PlantCyc, AraCyc, PoplarCyc and more . . .

Building databases and connecting to researchers at the Plant Metabolic Network

kate dreher

curator

PMN/TAIR

Introducing the PMN



PMN databases

- Current PMN databases: PlantCyc, AraCyc, PoplarCyc
 - Coming soon: databases for wine grape, maize, cassava, Selaginella, and more . . .
- Other plant databases accessible from the PMN:

PGDB	Plant	Source	Status
RiceCyc **	Ri P60. Co	nnecting the Solaneceae g	enome to the
SorghumCyc	Sc metabo	ic networks via SolCyc an	d MetaCyc.
MedicCyc **	Me	Anuradha Pujar	
LycoCyc **	Tomato	Sol Genomics Network	some curation
PotatoCyc	Potato	Sol Genomics Network	no curation
СарСус	Pepper	Sol Genomics Network	no curation
NicotianaCyc	Tobacco	Sol Genomics Network	no curation
PetuniaCyc	Petunia	Sol Genomics Network	no curation
CoffeaCyc	Coffee	Sol Genomics Network	no curation

** Significant numbers of genes from these databases have been integrated into PlantCyc

PMN data types





PMN database content statistics

- New PMN release . . . almost here!
 - Increased data content
 - shhh... sneak preview at pmn.plantcyc.org!

	PlantCyc 4.0	AraCyc 7.0	PoplarCyc 2.0
Pathways	685	369	288
Enzymes	11058	5506	3420
Reactions	2929	2418	1707
Compounds	2966	2719	1397
Organisms	343	1	1*

Upgraded Pathway Tools software (version 12.5 -> version 14.0)

Searching in PMN databases

PMN quick search bar

			Choline		PlantOyc 👻	search
About PMN Sea	rch Too	ls Downloads	Useful Sites	Submit Data	AreQyc PoplarCyc	Feedback

Pathway Tools quick search bar

PMN	Select a Database:	Choline Search Database PlantCyc <u>change</u>	Quick Search
	Arabidopsis thaliana col PlantCyc Populus trichocarpa		

Searching in PMN databases



Specific search pages

PMN	P q	and the second	ten de	1-1	Search Database	PlantCyc <u>cha</u>	Quick	Search
About PMN	Search	Tools	Downloads	Useful Sites	Submit Data	Help	Feedback	
	Databas	es Overview nds						
PlantCy	c Comp	oound Se	arch	lear Form	All Compounds			
	compound by r	same or ID						
Briver a compour LIGAND, PubChe generate a rubit Exempler: Nryp	nd name, or a data en or CAS. This dat tring search on com tophan", "C08036"	base identifier from th tabase may not contain pound names only (not	it database or firsm an e mappings to all of these on database identifiers)	internal darabaixe roch ac DN other darabaixer. Partial nan	IDI, No vili			
Search/Filt	ter by ontology	and the second second					(isa	ctive)
Search/Filt	ter by molecula	r weight					(ina	ctive)
► Search/Fit	ter by chemical	formula (partial	or full)				(ina	ctive)
Search by I	nchi string						{ina	ctive)
		5	ubmit Query C	lear Form List of	All Compounds			

Additional search options

Advanced Query Results

Your query in <u>Biologic</u> is html-sect-accending([[21"HAME, 21"HEACTION-LEFT, [e*HEC-WOMEER: e <- 12]) + 21C-ADA**Pathways, 12 += [22 + 22 <- 21"REACTION-LEFT, [+ 0 < #[23 : 25 <- 22"LEFT, (*C10" instaings: ID*CHERECAL-FORMULA)]) + (*2.4.1." instaings: 22"EC-MOMEER()], 0 < #12),1) [dick the guery to edit if in the Free Form Advanced Query Page)

This query resulted in a single table of 64 rows.

HAME A T	Reaction-List A 🕈	[(e ^? EC #UMDER) : e ~ [2] <u>\</u> \ \
abscicle acid glucosa ester biosynthesis	(+)-abscisate + UDP-D-glacose + abscisic acid glacose ester + UDP	2.4.1.1
ajugose biosynthesis i (galactinol dependent)	verbascose + galactinol + ajugose + myo-inositol, stachyose + galactinol + verbascose + myo-inositol	2.4.1,-
ajugose blenynthesis II (galactinol-independent)	2 raffinose = stachyose + sucrese, 2 verbascose = ajugese + stachyose, 2 stachyose + verbascose + raffinose	2.4.1 2.4.1 2.4.1
anthocyanih biosynthiasis (polargonidin 3-O-glucoside, cyanidin 3-O-glucoside)	$\label{eq:conversion} \begin{split} & \text{exconversion} + \alpha \cdot \text{ketsglutarate} + 0_2 = a \ \text{dihydroquercetin} + \text{succinate} + C0_2 + H_20, \ \text{UDP-D-glucose} + \text{synhidin} + \text{UDP} + \text{synhidin} - 3-0-8-0-glucoside, \\ & \text{leucosynhidin} + \alpha \cdot \text{ketsglutarate} + 0_2 + \text{synhidin} + \text{succinate} + C0_2 + 2\text{H}_20 + \text{H}^4, \ \text{UDP-D-glucose} + \text{pelargonidin} + \text{UDP} + \text{pelargonidin} - 3-0-8-0-glucoside, \\ & \text{D-glucoside} + \text{H}^4, \ \text{leucopelargonidin} + \alpha \cdot \text{ketsglutarate} + 0_2 + \text{H}^4 + \text{pelargonidin} + \text{succinate} + C0_2 + 2\text{H}_20 \end{split}$	2.4.1.115 2.4.1.115
arabidiol biorynthesis	(5)-2,3-Epoxytiqualene + H ₂ O + arabidiol	4.2.1
baruol biorynthesis	(5)-2,3-Epoxysigualene + baruot	5.4.99
berapyl-CoA biosynthesis	3-keto-3-phenylproplonyl-CoA + coenzyme A + benzzyl-CoA + acetyl-CoA, (E)-cinnamoyl-CoA + H ₂ O + 3-hydroxy-3-phenylproplonyl-CoA, 3-hydroxy- 3-phenylproplonyl-CoA + NADP [*] + 3-keto-3-phenylproplonyl-CoA + NADPH + H ⁴	
cellulose blorynthesis	a long-linear glucan + HyD + a short glucan + a short glucan, UDP-D-glucese + (1.4-8-D-glucesy() _(b) + UDP + [1,4-8-D-glucesy() _(b) +)	2.4.1.12
cholesterol biosynthesis (lanosterol biosynthesis, 4.ocarboxy-48-methyl-50-cholesta-8,24-dien-38-oi = NAD(P) ⁴ = 4.omethyl-50-cholesta-8,24-dien-3-one = NAD(P)H + CO ₂ , lathosterol = NAD(P)H + O ₂ = 7-dehydro-cholesterol = NAD(P) ⁴ = 2 H ₂ O, 4,4-dimethyl-5-o-cholesta-8,14,24-trien-3-8-ol = NADPH + H ⁴ = 4,4-dimethyl-50-cholesta-8,24-dien-38-ol = NADP ⁴ , eposystqualene biosynthesis, cholesterol = NADP ⁴ = 7-dehydro-cholesterol = NADPH + H ⁴ = 50-cholesta-7,24-dien-38-ol = NADP ⁴ , eposystqualene biosynthesis, cholesterol = NADP ⁴ = 7-dehydro-cholesterol = NADPH + H ⁴ , 50-cholesta-7,24-dien-38-ol = NADP ⁴ , eposystqualene biosynthesis, cholesta-7,24-dien-38-ol, 50-cholesta-8,24-dien-30-ore = NADPH + H ⁴ = zymosterol = NADP ⁴ , 40-carboxy-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 50-cholesta-8,24-dien-38-ol, 50-cholesta-8,24-dien-30-ore = NADPH + H ⁴ = zymosterol = NADP ⁴ , 40-carboxy-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 50-cholesta-8,24-dien-36-ol = NAD(P)H + CO ₂ , 40-formyl- 50-cholesta-8,24-dien-38-ol = NAD(P)H = O ₂ = 40-carboxy-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40-hydroxymethyl-50-cholesta- 8,24-dien-38-ol = NAD(P)H = O ₂ = 40-carboxy-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 2 H ₂ O, 40-methyl zymosterol = NAD(P)H = O ₂ = 40-hydroxymethyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 2 H ₂ O, 40-methyl zymosterol = NAD(P)H = H ⁴ = 4-0-methyl zymosterol = NADP ⁴ , 40-hydroxymethyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 0, 40-carboxy-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P)H = H ⁴ = 4-0-methyl zymosterol = NADP ⁴ , 40-formyl-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = 0, 40-carboxy-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40-dien-38-ol = NADP ⁴ = H ⁴ = 4-0-methyl zymosterol = NADP ⁴ , 40-formyl-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40-carboxy-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40-carboxy-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40-carboxy-48-methyl-50-cholesta-8,24-dien-38-ol = NAD(P) ⁴ = H ₂ O, 40	1, 14, 13, 70 1, 14, 13, 70 1, 14, 13, 70

Comparing across species

Use general Comparative Analyses tools

Comparative Analysis and Statistics

This page allows you to compute statistics for a single Pathway/Genome Database, and to compute comparisons across multiple Pathway/Genome Databases for the set of organisms listed below. Note that computing summary statistics for the organism databases you have chosen may take several minutes. For faster operation, install Pathway Tools on your own computer! <u>Click here</u> for details.

Click here to go to the Webinar page, where you can watch the Comparative and Omics Tools in BioCyc instructional videos.

Note: In addition to reflecting differences in biology of different organisms, these statistics will reflect differences in the levels of curation, data availability, and completeness of the PGDBs for these organisms.

Select which set(s) of comparative-analysis tables you wish to generate:

- Reactions: breakdowns by type of substrate, by EC Number, by number of isozymes, etc.
- Pathways: breakdown by pathway class, information on pathway holes.
- Compounds: small molecules that act as substrates, enzyme activators/inhibitors/cofactors.
- Proteins: breakdown of protein complexes by type and number of subunits, number of enzymes, number of enzymes with
 activators/inhibitors/cofactors, multifunctional enzymes.
- Orthologs: proteins shared among organisms or unique to an organism.
- Transporters: proteins that facilitate the movement of compounds across cell membranes.
- Transcription Units: number of genes per transcription unit, number of operons per pathway.

Select one or more organism databases:

Arabidopsis thaliana col PlantCyc Populus trichocarpa

Visualizing OMICs data

- Overlay "pre-cleaned" data sets on a metabolic map
 - Gene transcription data
 - Proteomic data
 - Metabolomic data
- Only available for single-species databases, not PlantCyc
- Demonstrations available at conference . . . please see me!



Visualizing OMICs data





Data and software downloads

Install a local copy of the Pathway Tools software









- The PMN houses a large quantity of data, but . . .
 - Numerous previously identified enzymes and pathways are missing
 - Exciting new discoveries are continuously made
 - Plant genome sequence data is becoming available at a rapid rate
 - How do we bring this information in?

Database

- Manual curatio
 - Curators reaResearchers

Computational

- Automated p
 - Predicted
 - Predicted
 - Predict a

Manual valid

- Remove
- Create or
- Add refer

Animes and Polyamines Martius Proteowski - Ruhr University Bochum, Germany Amino Acids. Thomas Levelais - Autores University; USA **Carbohydrales** Mater Ban Polled - Complex Caribohushate Research Center, University of Georgia, USA Coluctors, Prostellac Groups, and Electron Carriers Taiting Begine - Texas ABAP University: USA service Roman State Liniversity, USA enderation LAK You! Institute of Science and Technology, Japan Jaworski - Donald Danityth Plant Science Center, USA Basil Mikolau - Jowa State University, USA Ruth Welt - Kansas State University, USA Horasones and Regulators. Sue Abranis - National Research Council-Plant Biotechnology Institute. Canada Jerry Cohen - University of Minnesota USA Mary Wildersuch - University of California, Beckeley, USA **Nucleosides and Nucleotides** Rita Zrenner - Max Planck Inablute of Molécular Plant Physiology, Germany Secondary Metabolism Alkalokdy Peter Facihini - University of Calgary: Canada Ravanoida Erich Gestewold - Onio State University; USA Lioyt Summer - The Samuel Roberts Noble Foundation, USA Dranda Winkel - Virginia Tech, USA Oliver Yu - Donald Danforth Plant Science Center, USA Phonylpt rig-apoids Clint Chapple - Purdue University, USA David Dang - University of Aribona, USA **PohAutides** Joe him Schroeder - University of Preiburg, Germany Plant Volatilies Evan Pichersky - University of Michigan, US4

al articles

natic functions tways in PlantCyc *d enzymes for species X*



- Improvide
 Jonishan Page National Research Council-Plant Biotechnology Institute, Canada
 - Mark Lange Washington State University, USA

Building better databases together

- □ To submit data, report an error, or volunteer to help validate . . .
 - Send an e-mail: <u>curator@plantcyc.org</u>
 - Use data submission "tools"

					Plan	iCyc 💌	search
About Page	Search	Tools	Downloads	Useful Sites	Subwell Data	Ne	Feedback
Contractile		Data Sul	bmission		Feedback Form		

Meet with me individually at this conference

- P2,4,12,14,18,20,22,24,26,30,34,36,38,40,44,48,50,54,56,62,74,84,92,9,21,25, 29,31,33,39,41,43,5157,59,63,65,75,81,87,89,91,93
- **CS30**,15,32,27,28,13,14,7,4,12, 31, 35, 18, 36, 21, 1, 16, 11, 6,
- PS IVb, Vb, Vla, Vlla, Vlla, Vllb

Building better databases together

Details are very, very welcome!!

- Reactions:
 - All co-factors, co-substrates, etc.
 - EC suggestions partial or full
- Compounds
 - Structure visual representation / compound file (e.g. mol file)
 - Synonyms
 - Unique IDs (e.g. ChEBI, CAS, KEGG)
- Enzymes
 - Unique IDs (e.g. At2g46480, UniProt, Genbank)
 - Specific reactions catalyzed

Community gratitude



Plant metabolic NETWORKING

- Please use our data
- Please use our tools
- Please help us to improve our databases!
- Please contact us if we can be of any help!





curator@plantcyc.org

www.plantcyc.org

special pre-release sneak preview: pmn.plantcyc.org



Collaborators:

- Peter Karp (SRI)
- Ron Caspi (SRI)
- Suzanne Paley (SRI)
- SRI Tech Team
- Lukas Mueller (SGN)
- Anuradha Pujar (SGN)
- Gramene and MedicCyc

gements

Current Tech Team Members:

- Bob Muller (Manager)
- Larry Ploetz (Sys. Administrator)
- Cynthia Lee
- Shanker Singh
- Chris Wilks

Tech Team: recent alumni

- Raymond Chetty
- Anjo Chi
- Vanessa Kirkup
- Tom Meyer





Data and software downloads



How does the PMN acquire new data?



Additional search of



DLADT** pro	THE BLASTP AA DATE AA D
Datasets:	
ineut	• all kingdoms
QUAX SHOLE	nce FloriCyc Enzymes
	• experimental or computational support • plants only
Upload a file	Browse.
and	The Plat assester Options
lanced BLAST and salts Retain 0	Parameter Options
Manced BLAST pand coatts Retain 0 Worl Title	places
kanced BLASS and salts Return O Joy! The mail address	places
Nanced BLAST pand solits Return O April The mail address sturn Results	ptices optional, will be added to output for your use To your web browser
Annord BLAST pand Hoatts Reten O Upot Title mail address Hum Results stum Results	Parameter Options places options options To your web browser HTM, hypertext
Name of BLASS pand sollts Return O Joyd Title mail address etum Results coult Format	ptices optional, will be added to output for your use To your web browser

How can you put the PMN to work for you?

- Learn background information about particular metabolic pathways
- Create customized metabolic data sets
- Compare metabolism across plant species
- Analyze experimental OMICs data in a metabolic context
- Manipulate and study data offline
- Create new metabolic pathway databases



Comparing across species

Overview of the AraCyc Metabolic Map

This diagram provides a schematic of all pathways of AraCyc metabolism in the metabolite (see key to right). Lines represent reactions. Move the mouse over a r page or a related pathway page.	vacyc database. Nodes represent metabolites, with shape indicating class of olite icon to navigate to the metabolite Select one or more organisms:
Instructions Pathway Tools guery page Dense Viewer, Paul once data only this degram Species Comparison: Highlight reactions shared with other organisms	□ PlantCyc I Populus trichocarpa

Comparing across species



Additional search options

THEN CLOSED VIEW CONTRACTORS AND A CONTRACT	e table of 64 rows.	
HAME A 🕈	Reaction-List A T	[(e ^? EC #UMDER : e ~- 12] A 7
abscisic acid glucose ester- bioconthecis	(+)-abscisate + UDP-D-glacose + abscisic acid glacose ester + UDP	2.4.1.1
ajugose biosynthesis i (galactinol-dependent)	verbascose + galactinol + ajugose + myo-inositol, stachyose + galactinol + verbascose + myo-inositol	2.4.1,+
ajugose biesynthesis II (galactinol-Independent)	2 raffinose = stachyose + sucrose, 2 verbascose = ajugase + stachyose, 2 stachyose = verbascose + raffinose	2.4.1 2.4.1 2.4.1
anthocyanih biosynthesis (pelargoridin 3-O-glucoside, cyanidin 3-O-glucoside)	$\label{eq:conversion} \begin{split} &= \alpha \cdot ketsglutarate + O_2 = a \ dihydroquercetin + succinate + CO_2 + H_2O_1 \ UDP+D-glucose + orahidin + UDP + orahidin -3-O-8-O-glucoside, \\ &= ketsglutarate + O_2 = orahidin + succinate + CO_2 + 2 \ H_2O + M^4, \ UDP-D-glucose + pelargonidin = UDP + pelargonidin -3-O-8-O-glucoside, \\ &= D-glucoside + H^4, \ leucopelargonidin + o-ketsglutarate + O_2 + H^4 + pelargonidin + succinate + CO_2 + 2 \ H_2O \end{split}$	2.4.1.115 2.4.1.115
arabidiol biorynthesis	(5)-2,3-Epoxysqualene + HyO + arabidist	4.2.1.
baruol biosynthesis benzoyi-CoA biosynthesis	 [5]-2,3-Eporysqualene + baruol. 3-keto-3-phenylproplonyl-CoA + coerupme A + berzoyl-CoA + acetyl-CoA, (6)-cinnamoyl-CoA + H₂O + 3-hydroxy-3-phenylproplonyl-CoA, 3-hydroxy- 3-phenylproplonyl-CoA + NADP* + 3-keto-3-phenylproplonyl-CoA + NADPH + H* 	5.4.99
cellulose biorynthesis	a long-linear glucan + HyD = a short glucan + a short glucan, UDP-D-glucese + (1.4-8-D-glucesyDp) = UDP + (1.4-8-D-glucesyDp)	2.4.1.12
cholecterol biosynthesis (lanosterol biorynthech, 4o-carboxy-48-methyl-5o-cholesta-8,24-dien-38-ol + NAD(P)* = 4o-methyl-5o-cholesta-8,24-dien-3-one + NAD(P)H + CO ₃ , lathosterol + NAD(P)H + O ₂ = 7-dehydro-cholesterol + NAD(P)* + 2 H ₂ O, 4,4-dimethyl-5-o-cholesta-8,14,24-trien-3-8-ol + NADPH + H* = 4,4-dimethyl-5o-cholesta-8,24-dien-3-8-ol + NADP*, epoxystqualene biosynthesis, cholesta-8,14,24-trien-3-8-ol + NADPH + H*, 5o-cholesta-7,24-dien-38-ol + NADPH + H* = lathosterol + NADP*, zenosterol + 5o-cholesta-7,24-dien-38-ol, 5o-cholesta-8,24-dien-3-one + NADPH + H* = zenosterol + NADP*, 4o-carboxy-5o-cholesta-8,24-dien-38-ol + NAD(P)* + 5o-cholesta-8,24-dien-38-ol, 5o-cholesta-8,24-dien-3-one + NADPH + H* = zenosterol + NADP*, 4o-carboxy-5o-cholesta-8,24-dien-38-ol + NAD(P)* + HaD(P)* + Ho, 4o-hydroxymethyl-5o-cholesta- 5o-cholesta-8,24-dien-38-ol + NAD(P)H + O ₂ + 4o-carboxy-5o-cholesta-8,24-dien-38-ol + NAD(P)* + Ho, 4o-methyl zynosterol + NAD(P)H + O ₂ = 4o-hydroxymethyl-5o-cholesta-8,24-dien-38-ol + NAD(P)* + HaD, 4o-methyl zynosterol + NAD(P)H + O ₂ =	1.14.13.70 1.14.13.70 1.14.13.70

Data and software downloads

Get pathway data sets from pathway pages

