KEGG, BRENDA, etc.

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 5 tissue types. Mean centered expression levels are represented with green color denoting down regulation, and red color indicating upregulation. (B) Venn diagram of tissue-specific enrichment of metabolic pathways associated to the upregulated genes identified in Figure 1 and Supplementary Tables ST5 and ST6. As described by the authors (Sekhon et al., 2011), the RNA samples are from E=16DAP_embryo, D=16DAP_endosperm, R=V1_primary root, A=R1_anther and L=L1_pooled leaves. The Plant Ontology annotations of tissue samples are provided in Supplementary Table ST3.

Figure 4: OMICs viewer analysis of the overexpressed genes from five different tissue samples. (i) Cellular overview of the pathways, reactions and transporters painted with the gene expression dataset from leaf as an example. Purple color represents downregulation, and red color, upregulation. The view was generated on a local desktop version of the database using Pathway Tools software 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 (FMN) 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 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 or symbol, and column headers represent expression data from E=embryo, D=endosperm, R=root, A=anther and L=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 RNA samples. Majority of genes show root-specific expression. The expression data is provided in the supplementary table ST3.

Supplementary Files:

Supplementary Figure S1

The Omics Viewer description. (1) the cellular overview (available in the online and local desktop version of the Pathway tools) (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.

Supplementary Figure S2

A pop-up view on the desktop version of the tool, with reaction, compound and enzyme (EC number) details of the suberin biosynthesis pathway painted with expression data from the five different RNA samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally-installed version of Pathway Tools), rows correspond to gene ID or symbol, and column headers represent expression data from E=embryo, D=endosperm, R=root, A=anther and L=leaf.

Supplementary Figure S3

A pop-up view of the desktop version of the pathway tool, with reaction, compound and enzyme (EC number) details of the carotenoid biosynthesis pathway painted with expression data from the five different RNA samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally-installed version of Pathway Tools), rows correspond to gene ID or symbol, and column headers represent expression data from E=embryo, D=endosperm, R=root, A=anther and L=leaf.

Supplementary Figure S4

A pop-up view of the desktop version of the pathway tool, with reaction, compound and enzyme (EC number) details of the anthocyanin biosynthesis pathway painted with expression data from the five different RNA samples. Majority of genes show root-specific expression. In the expression data blocks (available only in the locally-installed version of Pathway Tools), rows correspond to gene ID or symbol, and column headers represent expression data from E=embryo, D=endosperm, R=root, A=anther and L=leaf.

Supplementary File SF1 (PDF format)

A MaizeCyc Tutorial, featuring a section describing how to perform gene expression data analyses with the OMICs viewer.

Supplementary File SF2 (PDF format)

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

Supplementary Table ST1 (MS Excel format)

The attributes of the functional annotation file in the PathoLogic format and source database and database element used to create these attributes. Ensembl Core 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 v32 and Ensembl Core D for maize was used.

File: ST1_source_data_and_xref_provider_list_.xlsx

Supplementary Table ST2 (MS Excel format)

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., 2011).

File: ST2_maize_atlas_tissue_to_PO_mapping.xlsx

Supplementary Table ST3 (Tab delimited txt format)

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

File: ST3_maize_gene_expression_only_upregulated_genes_from_five_tissues.txt

Supplementary Table ST4 (MS Excel format)

List of all the tissue specific genes (10,057 genes) from maize based on the 4-fold cutoff.

File: ST4_tissue_specific_gene_lists_all_v1.xlsx

Supplementary Table ST5 (MS Excel format)

Complete list of expressed gene IDs mapped to MaizeCyc reaction id and the MaizeCyc pathway name.

File: ST5_filtered_genes_mapped_to_reactions_and_pathways.xlsx

Supplementary Table ST6 (MS Excel format)

List of pathways and corresponding tissue-/organ-specific cluster. Counts used for creating the Venn Diagram (Figure 3b).

File: ST6_organ_clusters_and_pathways_for_Venn.xlsx

Supplementary Table ST7 (MS Excel format)

Complete list of pathways and the associated organs. Includes information on (1) Pathway name, (2) # of expressed genes mapped to the pathway, (3) # of reactions to which expressed genes map in the pathway, (4) Total # of reactions in the pathway, and (5) Total # of genes mapped in the pathway (6) Organ-specific cluster.

File: ST7_individual_pathways_scored_for_organ.xlsx

Supplementary Table ST8 (MS Excel format)

Overall counts for coexpression networks and conserved gene modules in maize overlaid onto MaizeCyc annotated genes.

File: ST8_overall_counts_conserved_clusters_pathway_overlay.xlsx

Table-1

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

	Embryo	Endosperm	Root	Anther	Leaf
(A): # of genes upregulated in organ-specific cluster	1,680	1,503	2,247	2,211	2,416
(B) # of upregulated genes mapped to pathways	129	90	245	257	341
(C) # of reactions mapped to upregulated genes	149	112	152	180	285
(D) # of pathways mapped to upregulated genes	128	106	146	159	199
(E) Average # 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 # of genes upregulated per reaction where F=(B/C)	0.87	0.80	1.61	1.43	1.20

The Plant Genome: Posted 23 Jan. 2013; doi: 10.3835/plantgenome2012.09.0025

Figure-1

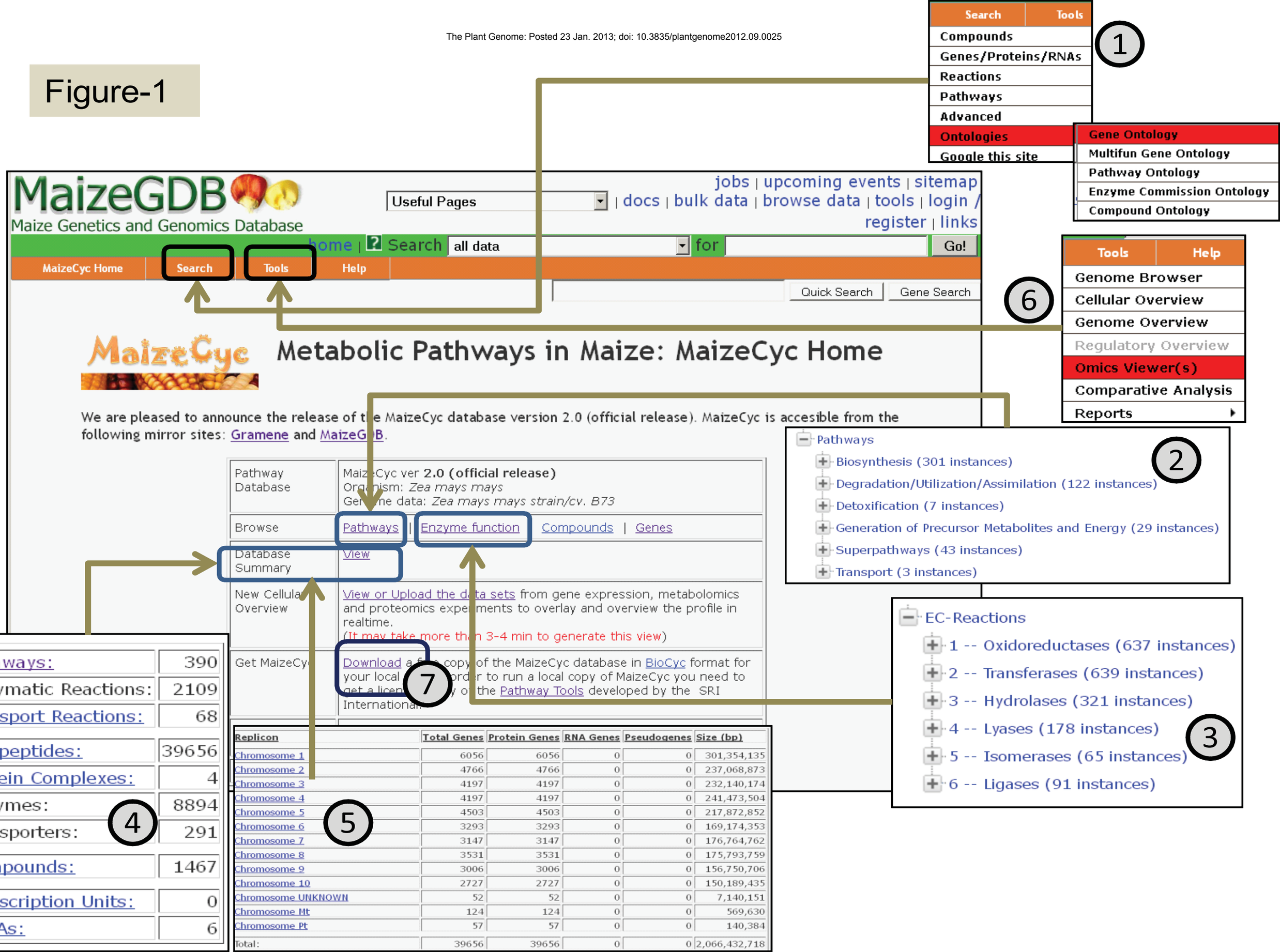


Figure-2

The Plant Genome: Posted 23 Jan. 2013; doi: 10.3835/plantgenome2012.09.0025

MaizeCyc Home | Search | Tools | Help

folate (1) NIL | Gene Search

Alternative searches:

- Full text search for **folate** on all pages in this database using Google

(A) Quick search results

Search Results for folate
using database *Zea mays mays*

Pathways (8) | Proteins (22) | Gene Ontology Terms (20) | Compounds (18) | Reactions (18) (2)

Pathways: Pathway pages contain: Depiction of metabolic pathway, of chromosomal locations of pathway genes, and of regulation of pathway genes.

- **folate** polyglutamylation
- **folate** transformations I
- **folate** transformations II (plants)
- formaldehyde oxidation V (tetrahydro**folate** pathway)
- superpathway of tetrahydro**folate** biosynthesis
- superpathway of tetrahydro**folate** biosynthesis and salvage
- tetrahydro**folate** biosynthesis
- tetrahydro**folate** salvage from 5,10-methenyltetrahydrofolate

(3)

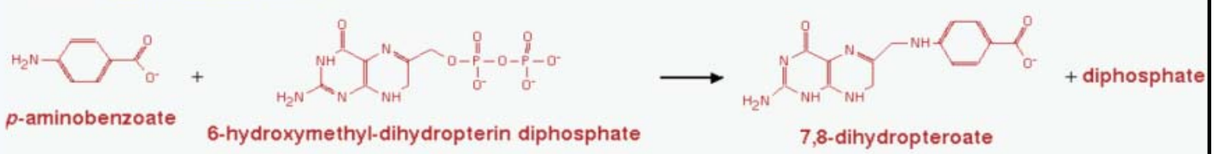
Zea mays mays Reaction: 2.5.1.15

Species Comparison

Superclasses: Reactions-Classified-By-Conversion-Type → Simple-Reactions → Chemical-Reactions → EC-Reactions → 2 -- Transferases → 2.5 -- Transferring alkyl or aryl groups, other than methyl groups → 2.5.1 -- Transferring alkyl or aryl groups, other than methyl groups
Reactions-Classified-By-Substrate → Small-Molecule-Reactions

Enzymes and Genes:
2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase : GRMZM2G095806_P01
ALG2-interacting protein X : GRMZM2G168855_P01
Dihydropteroate synthase : GRMZM2G082529_P01

In Pathway: tetrahydrofolate biosynthesis



The reaction direction shown, that is, A + B <==> C + D versus C + D <==> A + B, is in accordance with the Enzyme Commission system.

Enzyme Commission Primary Name for this Reaction: dihydropteroate synthase

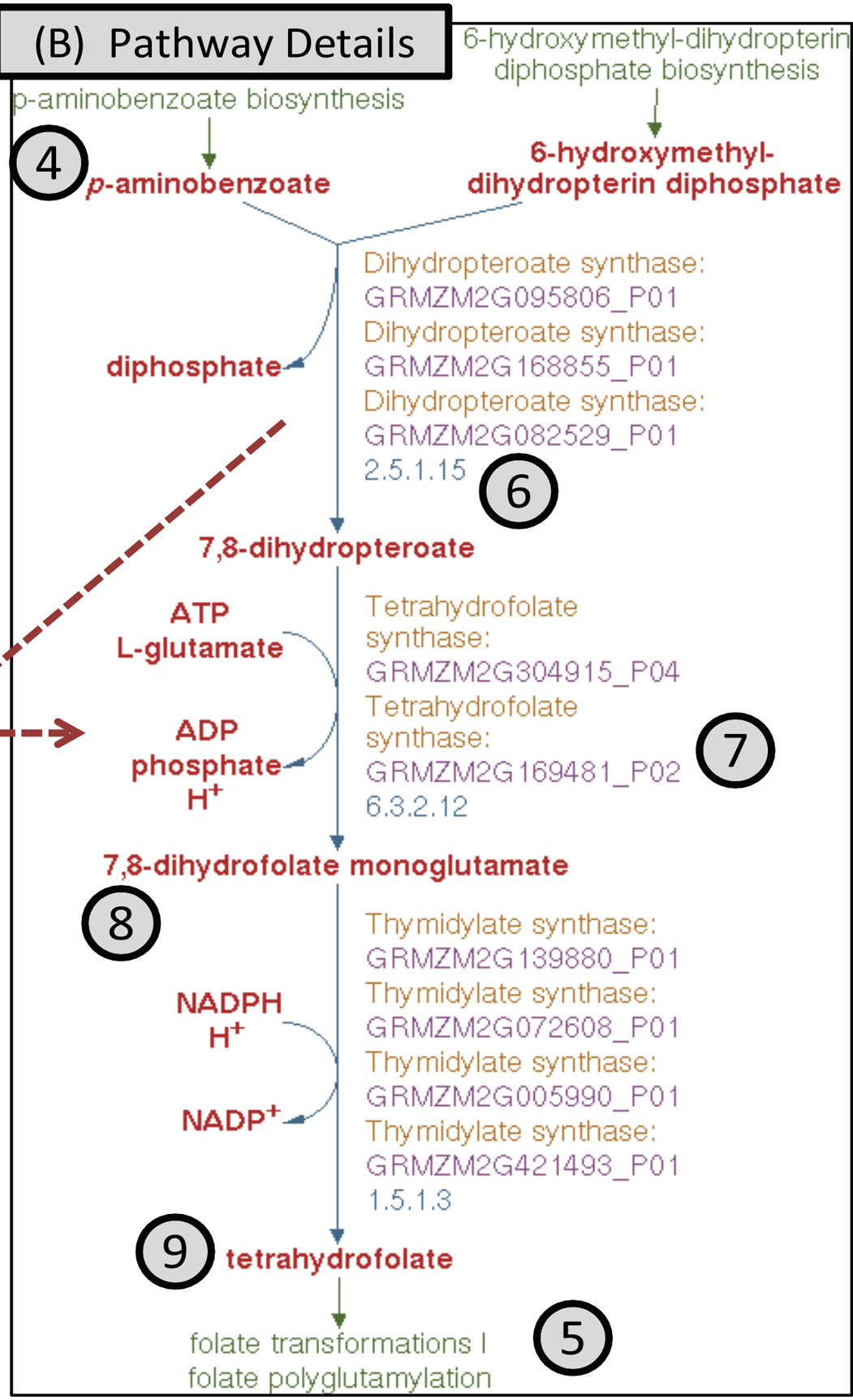
Enzyme Commission Synonyms for this Reaction: dihydropteroate pyrophosphorylase, DHPS, 7,8-dihydropteroate synthase, 7,8-dihydropteroate synthetase, 7,8-dihydropteroic acid synthetase, dihydropteroate synthetase, dihydropteroic synthetase, 2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine-diphosphate:4-aminobenzoate 2-amino-4-hydroxydihydropteridine-6-methenyltransferase

Gene-Reaction Schematic:

Unification Links: BRENDA:2.5.1.15 , ENZYME:2.5.1.15

(C) Reaction Details

(B) Pathway Details



The Plant Genome: Posted 23 Jan. 2013; doi: 10.3835/plantgenome2012.09.0025

Figure-3a

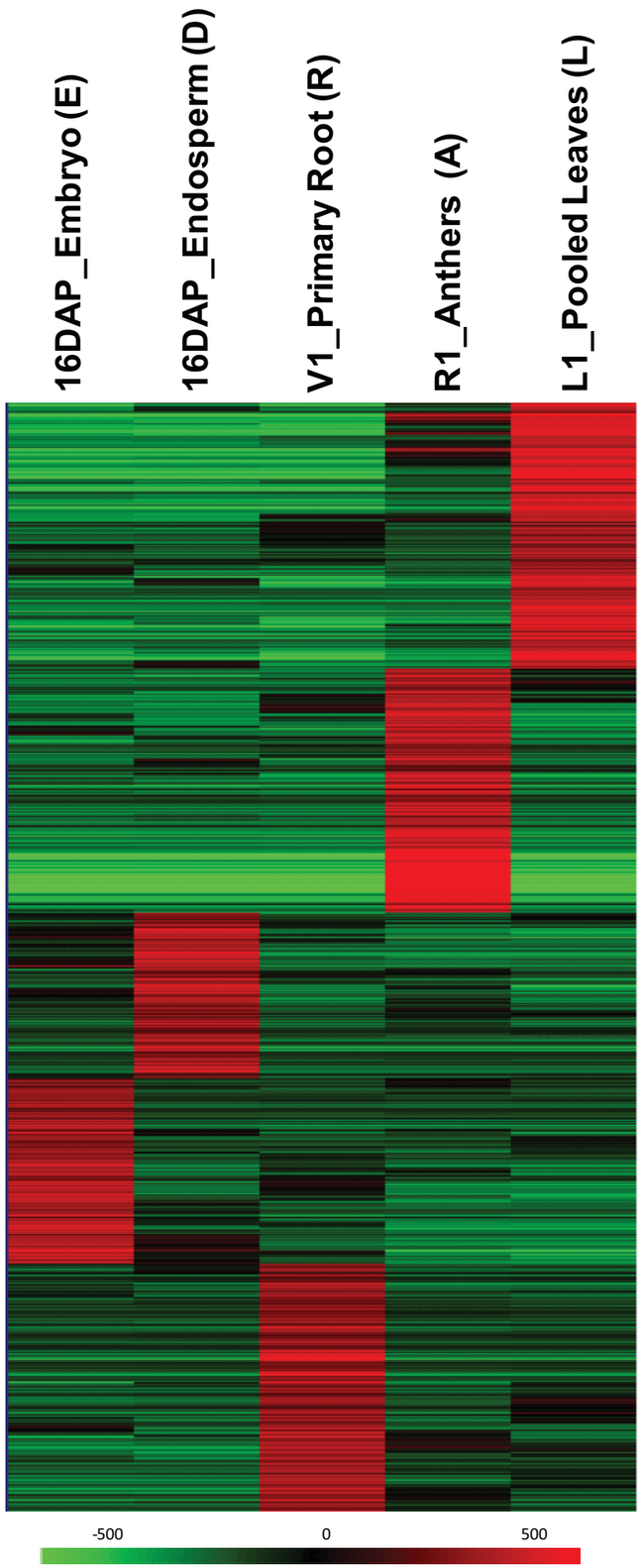
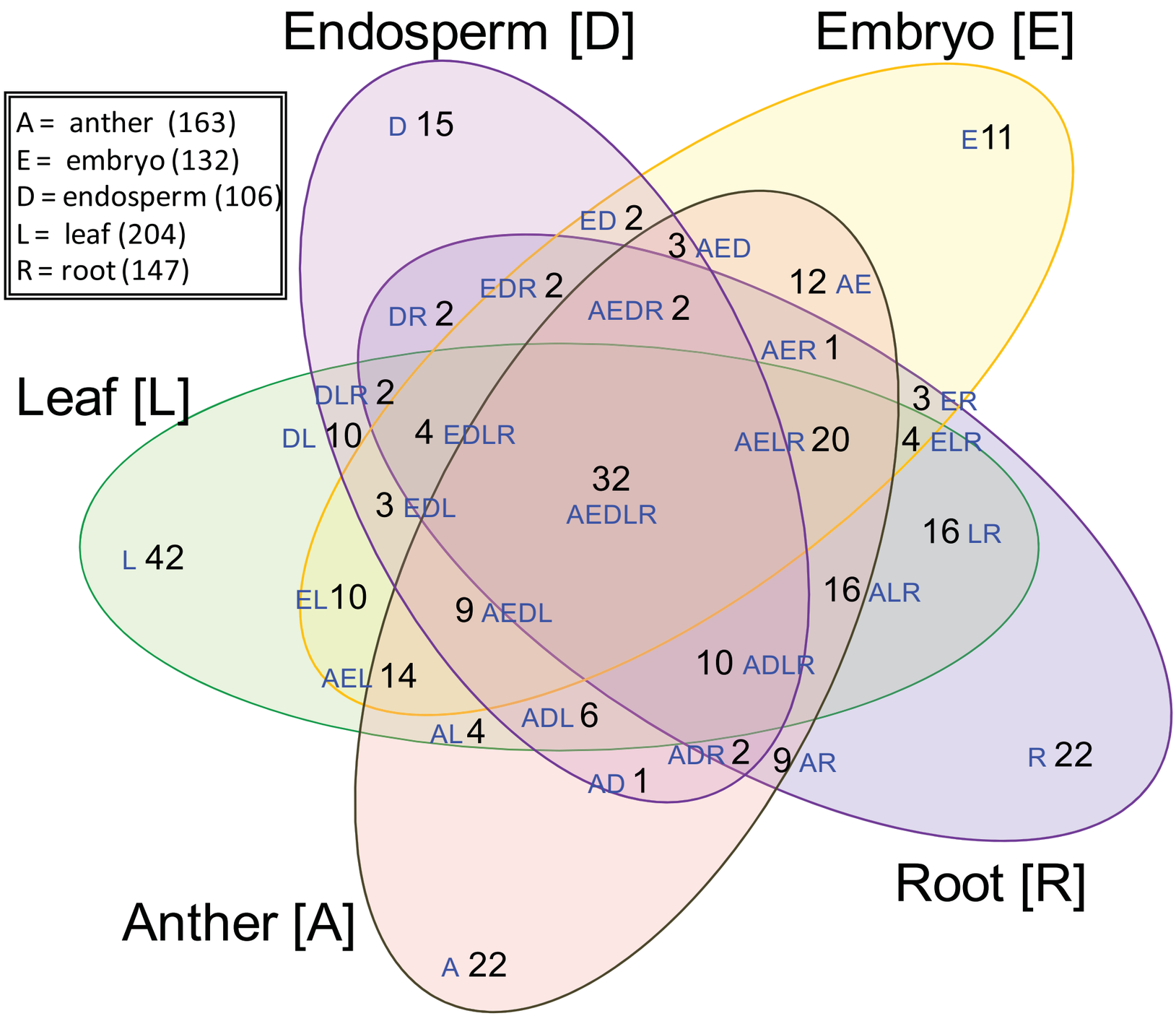
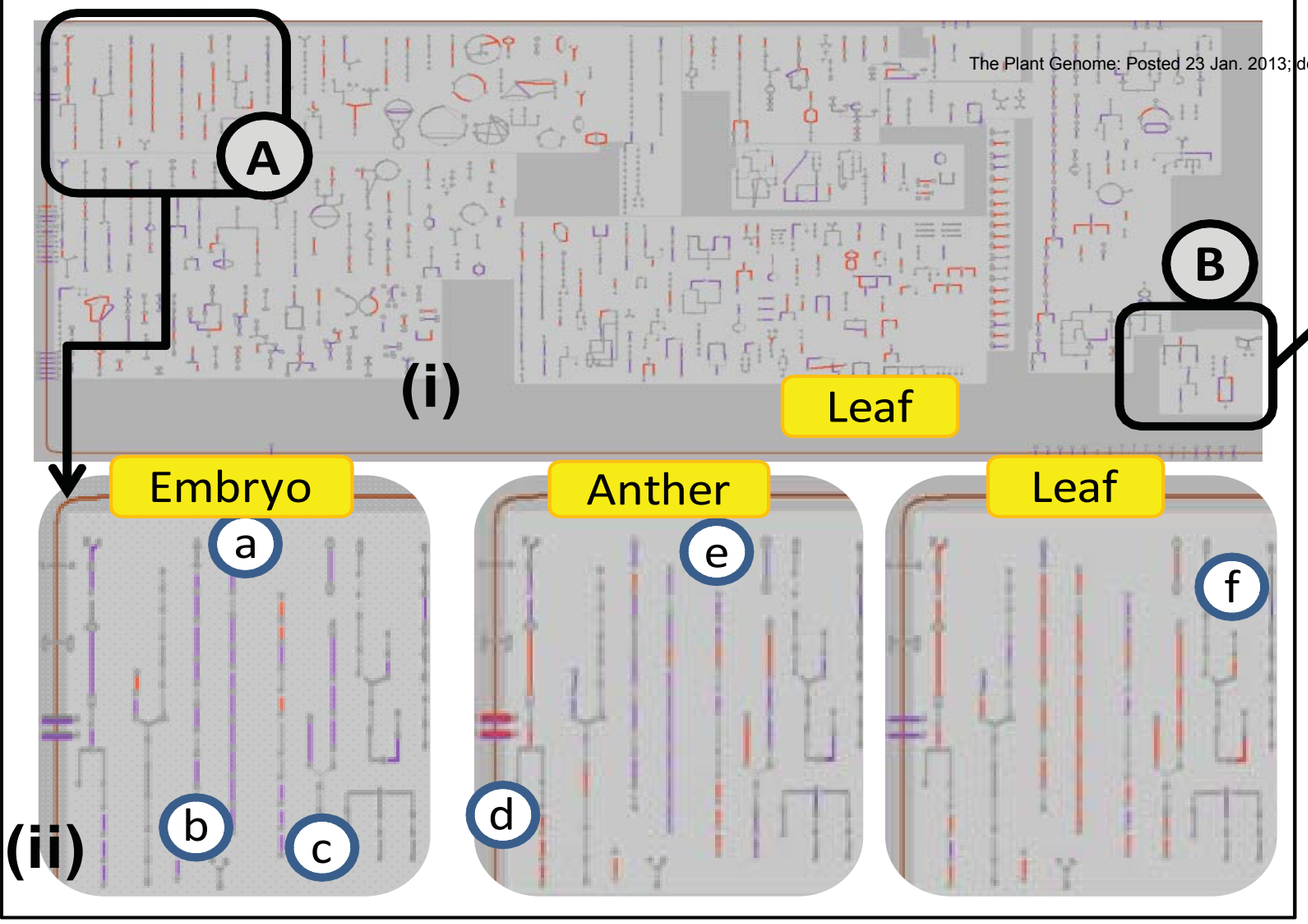
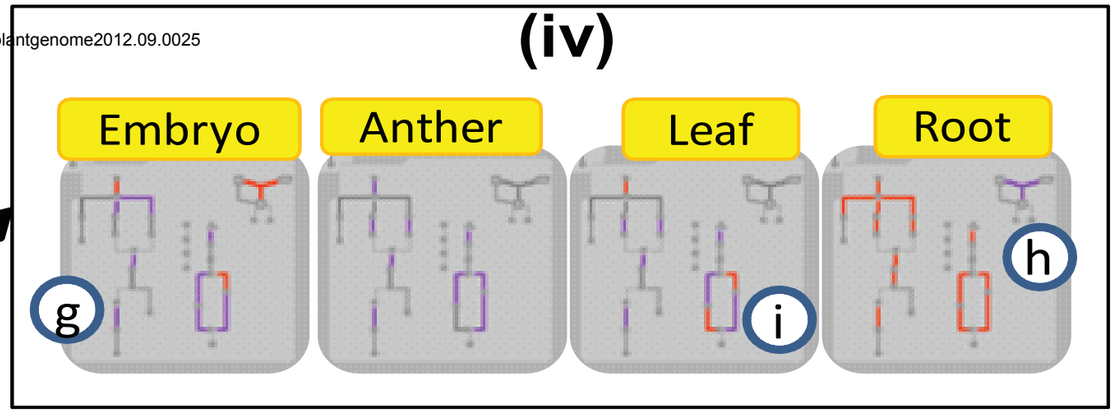


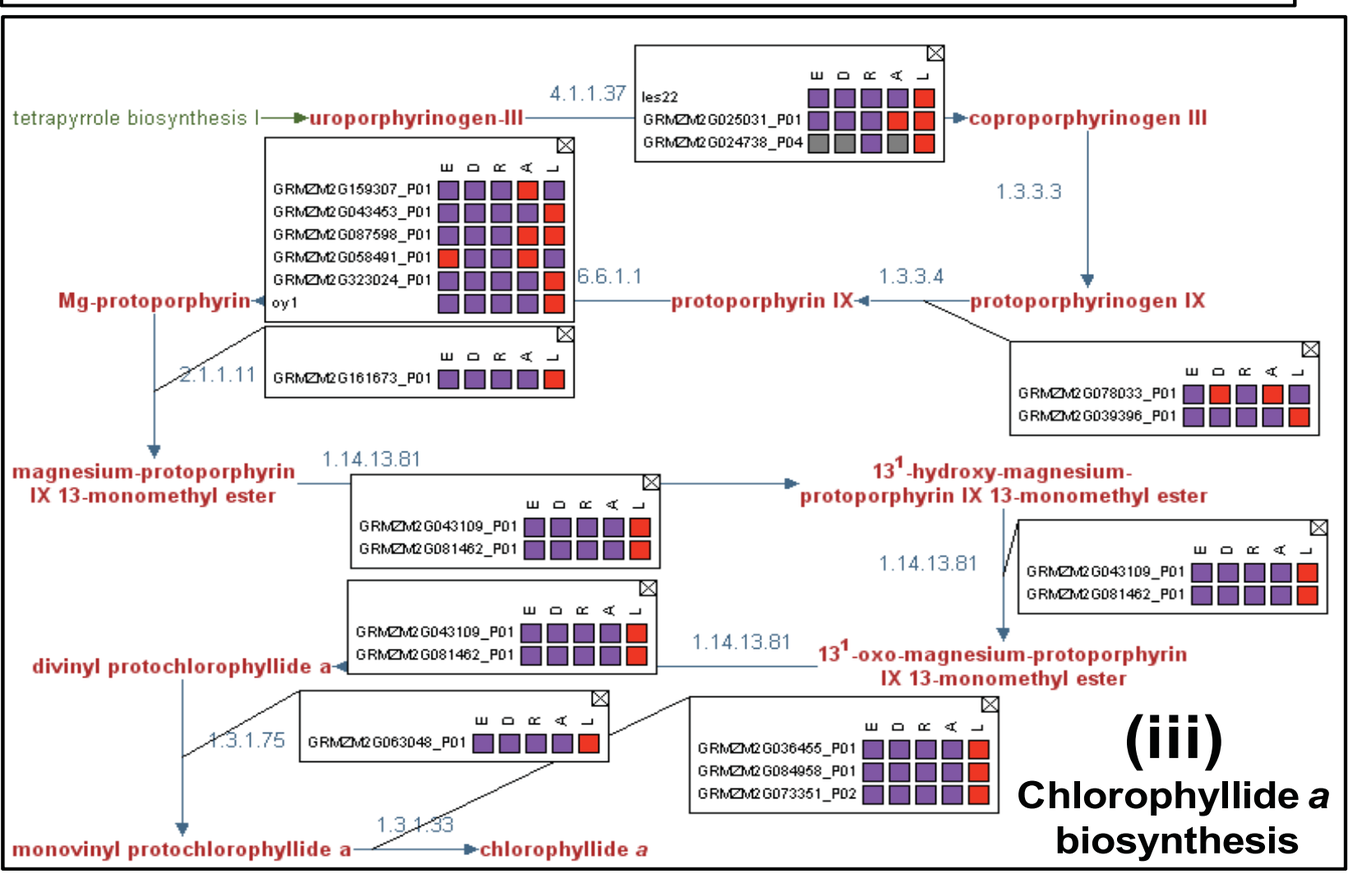
Figure-3b



The Plant Genome: Posted 23 Jan. 2013; doi: 10.3835/plantgenome2012.09.0025







(iii)
Chlorophyllide *a* biosynthesis

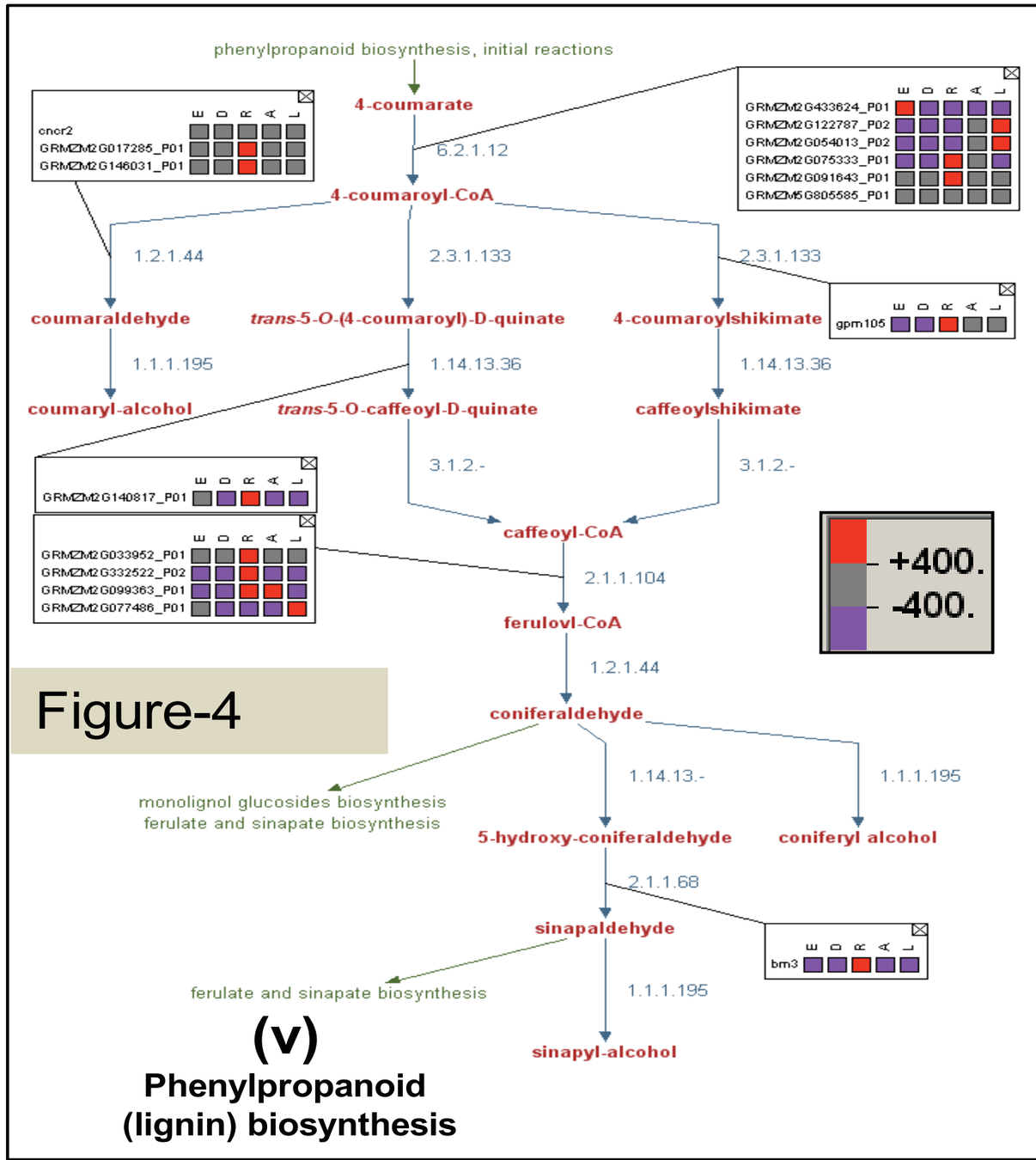


Figure-4

(v)
Phenylpropanoid (lignin) biosynthesis