**Supplementary Material**

**MaizeCyc Search and Browse Options**. On the entry page, users can search and browse the MaizeCyc database (Figure 4). Searching can be done using the “Search” link (Figure 4, Step 1) in the menu on the top of the page. For example, for “Gene/Protein/RNA search, the users can limit the search by sequence length, gene map position, product molecular weight, isoelectric point (pI), evidence code, cell component, GO term, among others. This advanced search capability enables users to obtain a list of molecules of their interest in a tabulated format and linkable to appropriate pages. If instead, the users would like to browse, they can use specific links next to the “Browse” row in the table (“Pathways” (Figure 4, Step 2), “Enzyme function” (Figure 4, Step 3), “Compounds”, and “Genes”). For example, when the pathways link is selected, the users will be directed to a page where pathways are displayed under functional classification provided in an expendable hierarchical view (Figure 4.2). Database summary provides a summary of total number of pathways, reactions, proteins, and pathways (Figure 4, Step 4), but also more granular information in a tabular format on the gene distributions (protein-coding, RNA, and pseudogenes) for each chromosome (Figure 4, Step 5).   
**Tools.** Several tools available at MaizeCyc, which can be found under the “Tools” menu on top of the page (Figure 4, Step 4). Among them are: Genome Browser, Cellular Overview, Genome Overview, and the Omics Viewer. The views of the last three tools are provided in the case study below in the Omics Viewer section.  
**Data downloads.** If interested users would like to get the whole MaizeCyc dataset, they can download it from the link provided in the table on the MaizeCyc page at [http://maizecyc.maizegdb.org](http://maizecyc.maizegdb.org/) (Figure 4, step-6).  We provide the complete database from the download section in the ZIP/TAR formats for installing local mirrors on users computers/webservers. However, this requires users to get a licensed software for installing the Pathway Tools available from the SRI (http://ecocyc.org/download.shtml). After the users have installed the local copy of Pathway Tools, they are encouraged to download the MaizeCyc database and unzip/untar in the directories \*/ptools-local/pgdb/users/ on their respective machines as a default directory. If the users are interested in getting a bulk download of the pathway, reaction, enzyme, compound, genes and gene products for simple analysis, a set of tab delimited files are also provided. For advanced users working on network modeling we provide data dumps in the standardized BioPax level-2 and level-3 and SBML formats. These files are compatible for viewing network using software like Cytoscape [ref].  
**Folate example**. To show the breath of the information that can be obtained using MaizeCyc, we will use folate (biosynthesis?) as an example. When users enter folate in the search box on the entry page (Figure 5, Step 1) in MaizeCyc, the web interface provides results shown in Figure 5i. On top of the page, Pathway Tools provides a bird eye’s categorized summary of the results for pathways, proteins, gene ontology terms, compounds, and reactions (Figure 5, Step 2). The page contains the linked results sectioned according to those categories. Users can reach more information by clicking on links. For example, following the “tetrahydrofolate biosynthesis” link (Figure 5, Step 3) the users can obtain the pathway details in an easy-to-understand schematic and clickable format: the users can reach information on upstream (Figure 5, Step 4) and downstream pathways (Figure 5, Step 5), EC numbers of enzymes (Figure 5, Step 6), gene products (Figure 5, Step 7), input (Figure 5, Step 8) and output (Figure 5, Step 9) metabolites. When the users click on a specific reaction, they can obtain more information about that reaction on a separate webpage, for example metabolite chemical structures (Figure 5, Step 10).    
  
**OMICs viewer:**  
The OMICs viewer tools enables users analyze the expression datasets from experiments on transcriptomics (microarray and RNA-Seq experiments), proteomics metabolomics, reaction flux and or any the data that can be assigned to genes, proteins, compounds, or reactions in reference to the metabolic network. Using this tool, users can upload their expression datasets and create a visual representation of cellular overview of all the pathways colored according to the values provided in the file and the cutoffs selected before uploading the datasets. We would like to mention that this tool is available with varying choices in the online vs. the locally installed desktop version of Pathway Tools. We will describe only the online version. While using the online version, user’s data are not compromised for security and are not available for sharing with other users, neither is it tracked by our webservers. Once users quit the session, the data views are discarded and users will be required to restart the session if desired. However, the data views can be saved in HTML and GIF image formats.  
The Omics viewer is accessible from a link on the MaizeCyc home page or from the ‘Tools’ option of the MaizeCyc section of the navigation bar highlighted in ‘orange’ colored stripe. Once on the OMICs viewer page, be prepared to have your own expression data (See standard format files HELP link on the OMICS viewer page), or download the gene expression data file used in this paper by downloading the supplementary file SF1 called ‘maize\_gene\_expression\_only\_upregulated\_genes\_from\_five\_tissues.txt’. the same file is also available from<http://ftp.maizegdb.org/MaizeGDB/FTP/MaizeCyc/>. [PJ3727] For simplification, here we choose 5 tissue types as examples and provide the columns they can be found (note that the first column in the data file is considered 0th column): (1) Embryo (16 days after pollination), (2) endosperm (16 days after pollination), (3) Root (v1 GH primary root), (4) Anther (R1 Anthers), and (5) leaf (V1 pooled leaves). The data file has 10,058 rows including the header row. The users need to download the file to their local computers before uploading on the OMICs viewer tool. Now select ‘Choose File’ button and select the appropriate expression data file by identifying the file at its location on your local computer. Then specify whether the data values are absolute or relative (for the example dataset, please choose absolute) and select single column or the ratio of two data columns as desired. Data values can either be ‘0-centered scale’ (e.g. log scale) or ‘1-centered scale’. The example dataset is not in logarithmic scale therefore, the users should choose ‘1-centered scale’. Next, inform the Omics Viewer what type of biological entities the data values are assigned to. Options include ‘Gene names and/or identifiers’, ‘Protein names and/or identifiers’, ‘Compound names and/or identifiers’, ‘Reaction identifiers and/or EC numbers’, and ‘Any of the above’. For our example set, choose ‘Gene name and/or identifiers’. The last option ‘Any of the above’ can be very useful for the users, in cases when users receive a large dataset from a collaborator and are not familiar with the type of the biological data that the data values are provided for in the column-0. By default column-0 (column-1 in user file is always the ID/name/synonym of the entity like gene, protein, compound). This is followed by entering the data column number users are interested in evaluating. It is worthwhile to remind the users one more time that the first column in the data file is actually considered 0th column by the Omics Viewer. Two boxes are located under the title ‘Single Experiment Time Step or Animated Time Series’.  The users should enter in the first box the data column(s) they are interested. If there is more than one column they would like to display, the column numbers can be entered one per each line. The second box is useful only when the user is interested in relative values; the columns entered in this box will be used as denominators. For the example set, the user should put any number between 1 and 5 in the first box. These columns are selected for tissue types mentioned above. Following the data selection, users get three options to select for color schema, (1) Full color spectrum, computed from data provided (default), (2) full color spectrum with a maximum cutoff, and (3) three color display with specified threshold. The first option is self-explanatory.  In the second option, the data values over the maximum cutoff are displayed in red, and the rest in the full color spectrum. In the third option, three colors are used. By default it uses red if data values are greater than the threshold, purple if data values are smaller than the inverse of the threshold, and grey for the values in between. You can also choose your threshold cutoff. For example for this paper we used 400. A more detailed explanation can be found on the Omics Viewer page. For the example set, the users can choose the first option. In the final step before hitting the submit button, users get to select the display type: (1) Paint data on cellular overview chart (default), (2) paint data on genome overview chart, and (3) generate a table of individual pathways exceeding threshold.  In this manuscript, we will focus on the cellular overview and now click the “Submit” button. The Omics Viewer will then start creating views. The time necessary to create these files depends on the size of the data file provided by the users. For the example data file, it might take about 10 minutes and 5 minutes to generate the cellular overview and genome overview respectively. On the cellular overviews with painted expression data, users will find legends explaining the color schema, cutoffs and if the data is from a time series or multiple columns (such as in the example file), users will see the moving images with expression levels marked for up-, down-regulated or no change. By hovering the mouse over the painted reaction and pathway it provides a pop-up option to zoom on the pathway view to visualize the expression for each gene associated to the given reaction. If for example there are multiple genes in a complex or multiple isomer (paralog) gene ids are listed each one will be colored respectively for its specific expression value. The ‘save views’ option is available at the bottom of the painted overview page.  
Supplementary Figure SX shows the cellular overview (Step 1) and genome overview chart of the NimbleGen expression data in Sekhon *et al.* (PMID: 21299659) for [shoot apical meristem and stem V4](http://www.maizegdb.org/cgi-bin/termrefs.cgi?id=2366338) expression (Step 2). The figures are colored according to a heat map based on expression levels: the higher the red content in the color, the higher the expression is. The cellular overview displays different types of molecules (amino acids, carbohydrates, proteins, purines, pyrimidines, cofactors, and tRNAs) with different symbols. When any of these molecules are phosphorylated, then the symbol is filled with black. There are advantages to both cellular and genome views. The cellular overview enables users to have a global view of the expression profiles of genes in the context of pathways they belong to. In contrast, the genome overview chart displays the expression of genes in the context of genomic location, which may or may not mean being in the same pathway. The genes that are part of the same operon are expected to be annotated in the same pathway; however, genomic proximity does not usually signify proximity of pathways. Having these two views therefore enables users to choose appropriate view for their research problem. It is also possible to use both views to identify operons: if a series of neighboring genes are often expressed as a unit in different conditions, an argument can be made that these genes are part of the same operon. In the same manner, it might be possible to enhance annotation in the pathways, because such genes might be also placed in the same pathway. Therefore Pathway Tools and specifically the Omics Viewer can be used to generate hypotheses for annotating genes (or proteins, compounds etc.) that can be tested experimentally.    
*Highlighting overexpressed genes.* The Omics viewer itself does not provide tools to test whether the expression level of a given gene is statistically significant. However, it allows quick visual comparisons of overexpressed genes and pathways. The users can choose a specific threshold for their data by going back to the Omics Viewer main page, and enter that the desired threshold next to “Three color display with specified threshold” under the Color Scheme section. The users can then see the highlighted pathways from a bird’s eye view. If they know where their pathways of interests are located, they can get a quick glimpse of their expression patterns. The Omics viewer also provides a list of overexpressed genes when “generate a table of individual pathways exceeding threshold” option is chosen on the main Omics page. For example, for the case of genes that have an expression value of 14 in the shoot apical meristem, the Omics Viewer enumerate the following pathways being activated: glycolysis III, glycolysis I, glycolysis IV (plant cytosol), aerobic respiration -- electron donor II, gluconeogenesis I, adenosine nucleotide de novo biosynthesis, UDP-glucose biosynthesis (from glucose 6-phosphate), glutamine biosynthesis III, TCA cycle variation III (eukaryotic), NAglycoD/NADH phosphorylation and dephosphorylation. Higher levels of activity in glycolysis and biosynthesis pathways are expected in the shoot apical meristem, where growth takes place.  
*Differential expression.* Instead of looking at each data column separately that specifies a single condition, the Omics Viewer allows displaying the *ratio* of two data columns that provides *relative* expression levels. Supplementary Figure SX shows the ratio of expressed genes in [shoot apical meristem and stem V4](http://www.maizegdb.org/cgi-bin/termrefs.cgi?id=2366338) vs. leaf base of expanding leaf V5 (Step 3a), and embryo, 24 days after pollination vs. kernel, 24 days after pollination (Step 3b). Looking at the relative expression levels can inform the users about the tissue-specific genes and pathways. A quick comparison of pathways reveals change in expression patterns. In Steps 3 and 4, suberin biosynthesis (red box) and C4 photosynthetic carbon assimilation cycle pathways (blue box) are highlighted to show tissue-specific expression differences.  
*Visualizing specific pathways.* Although the global view of the cellular pathways is definitely useful, the users may also want to look at specific pathways of their interest at a higher resolution for analysis or maybe for publications. The Omics Viewer has the capability of displaying each pathway individually, which enables to save them as image files.  At the end of the page of the colored cellular overview, there is a link that says “Instructions for saving this diagram to your local disk.” This link contains more than the instructions. It also has a link to the set of pathway image files. The users can click in the “Link” in the second bullet point that says. The link takes you to a webpage that has all the pathway images with expression levels displayed on them.  These figures can be copied and pasted in presentations and used in publications. As an example, Supplementary Figure SX  shows a detailed comparison of expressions of suberin biosynthesis pathways for [shoot apical meristem and stem V4](http://www.maizegdb.org/cgi-bin/termrefs.cgi?id=2366338) vs. leaf base of expanding leaf V5 (Step 3a), (Step 4a) and embryo, 24 days after pollination vs. kernel, 24 days after pollination (Step 4b). The figures also allow a visual comparison of expression levels in the isozymes associated with the suberin biosynthesis pathway, differentiated with colors as well as with the number of lines designating pathways.

\* TUTORIAL section (how to use the resource)

1). This section is a bit too long in the context of the paper. Consider minimizing it, include succinct description of the tools while moving depicted case examples to supplementary materials

2) Provide examples (ie, how the user can...) that highlight various uses of the resource (a & b could fall under a candidate-gene type analysis, while c is a more systems biology whole-centric/global genome-scale analyses):

   a) Web-server

   b) Desktop client

   c)  Standard formats for data exchange - pathologic-formatted files for example could be exported for use/integration with other tools