Title

Plant Metabolic Network 16: expansion of underrepresented plant groups and experimentally supported enzyme data.

Journal

Nucleic Acids Research, 53(D1)

Authors

Hawkins, Charles

Xue, Bo

Yasmin, Farida

et al.

Publication Date

2025-01-06

DOI

10.1093/nar/gkae991

Peer reviewed

Nucleic Acids Research, 2025, **53**, D1606–D1613
https://doi.org/10.1093/nar/gkae991
Advance access publication date: 18 November 2024
Database issue

Plant Metabolic Network 16: expansion of underrepresented plant groups and experimentally supported enzyme data

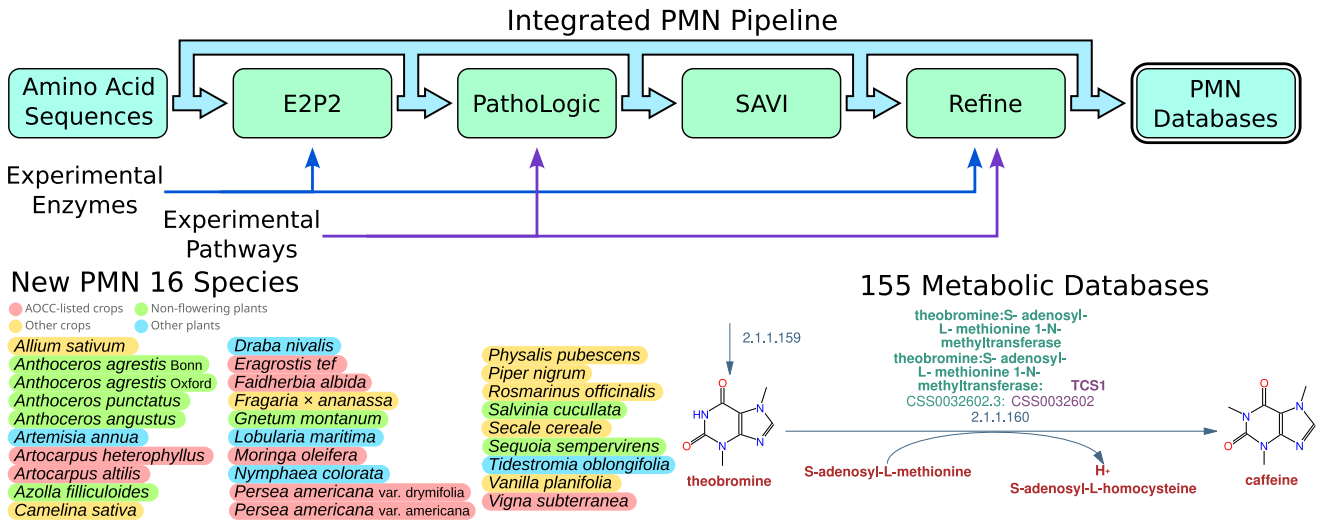
Charles Hawkins[1,2], **Bo Xue**[1,2], **Farida Yasmin**[3], **Gabrielle Wyatt**[3], **Philipp Zerbe**[3] and **Seung Y. Rhee**[1,2,4,5,*]

[1]Plant Resilience Institute, Michigan State University, 1066 Bogue St, East Lansing, MI 48824, USA
[2]Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Road, East Lansing, MI 48824, USA
[3]Department of Plant Biology, University of California-Davis, 1 Shields Avenue, Davis, CA 95616, USA
[4]Department of Plant Biology, Michigan State University, 612 Wilson Road, East Lansing, MI 48824, USA
[5]Department of Plant, Soil, and Microbial Sciences, Michigan State University, 1066 Bogue St, East Lansing, MI 48824, USA
*To whom correspondence should be addressed. Email: rheeseu6@msu.edu

Abstract

The Plant Metabolic Network (PMN) is a free online database of plant metabolism available at https://plantcyc.org. The latest release, PMN 16, provides metabolic databases representing >1200 metabolic pathways, 1.3 million enzymes, >8000 metabolites, >10 000 reactions and >15 000 citations for 155 plant and green algal genomes, as well as a pan-plant reference database called PlantCyc. This release contains 29 additional genomes compared with PMN 15, including species listed by the African Orphan Crop Consortium and nonflowering plant species. Furthermore, 52 new enzymes with experimentally supported function information have been included in this release. The single-species databases contain a combination of experimental information from the literature and computationally predicted information obtained through PMN's database generation pipeline for a single species, while PlantCyc contains only experimental information but for any species within *Viridiplantae*. PMN is a comprehensive resource for querying, visualizing, analyzing and interpreting omics data with metabolic knowledge. It also serves as a useful and interactive tool for teaching plant metabolism.

Graphical abstract



Introduction

Plant metabolism is fundamental to addressing major global challenges related to the climate crisis, pollution, agriculture, food safety, energy and healthcare. None of the currently used crops have been bred to withstand the variability and extremes of the current and projected climate in the near future. To

Received: August 19, 2024. **Revised:** October 8, 2024. **Editorial Decision:** October 10, 2024. **Accepted:** October 15, 2024

understand how plants respond to changing climates and identify and engineer adaptive traits rapidly, we need to reveal the genome sequence and metabolic capacity of a large diversity of species and populations. Advances in omics technologies and the growing application of these tools in plants rapidly increased the body of plant genome data. Most of the biological knowledge, however, exists in individual publications, limiting its access and use.

To enable a comprehensive understanding of plant metabolism and facilitate metabolic engineering and crop breeding efforts, the Plant Metabolic Network (PMN) was created to serve as a central resource for genome-scale annotations of metabolic enzymes, pathways and metabolites in plants and algae (1). PMN is an integrative metabolic database resource for 155 plant and algal genomes and a pan-plant reference database, PlantCyc. PMN has become a valuable resource, which is used widely by genome annotators, metabolic modelers and plant biologists. PMN has several unique attributes. First, the pipeline used to generate Pathway Genome Databases (PGDBs) can be scaled up to any number of genomes. Second, compared with other pathway databases such as Plant Reactome (2) and KEGG (3), PMN includes more experimentally validated reactions (1). Third, PMN has a large user base: 5719 registered users and a mailing list with 1324 members. PMN is used by a wide range of scientists and students as a general reference for plant metabolism, to gain metabolic insights from omics datasets, to aid in annotating new genomes and to answer questions about the evolution of plant metabolism.

Despite housing metabolic data for many species, PMN's coverage of certain plant groups is limited. In addition, there is still a very small portion of experimentally supported enzyme data. In this paper, we describe PMN 16 that introduces 29 new single-species databases from underrepresented plant groups as well as the addition of literature-curated, experimentally supported enzyme data.

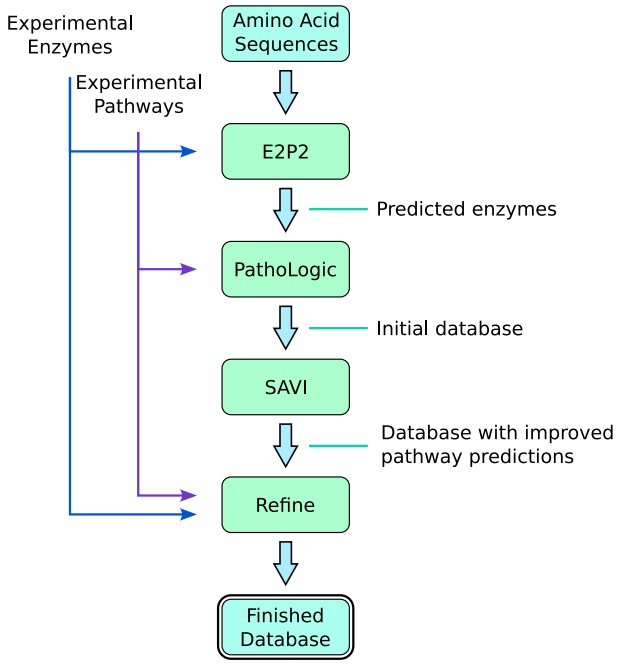


Figure 1. The PMN pipeline. An overview of the major steps in the PMN pipeline, used by the PMN team to generate each PMN single-species database. Amino acid sequences for the species are fed into E2P2 that predicts enzymes. PathoLogic then predicts pathways and generates the initial database. SAVI refines the predictions using a set of curator-defined rules. The refine steps add experimental data for the species and perform other finishing refinements.

Materials and methods

An advanced pipeline to generate PMN databases

The PMN databases are created using a bioinformatics pipeline (Figure 1). The pipeline takes the full set of protein sequences from an annotated genome for the organism and produces a PGDB. The principal stages are (i) the ensemble enzyme prediction pipeline (E2P2), which predicts enzymes from the protein sequences; (ii) PathoLogic, which uses the predicted enzymes to call presence/absence on each pathway in the reference database and creates the initial single-species database; and (iii) the semi-automated validation infrastructure (SAVI), which applies past curation decisions at the pathway level to each new database, serving as guardrails against highly implausible pathway predictions.

E2P2 (https://github.com/carnegie/E2P2) is an ensemble enzyme prediction framework produced by the PMN team. The new E2P2 version 5.0 was used to create PMN 16 and was released publicly alongside it. E2P2 works by combining the predictions of other classifier software, with each enzyme class weighted for each classifier according to its performance under cross-validation (4). Version 5.0 makes E2P2 modular by allowing the user to define new external classifiers by writing a configuration file and to select which of the available classifiers to use at runtime (in previous versions of E2P2, the classifiers were hardcoded and could not be changed without editing the E2P2 source code). E2P2 5.0 also removes PRIAM (5) as a default classifier and replaces it with DeepEC (6), which shows similar performance under cross-validation. The change was made because the PRIAM software is no longer maintained. BLAST remains the other default classifier, as it was in E2P2 4.0.

The predictions of E2P2 are made based on a reference protein sequence dataset (RPSD) assembled by the PMN team from various sources. It consists of protein sequences and the set of reactions each protein is known to catalyze based on experimental evidence. Additionally, a high-quality, trusted set of nonenzyme protein sequences is included. RPSD is assembled from MetaCyc (7), PlantCyc (1), GO (8,9), BRENDA (10), Expasy (11), TAIR (12) and UniProt (13). RPSD 5.2 was released alongside E2P2 5.0. RPSD 5.2 contains 169 251 enzymes and 17 002 enzymatic reactions ('enzyme function classes' in E2P2 terminology).

Generation of the initial databases is performed by PathoLogic, a function of the Pathway Tools suite produced by SRI International (14). PMN 16 was built using Pathway Tools 28.0. Each single-species database is built from reference databases as a source of pathway, reaction, and compound data. PlantCyc is used as the primary reference for all PMN databases and MetaCyc is used as an additional source. The function of PathoLogic in the PMN pipeline is to call presence or absence of each metabolic pathway in the target genome from the reference database on the basis of the pathway's complement of predicted enzymes. PathoLogic then imports the necessary reactions, compounds and other data to construct an initial database.

The last major step in the PMN pipeline is SAVI. SAVI is a piece of software developed by the PMN team to apply past curation decisions at the pathway level to a newly created single-species database. Each pathway that is predicted in any plant species is classified into five validation lists by the PMN curators. These lists are UPP (universal plant pathways), for pathways that are expected to be universal to all embryophytes and should be added if not predicted; CVP (common *Viridiplantae* pathways), similar to UPP but for chlorophytes (in spite of the name referring to *Viridiplantae* for historical reasons); NPP (nonplant pathways) for pathways not found in plants or green algae that should be removed if predicted; AIPP (accepted-if-predicted pathways) for pathways whose prediction or nonprediction by PathoLogic should be left as is; and CAPP (conditionally accepted plant pathways), for cases where a pathway should be removed unless it is within a specific taxon specified by the curators. SAVI is run on each new database and recommends pathways to be added to and removed from the initial database.

After SAVI comes a series of minor refining steps that apply SAVI's recommendations automatically, correct the naming of enzymatic reactions, add enzyme links to Phytozome (15) and other databases for genomes from those sources, add authorship information, add citations for E2P2 and SAVI, import pathways and enzymes from the reference databases that have experimental evidence for the target species, run Pathway Tools' consistency checker and generate the cellular overview diagram.

Past work has evaluated the results of this pipeline and found it to have good accuracy. Cross-validation on the results of E2P2 (5-fold cross-validation using the RPSD) previously found an *F*1 score of 0.735 for enzyme prediction (4). Subsequent work examined the accuracy of the final databases at the pathway level by having curators manually assign a randomly selected set of 120 pathways to their expected phylogenetic range based on the literature and comparing the results with the predicted phylogenetic range (1). Seventy-eight percent of the pathways' predicted range matched that assigned by the curators.

PMN 16 updates to streamline pipeline use

To build PMN 16, the pipeline as a whole was re-engineered. The previous version of the pipeline consisted of many steps that had to be executed manually by the user. There was a large amount of manual work involved in the process at many points, and the code was scattered among many scripts that had to be called individually by the user. The same information, such as the name and location of config files and executables, had to be entered multiple times in different steps. The new pipeline integrates everything using a single front-end script. Much of the manual work has been eliminated, and what remains is mostly front-loaded to the start of the pipeline, eliminating inefficient alternation between manual and automated steps. More importantly, all remaining manual steps are in the form of writing configuration files rather than direct manipulation of the data and software controls. The use of these config files greatly improves the reproducibility of results from the pipeline because the config files can be re-run to produce identical output. The pipeline can easily be re-run to adjust a parameter or correct a mistake without repeating the manual work. The pipeline has also been packaged into a Singularity container so that its results can be reproduced reliably on computer systems with different configurations. The new pipeline also uses much simpler commands and supports parallel execution of many steps that previously were obligately single-threaded. Finally, the pipeline features a presets system that allows common sets of configuration options to be applied to many databases at once without having to enter them repeatedly.

Results

PMN 16

The updated PMN 16 consists of 155 single-species databases (up from 126 species in PMN 15), each of which presents the metabolism of a single plant or green algal species, based on a combination of computational predictions and evidence curated from the literature. In addition, the network hosts PlantCyc, a pan-plant database containing only experimental information from any species within *Viridiplantae*. A total of 517 species are represented in PlantCyc by at least one enzyme or pathway. The principal objects stored by the PMN database are enzymes, reactions, compounds and metabolic pathways. PMN 16 has, across all single-species databases and PlantCyc, 1297 pathways, 10 389 reactions, 8189 compounds, 15 199 external publications cited and 1308 907 enzymes. The breakdown of the counts of these database objects for each database in PMN 16 is given in Supplementary Table S1, and a comparison with all prior PMN releases is given in Supplementary Table S2. The count of enzymes is much higher than the other metrics because unlike the other data types in the databases, each enzyme in a species is a different object (regardless of homology).

Increased coverage of underrepresented species

To the single-species databases already present in PMN 15, PMN 16 adds 29 new plant species and varieties (Supplementary Table S1), each with its own genome-wide computationally predicted database. Well-studied crop and model species like *Arabidopsis thaliana*, maize and soybean were already well served by early releases of PMN, but important crop species in the Global South have been underserved both by plant biology as a field and by PMN. Nonflowering plants have also been understudied and underrepresented in PMN in the past. To begin to address this shortcoming, the PGDBs for seven species listed by the African Orphan Crop Consortium, *Artocarpus altilis*, *Artocarpus heterophyllus*, *Moringa oleifera*, *Eragrostis tef*, *Faidherbia albida*, *Vigna subterranea* and *Persea americana* (two databases for this species, one for the americana and one for the drymifolia variety); and seven nonflowering plant species, *Anthoceros agrestis* (two varieties, Bonn and Oxford), *Anthoceros angustus*, *Anthoceros punctatus*, *Azolla filiculoides*, *Salvinia cucullata*, *Sequoia sempervirens* and *Gnetum montanum*, were added to PMN 16. These new metabolism databases can facilitate research into these species, as well as draw attention to and drive interest in studying them from the broader plant biology community.

Integration of experimental enzyme function data

The inclusion of experimentally validated enzyme function data is critical to improve the accuracy of enzyme function predictions and hence the overall quality of databases. How-

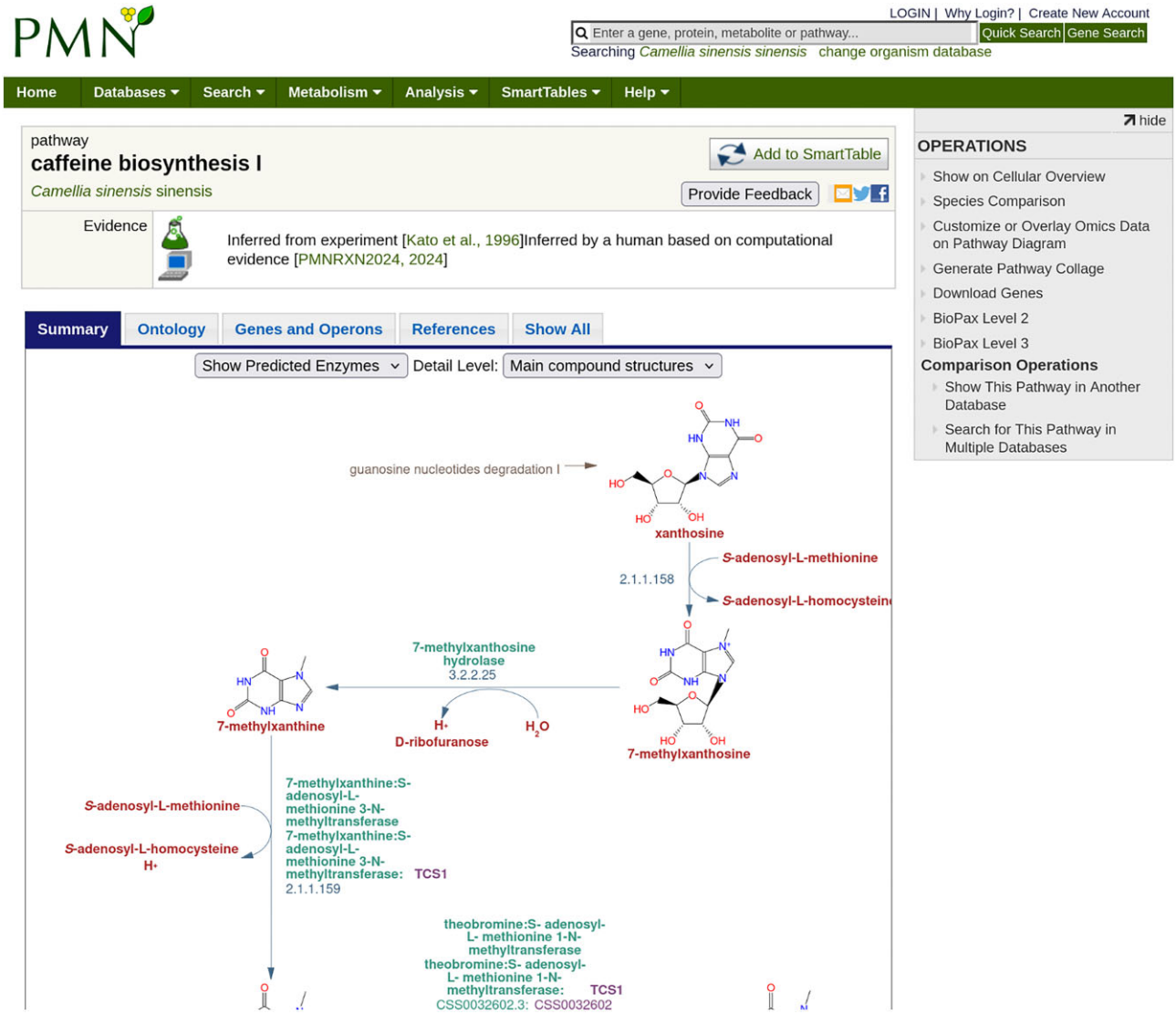


Figure 2. Example pathway view. Screenshot showing part of the caffeine biosynthesis pathway from TeaCyc (*Camellia sinensis*). The pathway structure is presented in the center; compounds, reactions and enzymes may be clicked to view details of those objects. Several operations are available in the Operations menu on the right. More functionality is under the Search, Metabolism, Analysis and SmartTables menus.

ever, while enzyme functions of general (primary) metabolism can typically be predicted with high reliability, computational functional annotation of enzymes of plant specialized (secondary) metabolism poses a significant challenge due to the large enzyme family sizes, high sequence identity and high functional diversity.

To address this issue, experimental enzyme functional data need to be expanded. To this end, PMN 16 introduces a new system for collaborators to submit experimental enzyme function data, using spreadsheets rather than requiring the use of the Pathway Tools software, which requires training. Using an initial trial of this system, we integrated 52 new enzyme data into PMN 16. This includes enzyme data for 39 terpene synthase (TPS) and 13 cytochrome P450 monooxygenase (P450) enzymes, representing core enzymes of general and specialized terpenoid metabolism. This dataset includes enzyme families that are challenging to functionally predict due to their high functional diversity and catalytic promiscuity (16,17). Fourteen TPSs and P450s were added to CornCyc (*Zea mays*) (18–20), 19 enzymes were added to SetariaCyc (*Setaria italica*) (21) and 19 TPSs and P450s were added to SwitchgrassCyc (*Panicum virgatum*) (22,23). All of these new enzymes were entered first into their respective single-species databases and then imported into PlantCyc. In addition to the enzymes currently appearing directly in the PMN databases, they will also be incorporated into the next RPSD (see 'Materials and Methods' section) and thereby contribute to E2P2's ability to predict additional enzymes with the same function in future PMN releases.

PMN information access and data analysis tools

PMN allows users to search and browse each database for enzymes, reactions, metabolites and metabolic pathways, which are presented in a visual format (Figure 2). PMN uses the Pathway Tools software (14) for generating the databases and presenting them in web-accessible form. Each data type links to related data; for example, a pathway page will link to the reactions, compounds and enzymes that make it up. An enzyme will link to the reactions it catalyzes. All pathways contain a

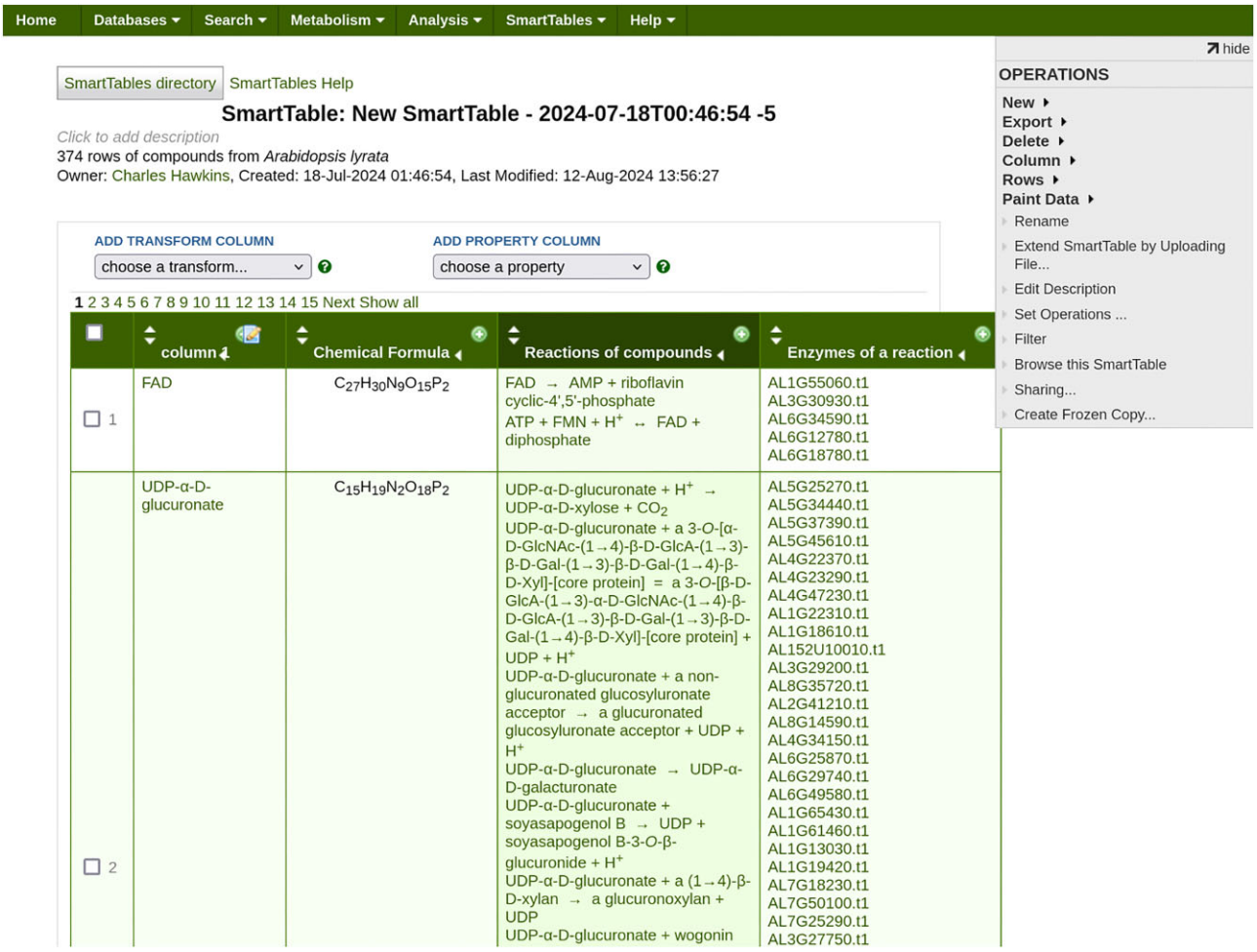


Figure 3. Example SmartTable. Screenshot showing a SmartTable of compounds from AlyrataCyc (*Arabidopsis lyrata*). From the initially uploaded list of compounds, a column for chemical formula was added, then a list of reactions involving each compound, then from there a list of enzymes catalyzing any of those reactions. This illustrates the utility of SmartTables in pulling bulk information out of PMN and transforming metabolic datasets.

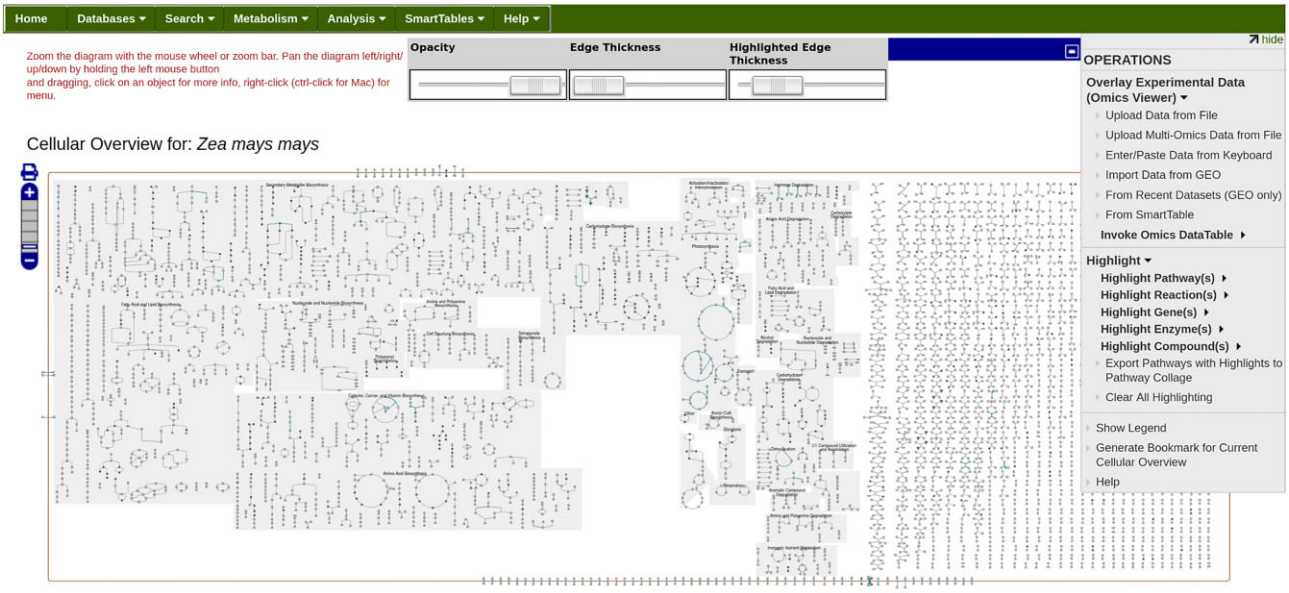


Figure 4. Cellular overview. Screenshot showing the cellular overview for CornCyc (*Z. mays*). The diagram shows all pathways in the selected PGDB. The view may be zoomed in for more detail. Uploaded omics data may be overlaid and shown using a color scale.

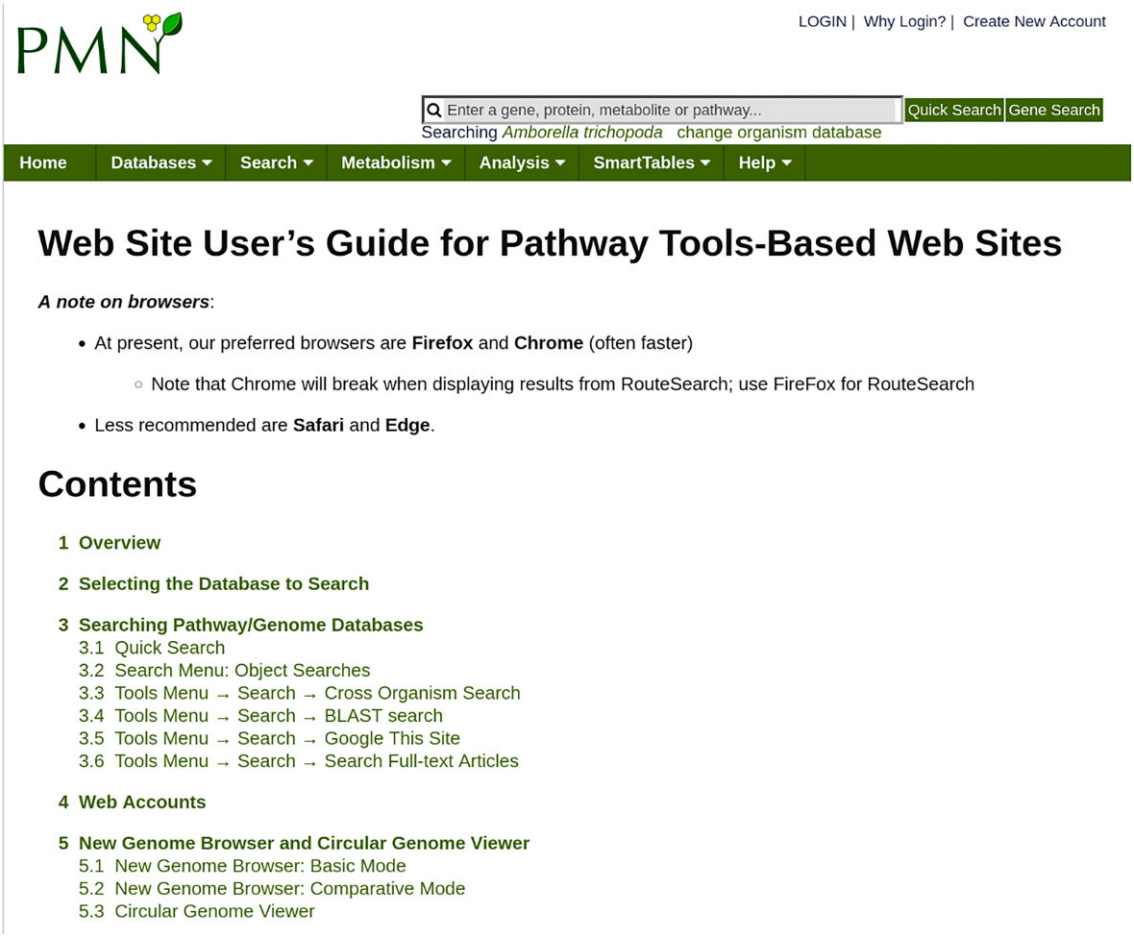


Figure 5. Website user's guide. The user's guide, accessible from the Help menu, details the functions of the website in a single, easily searchable document.

summary written by curators. Citations for relevant facts are presented, and icons indicate whether a given fact, such as that an enzyme catalyzes a reaction, is based on experimental evidence or computational prediction.

SmartTables is a functionality that allows users to create lists of database objects such as pathways, compounds, enzymes or reactions, and operate on them in bulk (Figure 3). Users can create these tables by uploading a list of various kinds of identifiers or assemble the tables manually. Once created, SmartTables can be used to get database fields for any object, such as by adding a column to a compound smart table with the SMILES code or molecular weight.

For example, a list of metabolites can be uploaded from a mass spectrometry experiment to access metabolite properties such as monoisotopic mass, chemical formula or PubChem ID and re-export these data as a table for all metabolites of interest. The user can also pull a list of enzymes that catalyze specific reactions involving those metabolites, transforming a metabolite list to an enzyme list using the 'Transform' function. This function allows users to add a column of related database objects and, if desired, turn the column into its own SmartTable. To examine broader metabolic trends in the data, the user can then perform a pathway enrichment analysis, which will result in a list of pathways and higher level classes of pathways that contain more of their metabolites than would be expected by chance for a set of metabolites of that size. SmartTables therefore enable large-scale operations to be performed using PMN and its data, and are a useful tool for generating new insights from an existing dataset. PMN 16 brings these abilities to all the newly released plant species.

The Cellular Overview tool (Figure 4) presents an overview of all pathways in the organism. Users can upload omics data (any numeric data associated with enzymes, reactions or compounds) and paint it onto the overview so that objects shown on the diagram will be colored according to the user's data. Such omics data can also be used to calculate a pathway perturbation score (PPS) to highlight pathways with the most extreme overall values for the uploaded enzyme, compound or reaction-associated data. Due to updating to Pathway Tools 28.0, PMN 16 now allows the painting of separate data (e.g. multi-omics data) onto node size, node color, line size and line color simultaneously on the cellular overview diagram.

PMN also includes a Pathway Co-expression Viewer tool, which is not included in the baseline Pathway Tools software. This function is available from the Operations menu on the pathway pages for nine species featured in ATTED-II: AraCyc (*Arabidopsis thaliana* Col), MtruncatulaCyc (*Medicago truncatula*), TomatoCyc (*Solanum lycopersicum*), Brapa_FpscCyc (*Brassica rapa*), OryzaCyc (*Oryza sativa* japonica), GrapeCyc

(*Vitis vinifera*), SoyCyc (*Glycine max*), PoplarCyc (*Populus trichocarpa*) and CornCyc (*Zea mays*). This tool pulls in data from the ATTED-II gene co-expression database (24) to show co-expression of all genes in the pathway.

Many other functions are available; PMN contains an online help document describing all functionality in detail (Figure 5; https://pmn.plantcyc.org/PToolsWebsiteHowto.shtml).

Discussion

PMN is the largest collection of plant metabolic databases in the world, and continues to grow. It is a significant resource for research involving plant metabolism and the underlying metabolites, enzymes and biosynthetic/degradation pathways. PMN serves to consolidate information that is otherwise dispersed throughout the literature while also extending knowledge from experimental data to provide genome-wide predicted metabolism data for now 155 *Viridiplantae* genomes. PMN is used as a general reference for specific information about pathways, metabolites, enzymes and reactions of interest; to transform datasets such as by converting a list of compounds from a metabolomics experiment into a list of enzymes in the species that catalyze reactions involving those compounds; and to highlight pathways or broad areas of metabolism that are implicated by a given dataset, such as by performing a pathway enrichment on a set of enzymes or calculating a PPS for numeric data associated with such a set.

The new pipeline represents a significant improvement in efficiency and usability over its previous versions. It will be used as the foundation for future PMN releases and presents the opportunity to scale up PMN in a way that would have been infeasible with the previous semi-manual pipeline. The next PMN release is expected to be significantly larger, in terms of the number of genomes and enzyme functions added, than any previous release.

Centralized resources like PMN integrate omics data into a comprehensive database of biological knowledge and a genome-scale gene annotation framework. These resources can greatly accelerate scientists' ability to understand complex data to inform their research and advance science and engineering.

Data availability

All PMN databases can be accessed online for free by all users without registration at https://plantcyc.org. All PMN databases are available for download upon receipt of a free license, which may be requested via the PMN website; all users are eligible for a free license regardless of affiliation. All downloads include both the 'native' representation of the database (.ocelot format), which can be loaded into the Pathway Tools software, and text-based 'flat file' dumps of most database data in tabular and key-value formats, as well as BioPAX level 3 dumps of data representable using that format. The source code for E2P2 version 5.0 is available at https://github.com/carnegie/E2P2 and https://doi.org/10.6084/m9.figshare.27214293.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

Much of the pipeline was run on the high-performance computing cluster maintained by the Institute for Cyber-Enabled Research (ICER) at Michigan State University. Hiroshi Maeda and Marcos Oliveira at the University of Wisconsin-Madison contributed to editing of the manuscript and testing of the enzyme submission form. Gaëlle Cassin-Ross at Michigan State University contributed administrative and logistical support, editing material on the PMN website and coordinating outreach related to PMN.

Funding

U.S. National Science Foundation [IOS-2312181 to S.Y.R., P.Z. and H.M., IOS-2406533 to S.Y.R., IOS-1546838 to S.Y.R., DBI-2213983 to S.Y.R., in part]; Department of Energy's Office of Biological and Environmental Research [DE-SC0018277 to S.Y.R., DE-SC0020366 to S.Y.R., DE-SC0023160 to S.Y.R., DE-SC0021286 to S.Y.R., in part]. Funding for open access charge: U.S. National Science Foundation and Michigan State University.

Conflict of interest statement

None declared.

References

1. Hawkins,C., Ginzburg,D., Zhao,K., Dwyer,W., Xue,B., Xu,A., Rice,S., Cole,B., Paley,S., Karp,P., *et al.* (2021) Plant Metabolic Network 15: a resource of genome-wide metabolism databases for 126 plants and algae. *J. Integr. Plant Biol.*, **63**, 1888–1905.
2. Naithani,S., Gupta,P., Preece,J., D'Eustachio,P., Elser,J.L., Garg,P., Dikeman,D.A., Kiff,J., Cook,J., Olson,A., *et al.* (2020) Plant Reactome: a knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res.*, **48**, D1093–D1103.
3. Kanehisa,M., Furumichi,M., Sato,Y., Kawashima,M. and Ishiguro-Watanabe,M. (2023) KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.*, **51**, D587–D592.
4. Schläpfer,P., Zhang,P., Wang,C., Kim,T., Banf,M., Chae,L., Dreher,K., Chavali,A.K., Nilo-Poyanco,R., Bernard,T., *et al.* (2017) Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.*, **173**, 2041–2059.
5. Claudel-Renard,C. (2003) Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.*, **31**, 6633–6639.
6. Wang,H., Zhang,H., Hu,J., Song,Y., Bai,S. and Yi,Z. (2020) DeepEC: an error correction framework for dose prediction and organ segmentation using deep neural networks. *Int. J. Intell. Syst.*, **35**, 1987–2008.
7. Caspi,R., Billington,R., Keseler,I.M., Kothari,A., Krummenacker,M., Midford,P.E., Ong,W.K., Paley,S., Subhraveti,P. and Karp,P.D. (2020) The MetaCyc database of metabolic pathways and enzymes—a 2019 update. *Nucleic Acids Res.*, **48**, D445–D453.
8. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T., *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
9. The Gene Ontology Consortium, Aleksander,S.A., Balhoff,J., Carbon,S., Cherry,J.M., Drabkin,H.J., Ebert,D., Feuermann,M., Gaudet,P., Harris,N.L., *et al.* (2023) The Gene Ontology knowledgebase in 2023. *Genetics*, **224**, iyad031.
10. Chang,A., Jeske,L., Ulbrich,S., Hofmann,J., Koblitz,J., Schomburg,I., Neumann-Schaal,M., Jahn,D. and Schomburg,D. (2021) BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Res.*, **49**, D498–D508.

11. Duvaud,S., Gabella,C., Lisacek,F., Stockinger,H., Ioannidis,V. and Durinx,C. (2021) Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Res.*, **49**, W216–W227.
12. Reiser,L., Bakker,E., Subramaniam,S., Chen,X., Sawant,S., Khosa,K., Prithvi,T. and Berardini,T.Z. (2024) The Arabidopsis Information Resource in 2024. *Genetics*, **227**, iyae027.
13. The UniProt Consortium, Bateman,A., Martin,M.-J., Orchard,S., Magrane,M., Ahmad,S., Alpi,E., Bowler-Barnett,E.H., Britto,R., Bye-A-Jee,H., *et al.* (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
14. Karp,P.D., Midford,P.E., Billington,R., Kothari,A., Krummenacker,M., Latendresse,M., Ong,W.K., Subhraveti,P., Caspi,R., Fulcher,C., *et al.* (2021) Pathway Tools version 23.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.*, **22**, 109–126.
15. Goodstein,D.M., Shu,S., Howson,R., Neupane,R., Hayes,R.D., Fazo,J., Mitros,T., Dirks,W., Hellsten,U., Putnam,N., *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.
16. Banerjee,A. and Hamberger,B. (2018) P450s controlling metabolic bifurcations in plant terpene specialized metabolism. *Phytochem. Rev. Proc. Phytochem. Soc. Eur.*, **17**, 81–111.
17. Karunanithi,P.S. and Zerbe,P. (2019) Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front. Plant Sci.*, **10**, 1166.
18. Ding,Y., Weckwerth,P.R., Poretsky,E., Murphy,K.M., Sims,J., Saldivar,E., Christensen,S.A., Char,S.N., Yang,B., Tong,A., *et al.* (2020) Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nat. Plants*, **6**, 1375–1388.
19. Murphy,K.M., Ma,L.-T., Ding,Y., Schmelz,E.A. and Zerbe,P. (2018) Functional characterization of two class II diterpene synthases indicates additional specialized diterpenoid pathways in maize (*Zea mays*). *Front. Plant Sci.*, **9**, 1542.
20. Murphy,K.M., Dowd,T., Khalil,A., Char,S.N., Yang,B., Endelman,B.J., Shih,P.M., Topp,C., Schmelz,E.A. and Zerbe,P. (2023) A dolabralexin-deficient mutant provides insight into specialized diterpenoid metabolism in maize. *Plant Physiol.*, **192**, 1338–1358.
21. Karunanithi,P.S., Berrios,D.I., Wang,S., Davis,J., Shen,T., Fiehn,O., Maloof,J.N. and Zerbe,P. (2020) The foxtail millet (*Setaria italica*) terpene synthase gene family. *Plant J.*, **103**, 781–800.
22. Muchlinski,A., Jia,M., Tiedge,K., Fell,J.S., Pelot,K.A., Chew,L., Davisson,D., Chen,Y., Siegel,J., Lovell,J.T., *et al.* (2021) Cytochrome P450-catalyzed biosynthesis of furanoditerpenoids in the bioenergy crop switchgrass (*Panicum virgatum* L.). *Plant J.*, **108**, 1053–1068.
23. Pelot,K.A., Chen,R., Hagelthorn,D.M., Young,C.A., Addison,J.B., Muchlinski,A., Tholl,D. and Zerbe,P. (2018) Functional diversity of diterpene synthases in the biofuel crop switchgrass. *Plant Physiol.*, **178**, 54–71.
24. Obayashi,T., Hibara,H., Kagaya,Y., Aoki,Y. and Kinoshita,K. (2022) ATTED-II v11: a plant gene coexpression database using a sample balancing technique by subagging of principal components. *Plant Cell Physiol.*, **63**, 869–881.

Received: August 19, 2024. **Revised:** October 8, 2024. **Editorial Decision:** October 10, 2024. **Accepted:** October 15, 2024