II.4 AraCyc: Overview of an Arabidopsis Metabolism Database and its Applications for Plant Research

S.Y. Rhee, P. Zhang, and H. Foerster[1]

1 Introduction

Metabolism is one of the most fundamental processes of life. Each organism possesses an intricate network of metabolic pathways, whose elaborate regulatory circuitry may be developmentally programmed and hard-wired to respond to changes in the environment. With the release of the fully sequenced plant genomes of Arabidopsis and rice (AGI 2000; Goff et al. 2002; Yu et al. 2002), and the initiation of many sequencing projects of other plant species, there is a growing desire to place the sequenced and annotated genomes in a metabolic context. AraCyc (http://arabidopsis.org/tools/aracyc/) was the first plant organism-specific metabolism database to be computationally predicted by the PathoLogic component of the Pathway Tools software using MetaCyc as the reference database (Mueller et al. 2003). With continued manual curation, the goal of AraCyc is to describe the complete set of metabolic pathways for *Arabidopsis thaliana* whilst placing genes and enzymes within their metabolic context. Though many enzymes in AraCyc have yet to be manually curated, most of the pathways have been manually validated and it is so far the only genome-wide, comprehensive metabolic database for a single plant species (Zhang et al. 2005).

The benefits of a species-specific metabolic pathway database are substantial: (1) it depicts the biochemical components of an organism; (2) it aids in comparative studies of pathways across species to facilitate metabolic engineering to improve crop metabolic traits; (3) it can be used as a platform to integrate and analyze data from large-scale experiments such as gene expression, protein expression, or metabolite profiling; finally (4) by presenting pathway steps lacking assigned genes, or having genes assigned but solely based on computational prediction, it allows the identification of the biochemical steps that remain to be identified and experimentally characterized. The manual, de novo creation of a pathway database can be labor intensive and time consuming. SoyBase (http://www.soybase.org/) is the only other plant pathway database, specific to soybean, which was manually created and made publicly available. Alternatively, there are metabolic pathway databases that cover a wide range of organisms. Examples of comprehensive pathway databases include Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) (Ogata

[1] Carnegie Institution, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305, USA, e-mail: rhee@acoma.stanford.edu

Biotechnology in Agriculture and Forestry, Vol. 57
Plant Metabolomics (ed. by K. Saito, R.A. Dixon, and L. Willmitzer)
© Springer-Verlag Berlin Heidelberg 2006

et al. 1999; Kanehisa 2002; Kanehisa et al. 2004), Enzymes and Metabolic Pathways (EMP, http://www.empproject.com/) (Selkov et al. 1996), and MetaCyc (www.metacyc.org) (Krieger et al. 2004). Each has its strengths and weaknesses, some of which have been reviewed (Maranas and Burgard 2001; Kanehisa 2002).

In this review, we describe the content and functionalities of AraCyc database as well as examples of applications that use the information contained in the database, in conjunction with functional genomics data to address systems-wide questions about metabolism. In addition, we discuss the current limitations and future directions of the database.

2 Database Content

AraCyc (version 2.5) currently features 197 pathways, comprising 979 unique reactions and 1071 compounds. Over 63% of the reactions have Arabidopsis genes/enzymes assigned and 1759 unique genes are assigned to the pathways. A metabolic pathway is a set of one or more enzymatic transformations, involved in processes such as biosynthesis, degradation, conversion,

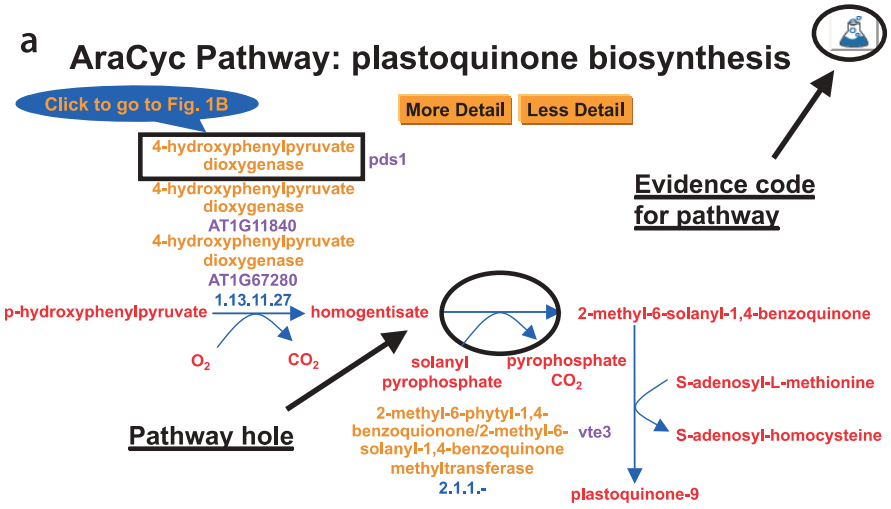

Fig. 1. An example of an AraCyc pathway: **a** pathway evidence, which could be either computational (*indicated by a computer icon*) or experimental (*indicated by a flask icon*), provides assertion of the existence of the pathway in Arabidopsis; **b** (see next page) similarly, evidence attached to an enzyme provides assertion of its catalytic activity involved in a specific reaction. Each piece of evidence is associated with a citation where the source of the evidence can be found (*inset*). Pathway can be zoomed to show ~~varies~~ levels of details. Compounds, reactions, enzymes and genes on a pathway detail page are clickable ~~for~~ more information. (Reprinted with permission from Plant Physiology)

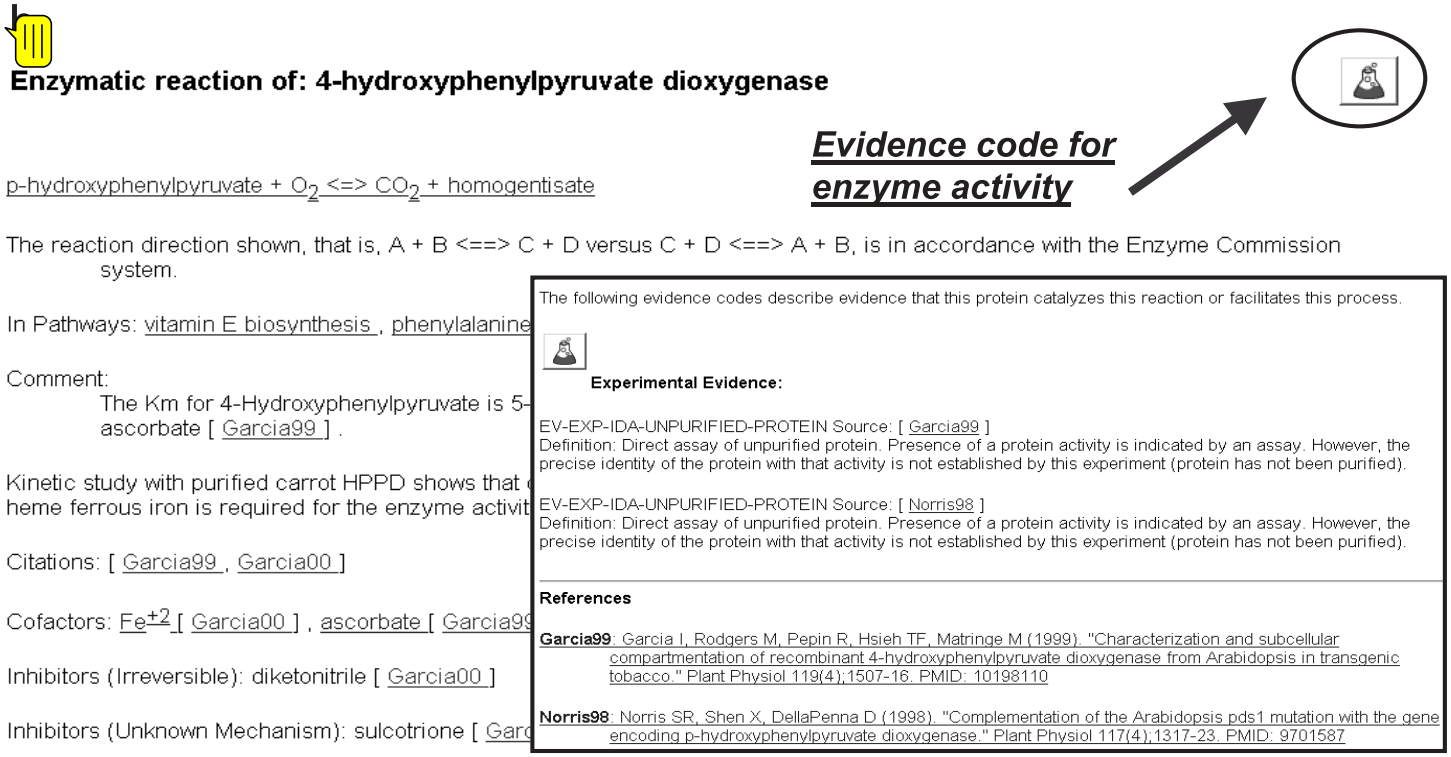

Fig. 1. (continued)

or utilization, as it occurs in a particular organism (Krieger et al. 2004). In addition to the 197 individual pathways, AraCyc has 15 super-pathways. A super-pathway is an aggregation of two or more individual pathways that are related in some way (Krieger et al. 2004). The reactions in AraCyc have EC numbers (Enzyme Commission Nomenclature, http://www.chem.qmul.ac.uk/iubmb/enzyme/) assigned, when available (Fig. 1a). Chemical structures are annotated to the compounds. The assignments of the enzymes to the reactions are based on the characterization of the enzymes or the functional annotations of the genes, which could be either experimentally determined or computationally derived. For example, cellulose synthases CesA1 and At3g02230 are both assigned to EC 2.4.1.12 on the cellulose biosynthesis pathway. However, the cellulose synthase activity is supported by functional studies of the enzyme only for CesA1, and the annotation of 'cellulose synthase' for At3g02230 comes from computational prediction based on sequence similarity. To distinguish the different levels of annotation qualities, an evidence code is provided along with the assignment of an enzyme to a reaction. These evidences can be easily recognized through the use of intuitive evidence icons (computer screen for computational, flask for experimental), which label each enzyme detail page (Fig. 1b). Citations are provided along with the evidence so that users can obtain more details about the source of the annotation.

The 197 pathways are classified into three main categories: "Biosynthesis", "Degradation/Utilization/Assimilation", and "Generation of Precursor Metabolites and Energy" (Table 1). Biosynthesis of all 20 protein amino acids, all DNA/RNA purine and pyrimidine nucleosides and nucleotides, commonly occurring sugars and polysaccharides, major fatty acid and lipid classes (including triacylglycerol, phospho- and glyco-lipids, cofactors, prosthetic groups and electron carriers~~)~~, and six known major plant hormone classes are represented. In addition, biosynthesis of the major molecules found in plant primary and secondary cell wall and epidermal structures, including cellulose, homogalacturonan (a component of pectin), lignin, suberin, wax and cutin are included. Pathways for central energy metabolism are well represented. It is not easy to assess the comprehensiveness of pathways under "Degradation/Utilization/Assimilation" as there is much less information available for catabolism than for biosynthesis in plants. Secondary metabolism is not yet covered comprehensively but this situation will change in the near future (see Sect. 6).

Data objects in AraCyc, such as pathways, reactions and compounds, as well as subcellular compartments (for annotating enzyme locations), and evidence types are structured in hierarchical ontologies (Gruber 1993; Karp 2000; Karp et al. 2004). Each ontology describes concepts (terms) and relationships between them. Terms are organized into classes~~, subclasses and instances~~ according to the primary 'is-a' relationship. The 'is-a' relationship classifies what type of a concept a term is. The broader concepts, or parent terms~~ such as classes~~, appear on ~~the top~~ level ~~on~~ the hierarchy tree. The more specific concepts, or children terms ~~such as subclasses and instances~~ are grouped under the broader

TS[e] Please provide a short running title

Editor's or typesetter's annotations (will be removed before the final TEX run)

Table 1. Summary of AraCyc database content

Total pathways ~~(~~excluding super-pathways~~)~~	210 (197)[a]	
Biosynthesis	137 (127)[a]	
Amino acids	36	
Cell structure	8	
Cofactors, prosthetic groups, electron donors	24	
Fatty acids and lipids	15	
Plant hormones	14	
Nucleosides and nucleotides	4	
Secondary metabolism	17	
Sugars and polysaccharides	9	
Others	10	
Degradation	58	(57)[b]
Amino acids	21	
Fatty acids and lipids	7	
Inorganic nutrients	5	
Sugar derivatives	2	
Sugars and polysaccharides	11	
Others	12	
Generation of precursor metabolites and energy	15[b]	
Total unique reactions of pathways	979	
Total unique compounds of pathways	1071	
Total unique genes of pathways	1759	

[a] Some pathways are classified to more than one pathway class. The numbers in parenthesis are unique number counts

[b] Two pathways are classified under both 'Degradation' and 'Generation of precursor metabolites and energy'

concepts. For example, within the pathway ontology, alanine biosynthesis is classified to Pathways→Biosynthesis→Amino acids→Individual amino acids (http://www.arabidopsis.org:1555/ARA/NEW-IMAGE?type=PATHWAY{\&}object=ALANINE-SYN2-PWY). Unlike terms in a simple list, the organization of terms into a hierarchical ontology allows more robust queries and makes retrieval of related information easier.

3 Search, Browse, and Analyze Functionalities

AraCyc can be freely accessed through the web (http://arabidopsis.org/tools/aracyc) using a common web browser, or downloaded as text files. The database is available to download with an 'open source' license (http://arabidopsis.org/aracyc/form.html). Installing AraCyc database and desktop version of the software on a local computer has a few advantages such as allowing

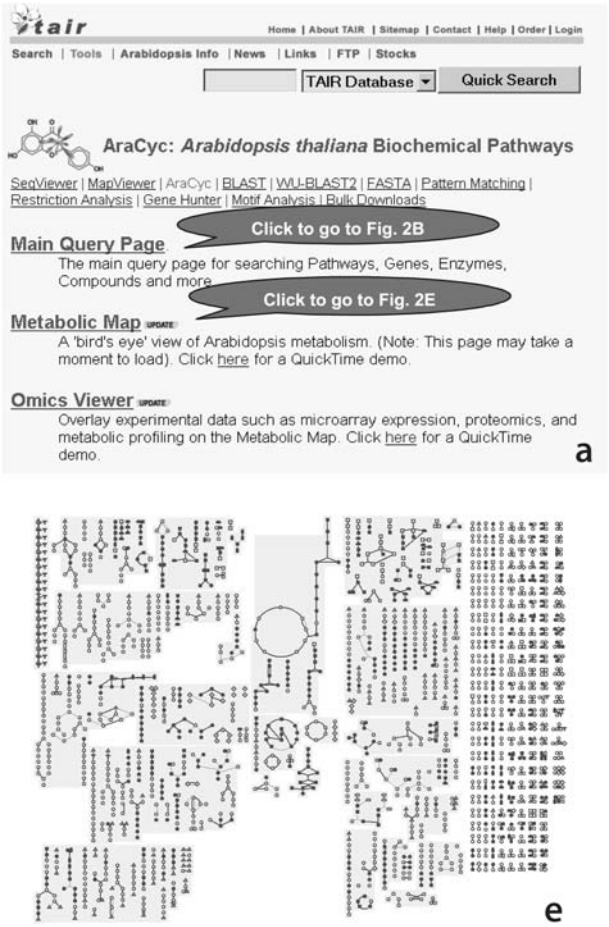

Fig. 2. Accessing the metabolic data: **a,b** from the AraCyc main page, data can be queried by names and browsed from alphabetic lists or hierarchy ontologies; **c** query results are displayed grouped by data types–different data types such as compounds, reactions, pathways, enzymes and genes are interlinked from individual data detail pages; **d** an example of a compound detail page; **e** the Metabolic Map depicts all the pathways ~~all~~ in one diagram. Clicking on a pathway glyph will open up the pathway detail page. Experimental data such as gene expression and metabolic profiling can be painted onto the Metabolic Map using the Omics Viewer tool

the user to update the database with proprietary data and to perform more advanced queries. Using either the web or the desktop version, a user can browse, query, and visualize the data (Fig. 2a). One can navigate through all of the pathways, for example, from an alphabetic list or from the hierarchy ontology browser (http://www.arabidopsis.org:1555/ARA/class-instances?object=Pathways). Substrings, or partial words, for example, 'gibber' in 'gibberellin', can be queried against names of a specific data type such as com-

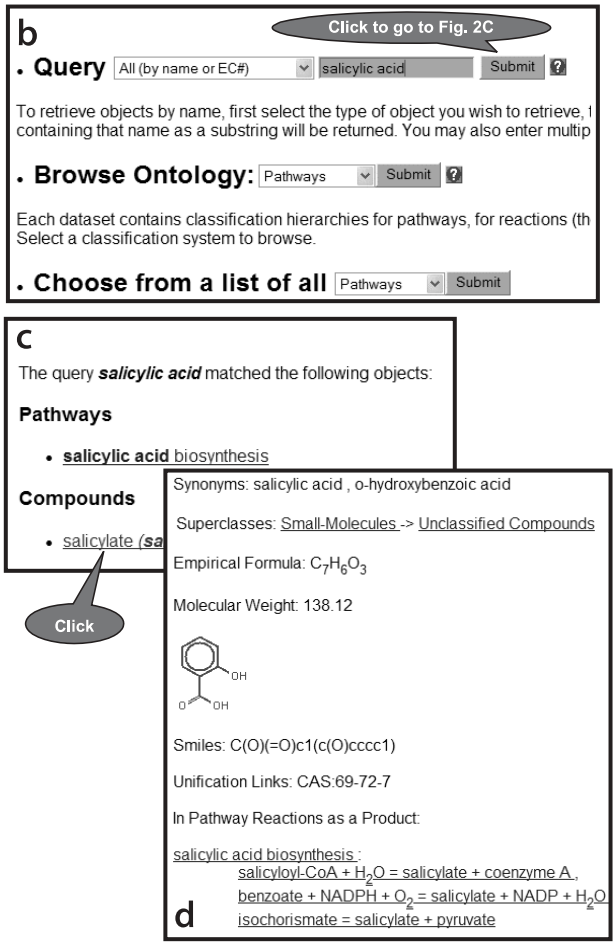

Fig. 2. (continued)

pounds only, or all of the data types (Fig. 2b). In the latter case, results are grouped according to ~~the~~ different data types (Fig. 2c). Each result is linked to its corresponding detail page. In the example shown in Fig. 2, clicking on the compound '~~auxin~~' from the query result page opens up the compound detail page for ~~auxin~~ (Fig. 2d). Many data are interconnected by hyperlinks. Pathways and reactions shown on the compound ~~auxin~~ detail page are linked to the corresponding pathway detail pages and reaction detail pages, and vice versa. The 'Metabolic Map' tool shows a bird's eye view of all of the pathways grouped by the pathway classes (Fig. 2e). In addition to the searching and browsing options using a web browser or the desktop application, datasets are also provided as downloadable text files, (ftp://ftp.arabidopsis.org/home/tair/Pathways/), such as a pathway dump file that lists all of the pathways and the genes and enzymes

assigned to each pathway. A user who has a list of genes of interest can use this file to quickly sort out what pathways the genes are involved in. The files are updated with each AraCyc release.

An important component of the Pathway Tools software package (Karp et al. 2002) is the Omics Viewer (http://aracyc.stanford.edu:1555/expression.html). It allows the analysis of changes in the levels of transcripts, proteins, and

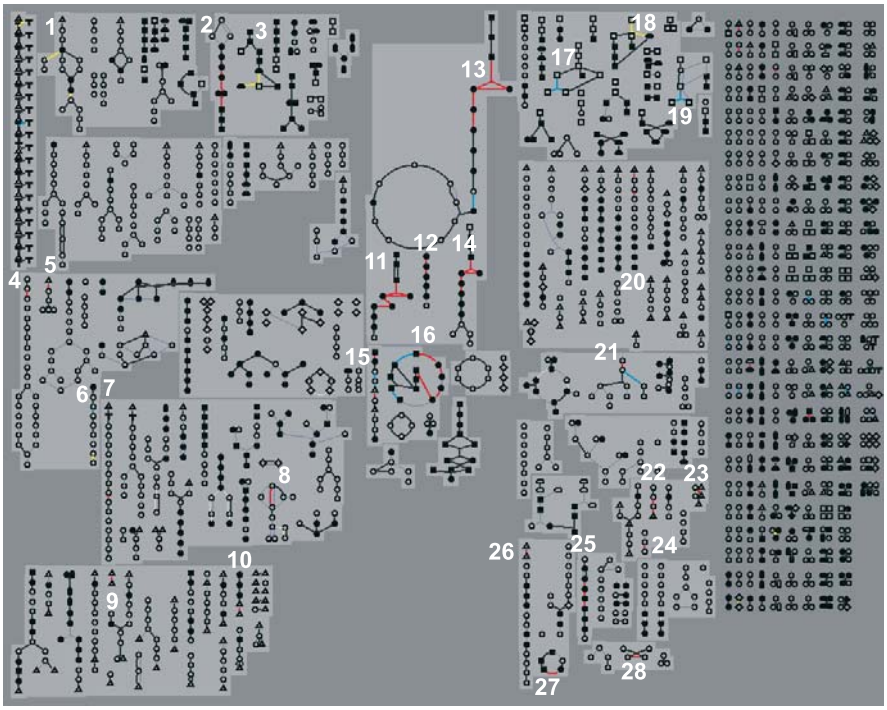

Fig. 3. Proteomics data from Weckwerth et al. (2004) overlaid on AraCyc Omics Viewer. Each glyph represents a pathway in which reactions are represented by *lines* and metabolites are represented by *triangles* (amino acids), *squares* (carbohydrates), or other metabolites (*circles*). Individual reactions that are not placed within a metabolic pathway are *listed on the right side of the diagram*. The 281 proteins that were ~~found~~ in the leaves of Columbia ~~(*red lines*)~~ and C24 accessions were overlaid onto the metabolic map ~~and 45~~ proteins were found to carry out reactions ~~of~~ 28 pathways on the map ~~and~~ are highlighted. *Red lines*, proteins ~~that~~ were ~~found~~ only in Columbia; *blue lines*, ~~reactions~~ that were ~~found~~ in both ecotypes, and *yellow lines*, ~~reactions~~ that were ~~found~~ only in C24. Labeled pathways are: *1*, lignin biosynthesis; *2*, gluconeogenesis; *3*, sucrose biosynthesis; *4*, brassinosteroid biosynthesis; *5*, ethylene biosynthesis; *6*, jasmonic acid biosynthesis; *7*, chlorophyll biosynthesis; *8*, formylTHF biosynthesis; *9*, glutamine biosynthesis I; *10*, serine biosynthesis; *11*, glycolysis IV; *12*, glyceraldehyde 3-phosphate catabolism; *13*, glycolysis; *14*, glycolysis; *15*, phosphorespiration; *16*, Calvin cycle; *17*, starch degradation; *18*, sucrose degradation; *19*, glycogen catabolism; *20*, methionine degradation I; *21*, lipoxygenase pathway; *22*, nitrate assimilation; *23*, ammonium assimilation; *24*, cyanate catabolism; *25*, aerobic glycerol catabolism; *26*, serine isocitrate lyase pathway; *27*, xylulose monophosphate cycle; *28*, removal of superoxide radicals

metabolites by overlaying results of genome-wide gene expression, proteomics, or metabolite profiling data onto the metabolism overview diagram (Fig. 3). Each reaction (represented as a line connecting the compounds) can be color-coded according to the expression level of the gene or protein that catalyzes the reaction. Metabolite levels can be depicted by color-coding the symbols for compounds (represented as squares or triangles connected by the reaction lines). Note that only those genes and compounds that are included in AraCyc can be displayed on the metabolic map. However, it is possible to extrapolate from the Omics Viewer to identify additional components of a pathway. For example, if a set of genes from an expression array appeared to be all involved in the same pathway and showed similar changes in expression values, one could cluster the original dataset to identify other genes having a similar expression profile. These genes, in turn, may represent components of the pathway that are missing from AraCyc. Specific usage examples of this tool are described in the next section.

4 Applications of AraCyc

4.1 Putting Functional Genomics Data into a Metabolic Network Framework

The Omics Viewer is a convenient way of quickly assessing the metabolic changes from a large-scale experiment such as gene expression profiling or metabolite profiling under different environmental or genotypic conditions, either to test a specific hypothesis or explore the trends in a large-scale data set. Alternatively the viewer can be used to annotate a set of genes grouped by certain criteria such as gene families or co-expressed genes. For example, Arabidopsis accessions Columbia (Col-0) and C24 are known to have a number of polymorphisms that lead to differences in a variety of phenotypic traits (Rohde et al. 2004). For example, Columbia can acclimate to cold and tolerate freezing much better than C24. However, the exact molecular, biochemical, and physiological differences that result in this phenotypic trait are not known. One hypothesis is that their metabolic state is different. Weckwerth and colleagues compared the protein and metabolite content of Columbia and C24 leaves (Weckwerth et al. 2004). They ~~found~~ 297 proteins, of which 153 were ~~found~~ in both accessions and 144 were detected in only one of the ecotypes. An equivalent number of proteins was found in each accession (30% specific in Columbia, 30% in C24, and 40% detected in both). Overlaying 281 of the differentially expressed proteins on the AraCyc Omics viewer shows that about 37% of the proteins (104/281) were placed within the metabolic framework, specifically onto 28 pathways (Fig. 3). The mapping shows that Columbia is much more active metabolically than C24, with 19 out of 28 pathways that were mapped with proteins specific to Columbia. It is possible that Columbia's more

active metabolic state may be relevant for its ability to acclimate to cold better than C24. Pathways that were more prominent in C24 than in Columbia include lignin biosynthesis, sucrose biosynthesis, and sucrose degradation. While it is premature to derive definitive conclusions from this exercise, it demonstrates the ability of this tool to explore quickly and efficiently ~~large-scale datasets~~.

Similarly, the Omics Viewer can be used to address hypotheses using large-scale data such as gene expression microarrays. For example, it is well-established that CBF (CRT/DREB Binding Factor) transcription factors play an important role in cold acclimation (Thomashow 1999; Cook et al. 2004; Gilmour et al. 2004). Also, increases in sugars such as sucrose, glucose, fructose, and raffinose, and in the amino acid proline, are correlated with the ability to tolerate freezing, perhaps because these compounds act as osmoprotectants (compatible solutes) (Strand et al. 1999; Taji et al. 2002; Shinozaki et al. 2003; Uemura et al. 2003; Zuther et al. 2004). It is, however, not known whether the cold acclimation process via the CBF pathway directly affects the increase in the production of these metabolites (Stitt and Hurry 2002). Recently Vogel and colleagues asked which genes are regulated by the CBF pathway in the cold acclimation process by examining the global gene expression profiles of cold-treated wild type plants and lines overexpressing the *CBF2* gene (Vogel et al. 2005). They found that 93 genes were affected in both of these lines as compared to wild type and considered these genes to be involved in the CBF cold-response pathway. By overlaying the expression profiles of these genes on the Omics Viewer, we can quickly assess which metabolic pathways are affected via the CBF cold-response pathway (Fig. 4). The results show that 31 out of 93 genes were placed within the metabolic context. Six pathways are induced by CBF overexpression and cold treatment and one pathway, glucosinolate biosynthesis, is reduced by CBF overexpression and cold treatment. Pathways in which genes encoding enzymes ~~performing one or more~~ reactions are induced include sucrose biosynthesis, flavonoid biosynthesis, flavonol biosynthesis, anaerobic glycolysis, homogalacturonan degradation, and sucrose degradation. This result suggests that sucrose biosynthesis may be directly affected by the CBF pathway.

4.2 Application of AraCyc Data to Other Software Environments

In addition to addressing specific questions using the data and tools in AraCyc as exemplified above, AraCyc data can be used in other software environments to explore metabolism data in conjunction with other data such as gene expression profiling, metabolite profiling, and proteomics data. Examples of third-party software that have used data from AraCyc include MapMan (Thimm et al. 2004), MetNetDB (Wurtele et al. 2003), FCModeler (Wurtele et al. 2003), and 3D virtual reality visualization environment (Dickerson et al. 2003). MapMan is a user-driven software environment that allows the display and analysis of large-scale datasets ~~onto metabolic pathway context~~. It is com-

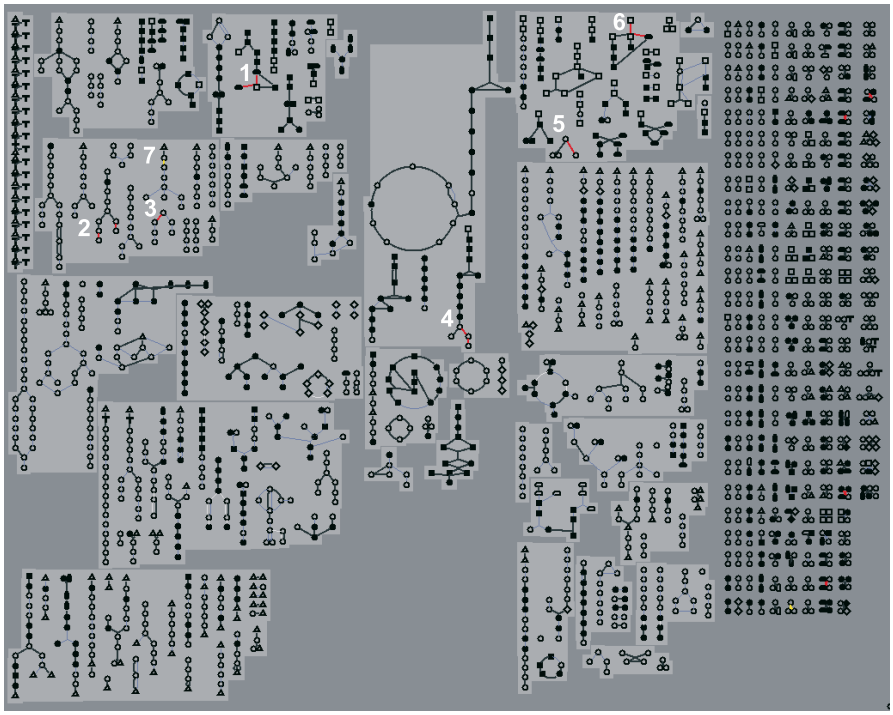

Fig. 4. Genes that belong to the CBF regulon in Arabidopsis from Vogel et al. (2005) overlaid on AraCyc Omics Viewer. Ninety three genes were used as input and 31 genes were placed on the metabolic framework into seven pathways. *Red* (induced) and *yellow* (reduced) *lines* indicate reactions that are performed by the proteins encoded by genes whose transcripts are induced or reduced by the CBF transcription factor, respectively. Six pathways are induced by CBF overexpression and cold treatment and one pathway, glucosinolate biosynthesis, is reduced by CBF overexpression and cold treatment. Labeled pathways are: *1*, sucrose biosynthesis; *2*, flavonoid biosynthesis; *3*, flavonol biosynthesis; *4*, anaerobic glycolysis; *5*, homogalacturonan degradation; *6*, sucrose degradation; *7*, glucosinolate biosynthesis

posed of two modules, SCAVENGER and IMAGEANNOTATOR. SCAVENGER designates measured values of an experiment onto a set of metabolic pathways and other processes that are organized into bins and IMAGEANNOTATOR allows users to generate a custom view of the annotated bins and the measured parameters according to their specifications and needs. MetNetDB is an Arabidopsis interactions database that is used as a basis for FCModeler software. FCModeler software uses fuzzy cognitive maps to allow biologists to generate models of regulatory and metabolic pathways from data in MetNetDB and large-scale datasets such as those resulting from genome-wide gene expression profiling experiments. The 3D virtual reality environment uses the data in MetNetDB and allows visualization of the network data in 3D in a virtual reality cave such that users can 'get inside' the pathways and explore the data

from particular areas within the network. All of the applications mentioned here have imported the AraCyc data to be visualized and analyzed in a number of flexible and creative ways.

5 Current Issues and Future Directions

About 86% (170 pathways) of the pathways have been manually validated, meaning that the pathway diagrams have been validated and corrected according to the latest literature information. The remaining 27 pathways were predicted to exist in Arabidopsis but no experimental support was found in the literature irrevocably confirming their existence in plants. Pathways of secondary metabolism are under-represented in AraCyc. Curation of new secondary metabolic pathways is an ongoing task. In addition, we plan to curate and integrate transporters into their relevant pathways. Users are encouraged to contact us (curator@arabidopsis.org) for data submissions, including updating or correcting an existing pathway, or submitting a new pathway.

Starting in 2005, updates to AraCyc are released on a quarterly basis. Each release includes manual updates, corrections of the existing pathway data, and manual curation of new pathway data. A major release at the end of each year is planned to take advantage of the progress in the functional annotations of the Arabidopsis genome (Berardini et al. 2004; Zhang et al. 2005). A gene whose function was previously unknown may now have an annotated function and thus may be assigned to a corresponding AraCyc pathway.

Many enhancements to the data visualization capabilities provided by the Pathway Tools software are planned, such as the ability to display pathways in the context of subcellular location information.

6 Conclusions

Currently we are experiencing a rapidly increasing rate of production of large-scale data such as genome sequences, genome-wide gene expression profiles, proteomics and metabolomics data. The necessity to organize all of these data into a biological framework has been, in part, the motivation for the work described in this review. While we have created a comprehensive database that describes the metabolic network of a model plant species, *Arabidopsis thaliana*, the database is far from being either complete or error-free. Many of the pathways are in need of manual curation using the current literature and many more pathways, particularly those for secondary metabolism and those that include transport reactions, need to be brought into the database. As with any other database project, the content of the AraCyc database is dynamic and will continue to undergo enhancement, additions, and modifications to make it more useful.

Acknowledgements. We are grateful to Tanya Berardini and Leonore Reiser for their careful reading of the manuscript. Work described in this project is funded in part by grant #1-R01-GM65466–01 from the NIH National Institute of General Medical Sciences and grant # 0417062 from NSF Biological Database and Informatics division.

References

AGI (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Berardini TZ, Mundodi S, Reiser L, Huala E, Garcia-Hernandez M, Zhang P, Mueller LA, Yoon J, Doyle A, Lander G, Moseyko N, Yoo D, Xu I, Zoeckler B, Montoya M, Miller N, Weems D, Rhee SY (2004) Functional annotation of the Arabidopsis genome using controlled vocabularies. Plant Physiol 135:745–755

Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of Arabidopsis. Proc Natl Acad Sci USA 101:15243–15248

Dickerson JA, Yang Y, Blom K, Reinot A, Lie J, Cruz-Neira C, Wurtele ES (2003) Using virtual reality to understand complex metabolic networks. Atlantic symposium on computational biology and genomic information systems and technology, september, pp 950–953

Gilmour SJ, Fowler SG, Thomashow MF (2004) Arabidopsis transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. Plant Mol Biol 54:767–781

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science 296:92–100

Gruber TR (1993) A translation approach to portable ontology specifications. Knowledge Acquisition 5:199–220

Kanehisa M (2002) The KEGG database. Novartis Found Symp 247:91–101; discussion 101–103, 119–128, 244–152

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–D280

Karp PD (2000) An ontology for biological function based on molecular interactions. Bioinformatics 16:269–285

Karp PD, Paley S, Romero P (2002) The pathway tools software. Bioinformatics 18 [Suppl 1]:S225–S232

Karp PD, Paley S, Krieger CJ, Zhang P (2004) An evidence ontology for use in pathway/genome databases. Pacific Symp Biocomput 9:190–201

Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res 32:D438-D442

Maranas CD, Burgard AP (2001) Review of EcoCyc and MetaCyc databases. Metab Eng 3:98–99

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

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27:29–34

Paley SM, Karp PD (2002) Evaluation of computational metabolic-pathway predictions for *Helicobacter pylori*. Bioinformatics 18:715–724

Rohde P, Hincha DK, Heyer AG (2004) Heterosis in the freezing tolerance of crosses between two *Arabidopsis thaliana* accessions (Columbia-0 and C24) that show differences in non-acclimated and acclimated freezing tolerance. Plant J 38:790–799

Selkov E, Basmanova S, Gaasterland T, Goryanin I, Gretchkin Y, Maltsev N, Nenashev V, Overbeek R, Panyushkina E, Pronevitch L, Selkov E Jr, Yunus I (1996) The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. Nucleic Acids Res 24:26–28

Shinozaki K, Yamaguchi-Shinozaki K, Seki M (2003) Regulatory network of gene expression in the drought and cold stress responses. Curr Opin Plant Biol 6:410–417

Stitt M, Hurry V (2002) A plant for all seasons: alterations in photosynthetic carbon metabolism during cold acclimation in Arabidopsis. Curr Opin Plant Biol 5:199–206

Strand A, Hurry V, Henkes S, Huner N, Gustafsson P, Gardestrom P, Stitt M (1999) Acclimation of Arabidopsis leaves developing at low temperatures. Increasing cytoplasmic volume accompanies increased activities of enzymes in the Calvin cycle and in the sucrose-biosynthesis pathway. Plant Physiol 119:1387–1398

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K (2002) Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. Plant J 29:417–426

Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37:914–939

Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. Annu Rev Plant Physiol Plant Mol Biol 50:571–599

Uemura M, Warren G, Steponkus PL (2003) Freezing sensitivity in the sfr4 mutant of Arabidopsis is due to low sugar content and is manifested by loss of osmotic responsiveness. Plant Physiol 131:1800–1807

Vogel JT, Zarka DG, Van Buskirk HA, Fowler SG, Thomashow MF (2005) Roles of the CBF2 and ZAT12 transcription factors in configuring the low temperature transcriptome of Arabidopsis. Plant J 41:195–211

Weckwerth W, Wenzel K, Fiehn O (2004) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics 4:78–83

Wurtele ES, Li J, Diao L, Zhang H, Foster CM, Fatland B, Dickerson JA, Brown A, Cox Z, Cook D, Lee E-K, Hofmann H (2003) MetNet: Software to build and model the biogenetic lattice of Arabidopsis. Comp Funct Genom 4:239–245

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Li J, Liu Z, Qi Q, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Zhao W, Li P, Chen W, Zhang Y, Hu J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Tao M, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science 296:79–92

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

Zuther E, Buchel K, Hundertmark M, Stitt M, Hincha DK, Heyer AG (2004) The role of raffinose in the cold acclimation response of *Arabidopsis thaliana*. FEBS Lett 576:169–173