

Curation, implementation, and visualization of biological information in metabolic networks of taxon- and species-specific databases at the Sol Genomics Network (SGN)

Hartmut Foerster, Lukas A. Mueller
Boyce Thompson Institute (BTI)

James N.D. Battey, Nicolas Sierro, Nikolai V. Ivanov
PMI R&D, Philip Morris Products S.A., Quai Jeanrenaud 5,
CH-2000 Neuchâtel, Switzerland
(Part of Philip Morris International group of companies)



sol genomics network

SRI International



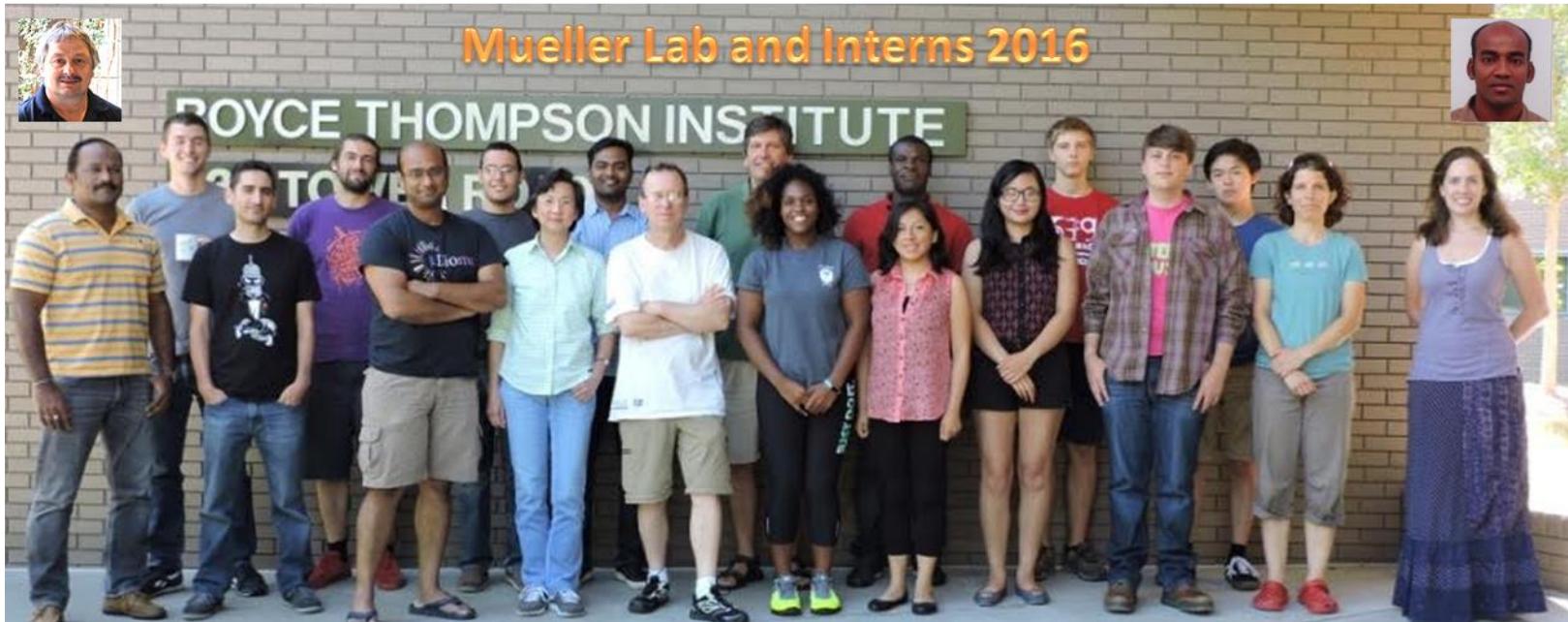
PHILIP MORRIS INTERNATIONAL



Boyce Thompson Institute (BTI)



The Bioinformatics Group at BTI (Lukas A. Mueller)



Databases for managing Big Data – access to transcriptomic, genotypic and phenotypic data

Databases supporting researcher and breeder:

Sol Genomics Network (Solanaceae)

Cassavabase, Yambase, Sweetpotatobase, Musabase, Rtbbase

Genomic and Open-source Breeding Informatics Initiative ([GOBII](#))

Genome Sequencing Projects (*Solanaceae* and related *Asterids*)

<https://btiscience.org/explore-bti/directory/lukas-mueller/>



sol genomics network



BTI Computational Biology Center (BCBC)



The screenshot shows the homepage of the BTI Computational Biology Center. At the top, the BTI logo is on the left, and a navigation bar with links to Explore, Our Research, Education & Outreach, Get Involved, and Login, along with a search icon is on the right. The main title "BTI Computational Biology Center" is displayed in large white text over a background image of green leaves. Below the title, a breadcrumb navigation shows Home, Our Research, and BTI Computational Biology Center. The page is titled "Members" and lists the Director and Consulting Team. Each team member is shown with a photo and name: Suzy Strickler, Yi Zheng, Adrian Powell, and David Lyon. Below this, the BTI Faculty Core Members are listed: Fay-Wei Li, Lukas Mueller, and Zhangjun Fei. The photos of Lukas Mueller and Suzy Strickler are highlighted with orange outlines.



The Sol Genomics Network

(<https://solgenomics.net/>)

Sol Genomics Network

Search Maps Genomes Projects Tools About

Tomato Genome

The new Tomato Genome assembly (SL3.0) and annotations (iTAG3.0) are now available on SGN

Browse the tomato genome
Find sequences by similarity
Download the annotations
Download the Genome Sequence

Tomato Potato Pepper *N. benthamiana* Petunia Eggplant SOL Meeting

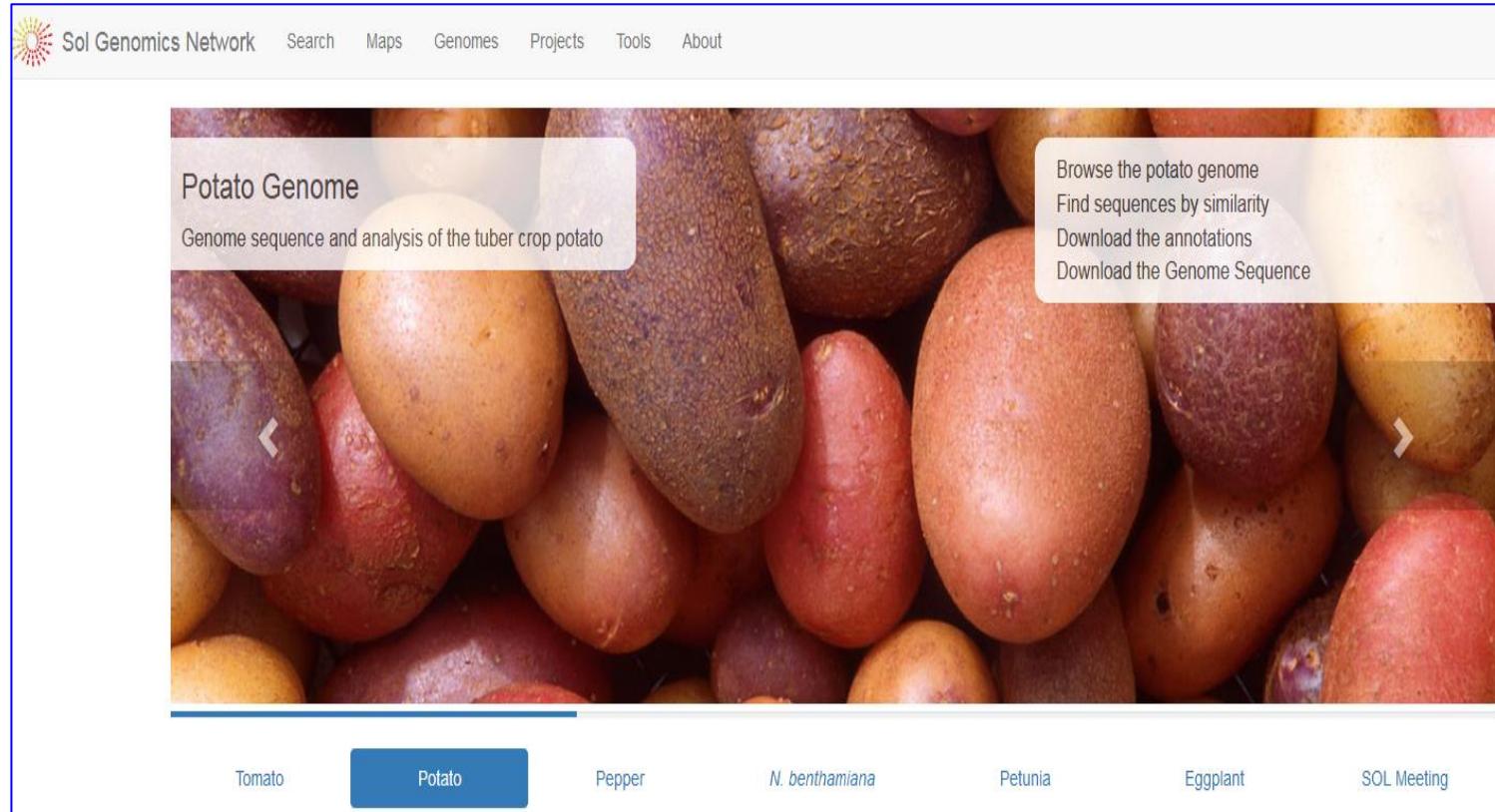


sol genomics network



The Sol Genomics Network

(<https://solgenomics.net/>)



Potato Genome
Genome sequence and analysis of the tuber crop potato

Browse the potato genome
Find sequences by similarity
Download the annotations
Download the Genome Sequence

Tomato Potato Pepper *N. benthamiana* Petunia Eggplant SOL Meeting

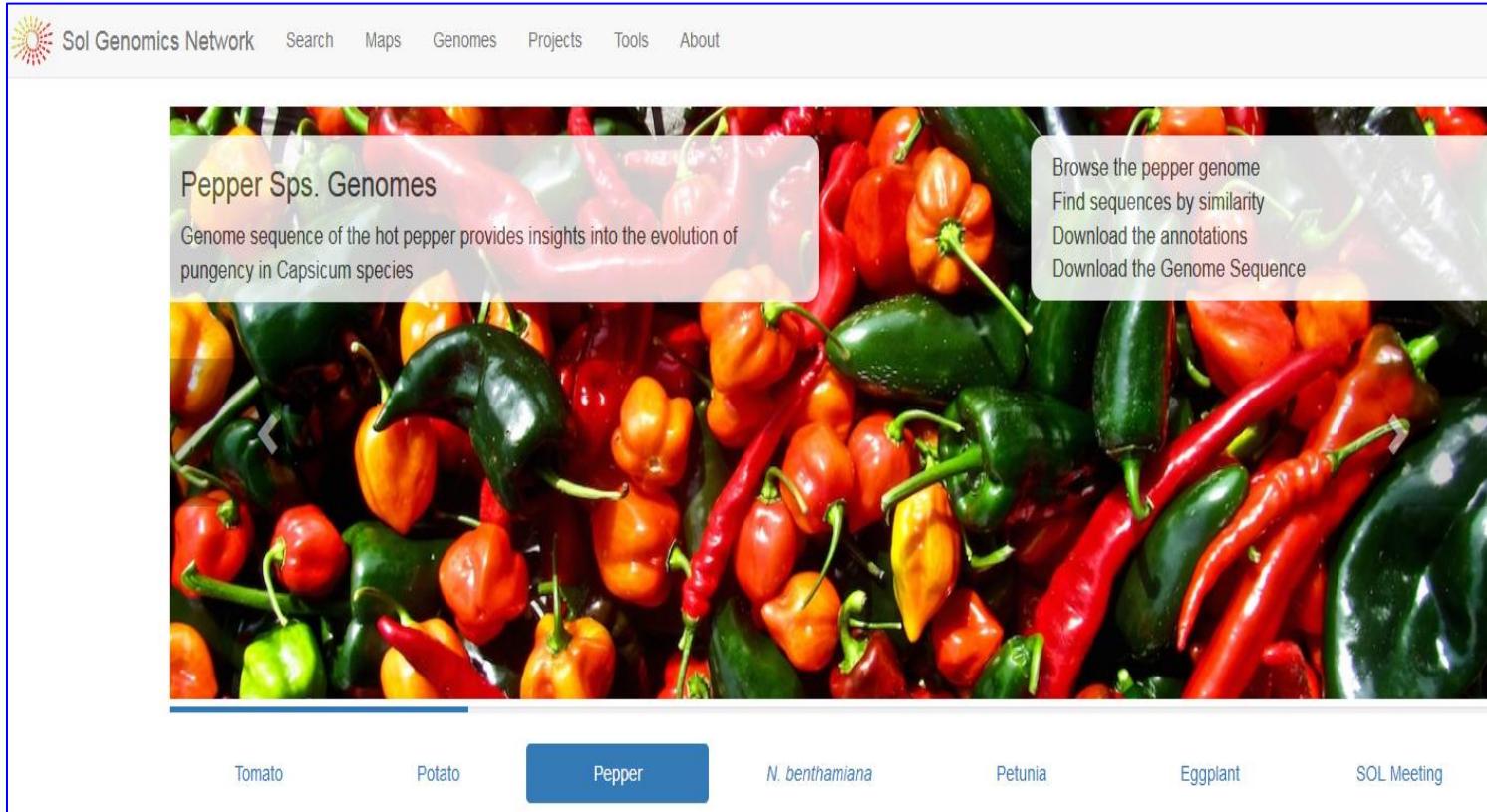


sol genomics network



The Sol Genomics Network

(<https://solgenomics.net/>)



The screenshot shows the Sol Genomics Network website for the Pepper species page. The header includes the logo and navigation links: Sol Genomics Network, Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of various chili peppers (red, green, orange) and a callout box for "Pepper Sps. Genomes". The callout box contains the text: "Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species". To the right of the image is another callout box with links: "Browse the pepper genome", "Find sequences by similarity", "Download the annotations", and "Download the Genome Sequence". Below the main image is a navigation bar with links: Tomato, Potato, Pepper (highlighted in blue), *N. benthamiana*, Petunia, Eggplant, and SOL Meeting.

The Sol Genomics Network

(<https://solgenomics.net/>)

The screenshot shows the Sol Genomics Network homepage. At the top, there is a navigation bar with links: Sol Genomics Network, Search, Maps, Genomes, Projects, Tools, and About. Below the navigation bar, there is a main content area featuring a central image of a *N. benthamiana* plant and several leaves. A callout box on the left side of the image contains the text: "N. benthamiana Genome" and "A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research". Another callout box on the right side contains the text: "The *N. benthamiana* sequencing project", "Browse the genome", "Find sequences by similarity", "Design VIGS constructs", and "Downloads". At the bottom of the main content area, there are links for other genome projects: Tomato, Potato, Pepper, *N. benthamiana* (which is highlighted in a blue box), Petunia, Eggplant, and SOL Meeting.

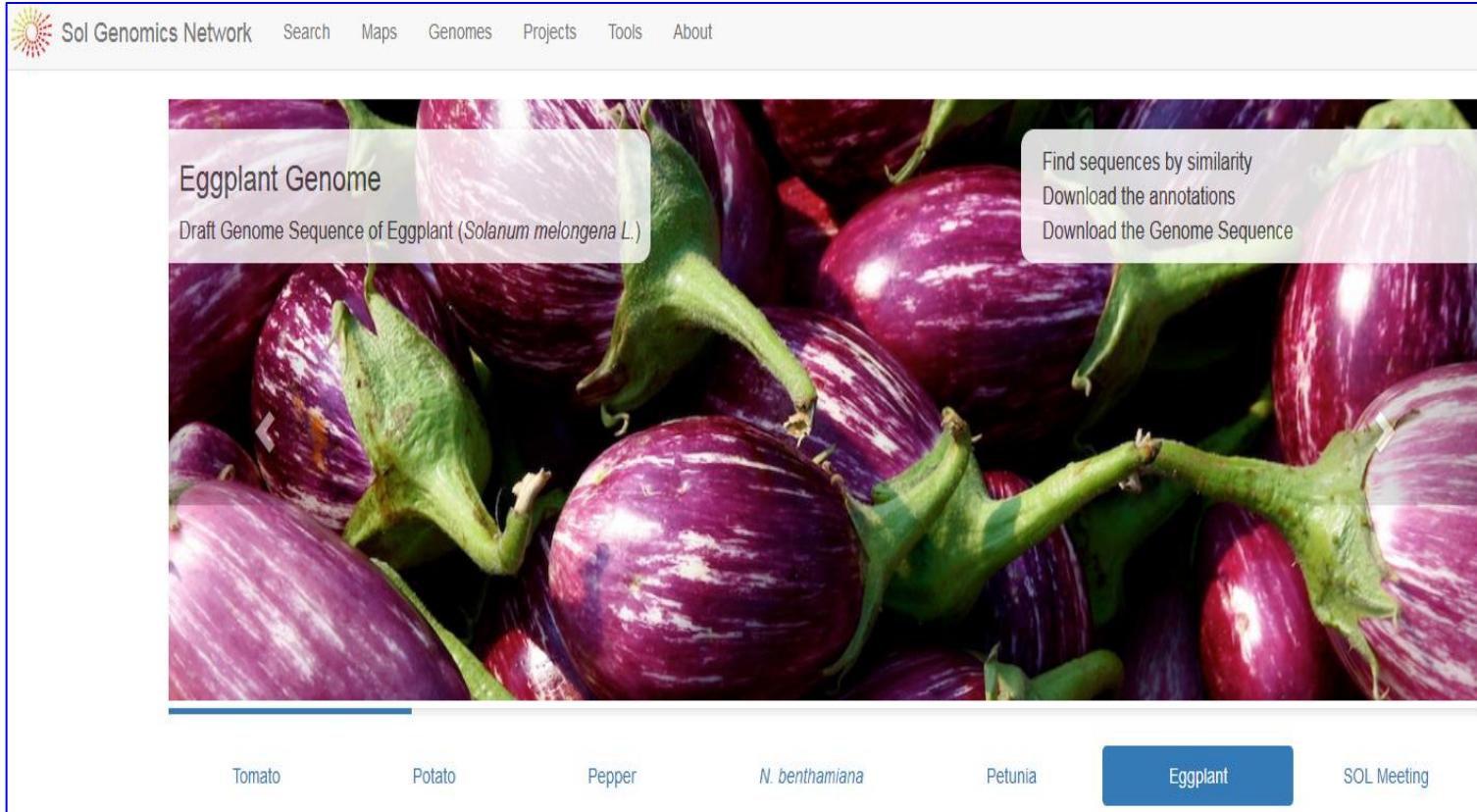
The Sol Genomics Network

(<https://solgenomics.net/>)

The screenshot shows the Sol Genomics Network website. At the top, there is a navigation bar with links: Sol Genomics Network, Search, Maps, Genomes, Projects, Tools, and About. Below the navigation bar is a large image of various Petunia flowers in different colors (pink, red, purple, yellow). Overlaid on this image is a white callout box containing the text: "Petunia Sps. Genomes" and "Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida". To the right of the image is a sidebar with links: "Browse the *P. axillaris* genome", "Browse the *P. inflata* genome", "Find sequences by similarity", "Design VIGS constructs", "*P. axillaris* Downloads", and "*P. inflata* Downloads". At the bottom of the page, there are links for other Sol species: Tomato, Potato, Pepper, *N. benthamiana*, Petunia (which is highlighted in a blue box), Eggplant, and SOL Meeting.

The Sol Genomics Network

(<https://solgenomics.net/>)



The screenshot shows the Sol Genomics Network homepage for the Eggplant Genome. The main image is a close-up of several purple eggplants with green stems. A white callout box on the left contains the text "Eggplant Genome" and "Draft Genome Sequence of Eggplant (*Solanum melongena L.*)". Another callout box on the right contains links: "Find sequences by similarity", "Download the annotations", and "Download the Genome Sequence". Below the image is a horizontal navigation bar with buttons for Tomato, Potato, Pepper, *N. benthamiana*, Petunia, Eggplant (which is highlighted in blue), and SOL Meeting.

Sol Genomics Network

Search Maps Genomes Projects Tools About

Eggplant Genome

Draft Genome Sequence of Eggplant (*Solanum melongena L.*)

Find sequences by similarity
Download the annotations
Download the Genome Sequence

Tomato Potato Pepper *N. benthamiana* Petunia Eggplant SOL Meeting

The Sol Genomics Network

(<https://solgenomics.net/>)

Sol Genomics Network Search Maps Genomes Projects Tools About

The 15th Solanaceae Conference
September 30th - October 4th 2018, Chiang Mai, Thailand

Registration
Preliminary Program
Venue

The 15th Solanaceae Conference

Applied Genomics, Accelerated Breeding, Gene Targeting

EAST-WEST SEED

September 30th - October 4th 2018 - Chiang Mai, Thailand



Tomato Potato Pepper *N. benthamiana* Petunia Eggplant [SOL Meeting](#)

Accessing SolCyc

(<https://solgenomics.net/tools/solcyc/index.pl>)

The screenshot shows the Sol Genomics Network website. The main navigation bar includes links for Sol Genomics Network, Search, Maps, Genomes, Projects, Tools (which is currently selected and highlighted in blue), and About. The main content area features a section for the *N. benthamiana* Genome, which is described as a draft genome sequence of *Nicotiana benthamiana* to enable microbe biology research. Below this text are images of *N. benthamiana* leaves. A dropdown menu is open over this section, listing various tools and databases. The menu categories include Sequence Analysis, Mapping, Molecular Biology, Systems Biology, Breeder Tools, and Bulk Query. The Systems Biology category is currently active, as indicated by a red box around the "SolCyc Biochemical Pathways" link. Other visible links in this category include Coffee Interactomic Data, SGN Ontology Browser, Breeders Toolbox, and Unigene Converter. The main content area also features images of leaves from other species: Tomato, Potato, Petunia, Eggplant, and SOL Meeting. A news section at the bottom left discusses SGN GDPR compliance update, and a SOL Meeting 2018 announcement is at the bottom right.

N. benthamiana Genome
A draft genome sequence of *Nicotiana benthamiana* to enable microbe biology research

Tools

- Sequence Analysis
 - BLAST
 - VIGS Tool
 - Alignment Analyzer
 - Tree Browser
- Mapping
 - Genome Browser (Jbrowse)
 - Comparative Map Viewer
 - CAPS Designer
 - solQTL: QTL Mapping
- Molecular Biology
 - In Silico PCR
 - Tomato Expression Atlas (TEA)
 - Tomato Expression Database (TED)
- Systems Biology
 - SolCyc Biochemical Pathways
 - Coffee Interactomic Data
 - SGN Ontology Browser
- Breeder Tools
 - Breeders Toolbox
- Bulk Query
 - FTP Site
 - Download Gene Sequences
 - Clones, Arrays, Unigenes and BACs
 - Unigene Converter

News

SGN GDPR compliance update
SGN was updated to comply with the new European General Data Protection Regulation (GDPR). Please let us know about any issues or concerns. [May 24, 2018]

SOL Meeting 2018 in Chiang Mai, Thailand
The SOL Meeting 2018 will be held in Chiang Mai, Thailand, from September 30 - Oct 4. [Registration](#) is now open!

Actively curated *Solanaceae* Cyc's at SGN

Pathway Tools version 22.0

File Overviews Pathway Reaction Protein RNA Gene Compound Chromosome SmartTables Tools Help

Select Organism Home Back Forward History Next Answer Clone

Pathway Tools -- Available Databases

Organisms	Pathways	Genes (ORF %)	Genome Size (bp)
<i>Nicotiana attenuata</i> <i>Nicotiana benthamiana</i> <i>Nicotiana sylvestris</i> <i>Nicotiana tabacum</i> <i>Nicotiana tabacum K327</i> <i>Nicotiana tomentosiformis</i> NicotianaCyc SolanaCyc <i>Solanum lycopersicum</i> <i>Solanum tuberosum</i>			

Copyright Notice

Curation progress in SGN's manually curated taxon-specific databases

<i>SolanaCyc</i>	Pathways	Reactions	Enzymes	Transporter	Compounds
version					
1.0	169	835	257	1	1441
2.0	246	1093	491	72	1769

5168 citations

<i>NicotianaCyc</i>	Pathways	Reactions	Enzymes	Transporter	Compounds
version					
1.0	18	75	32	0	260
2.0	100	407	139	29	646

3054 citations

Curation progress in SGN's manually curated taxon-specific databases



Database, 2018, 1–13
doi: 10.1093/database/bay035
Original article



Original article

SolCyc: a database hub at the Sol Genomics Network (SGN) for the manual curation of metabolic networks in *Solanum* and *Nicotiana* specific databases

**Hartmut Foerster¹, Aureliano Bombarely², James N.D. Battey³,
Nicolas Sierro³, Nikolai V. Ivanov³ and Lukas A. Mueller^{1,*}**

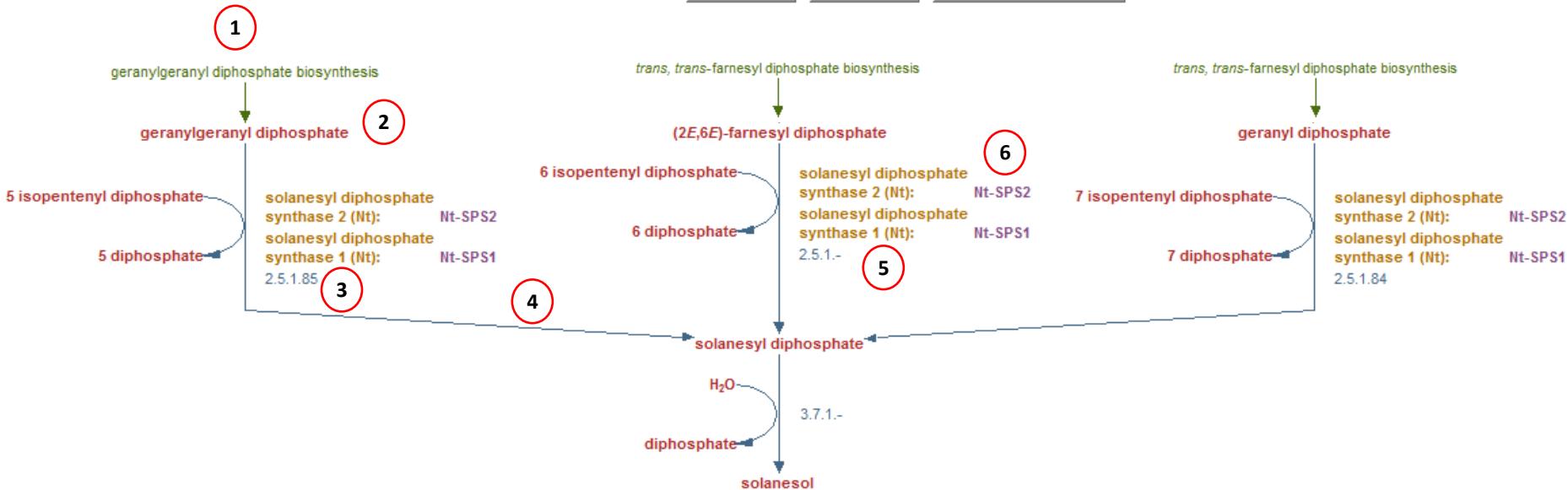
¹Boyce Thompson Institute, 533 Tower Road, Ithaca, New York, 14853-1801, USA, ²Department of Horticulture, Virginia Polytechnic Institute and State University, 220 Ag Quad Lane, Blacksburg, VA 24061, USA and ³PMI R&D, Philip Morris Products S.A (Part of Philip Morris International group of companies), Quai Jeanrenaud 6, Neuchâtel CH-2000, Switzerland

Information ‘at a glance’ of curated pathways



NicotianaCyc Pathway: solanesol biosynthesis

All Organisms



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

Information ‘at a glance’ of curated pathways cont’d

Some taxa known to possess this pathway include  **Nicotiana tabacum**  [Yan16], **Solanum lycopersicum**  [Jones13]

8

Expected Taxonomic Range: **Magnoliophyta**

9

Superclasses: **Biosynthesis** -> **Secondary Metabolites Biosynthesis** -> **Terpenoids Biosynthesis**

10

Summary:

11

Solanesol is a linear terpene alcohol, which belongs to the wide spread compound class of isoprenoids with over 40.000 members found in nature [Adam98, Taylor11]. This compound has been first identified in tobacco [Rowland56] and shown to be present in other members of the night shade family as well, where it occurs in both free and esterified form. Among the *Solanaceae*, *Nicotiana tabacum* has been reported to be the species in that plant family that accumulates **solanesol** at the highest concentration [Yan15, Taylor11].

The biosynthetic route of **solanesol** has been shown to originate from the plastidial **methyerythritol phosphate pathway I** [Fukusaki04] and extends further towards plastoquinones. The **solanesol** molecule is a sought-after precursor used in the pharmaceutical industry for the production of isoprenylated benzoquinones such as coenzyme Q10 (or ubiquinone 10) involved in the electron carrier chain of aerobic respiration [Parmar15, Yan15], vitamin K2 used in human diet to prevent several chronic diseases [Hey15] and anticancer drugs such as N-solanesyl-N,N'-bis(3,4-dimethoxybenzyl)ethylenediamine (SDB) [Suzuki90, Tomida90, Sidorova02].

Although **solanesol** has also been shown to be a precursor of carcinogenic compounds (PAHs - polynuclear aromatic hydrocarbons) [Scholtzhauer76, Ishikawa16], this compound has also demonstrated beneficiary effects such as protecting human hepatic L02 cells from oxidative injury [Yao15] and increasing the resistance against tobacco mosaic virus (TMV) [Bajda09] and other pathogens [Maciejewska02]. Moreover, paralogous solanesyl diphosphate synthases, i.e. catalyzing enzymes in **solanesol** biosynthesis, have been shown by functional modeling to be involved in modifying the side chain of plastoquinone-9, representing an essential redox cofactor of photosynthesis [Block13].

Information ‘at a glance’ of curated pathways cont’d

Credits: Created 14-May-2018 by Foerster H, Boyce Thompson Institute

References

12

Adam98: Adam K-P, Zapp J (1998). "Biosynthesis of the isoprene units of chamomile sesquiterpenes." *Phytochemistry* 48(6):953-959.

Bajda09: Bajda A, Konopka-Postupolska D, Krzymowska M, Hennig J, Skorupinska-Tudek K, Surmacz L, Wojcik J, Matysiak Z, Chojnicki T, Skorzynska-Polit E, D M, Danikiewicz W, Piotrowska W, Swiezewska E (2009). "Role of polyisoprenoids in tobacco resistance against biotic stresses." *Physiol Plant* 135(1):1-10.

Block13: Block A, Fristedt R, Rogers S, Kumar J, Barnes B, Barnes J, Elowsky CG, Wamboldt Y, Mackenzie SA, Redding K, Merchant SS, Basset GJ (2013). "Functional solanesyl-diphosphate synthases that assemble the side chain of plastoquinone-9 in plastids." *J Biol Chem* 288(38):27594-606. PMID: 23913686

Campbell16: Campbell R, Freitag S, Bryan GJ, Stewart D, Taylor MA (2016). "Environmental and Genetic Factors Associated with Solanesol Accumulation in *Arabidopsis thaliana*." *Plant Physiol* 171(3):1320-1333. PMID: 27040000

Fukusaki04: Fukusaki E, Takeno S, Bamba T, Okumoto H, Katto H, Kajiyama S, Kobayashi A (2004). "Biosynthetic pathway for the C45 polypropenol, solanesol." *Plant Physiol* 134(3):1320-1333. PMID: 15388978

Hey15: Hey H, Brasen CL (2015). "[Vitamin K2 influences several diseases]." *Ugeskr Laeger* 177(32):V12140700. PMID: 26321585

Hirooka03: Hirooka K, Bamba T, Fukusaki E, Kobayashi A (2003). "Cloning and kinetic characterization of *Arabidopsis thaliana* solanesyl diphosphate synthase." *Plant Physiol* 132(3):1320-1333. PMID: 14602000

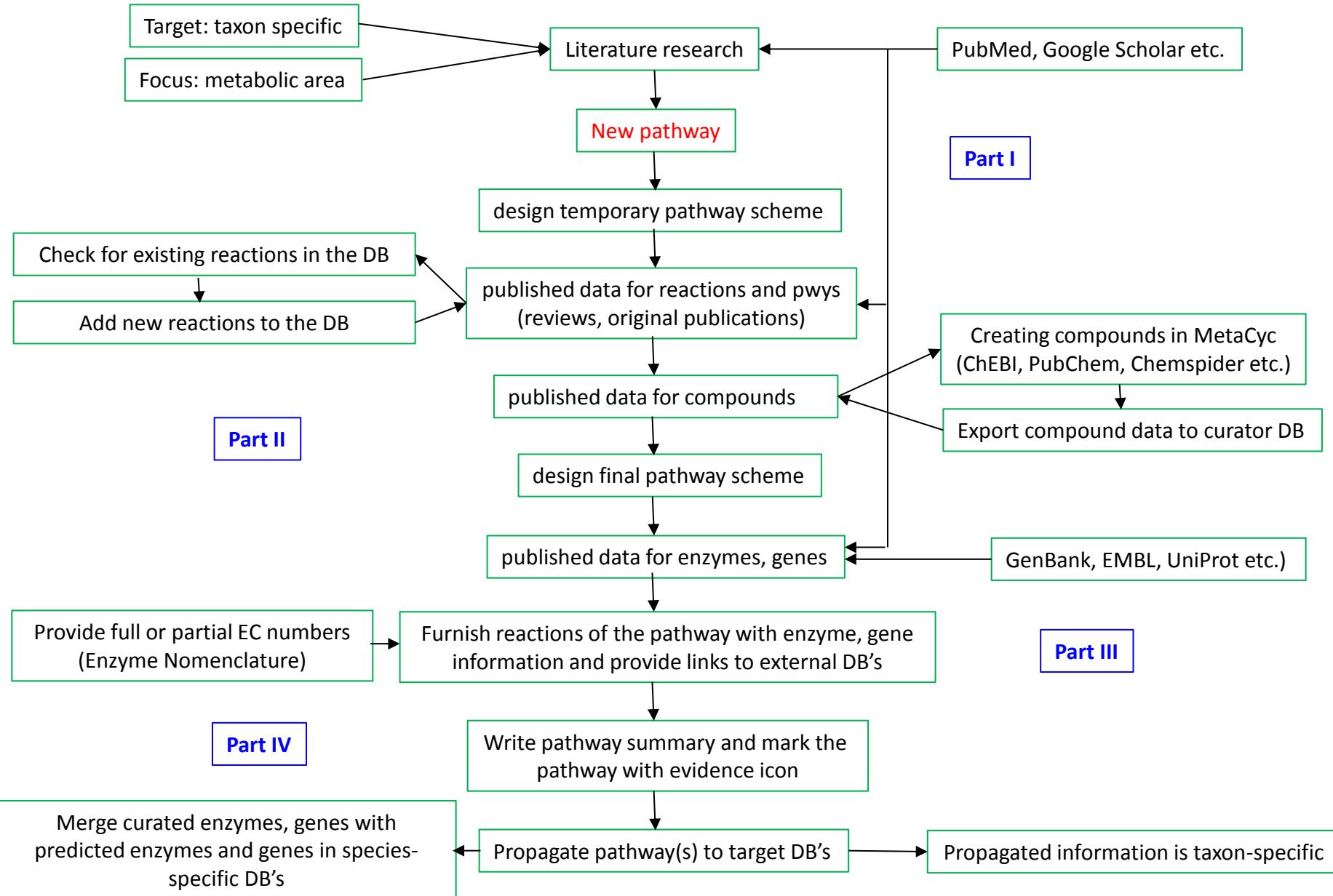
Hirooka05: Hirooka K, Izumi Y, An CI, Nakazawa Y, Fukusaki E, Kobayashi A (2005). "Functional analysis of two solanesyl diphosphate synthases from *Arabidopsis thaliana*." *Plant Physiol* 138(3):1578-1589. PMID: 15784989

Ishikawa16: Ishikawa S, Kanemaru Y, Nara H, Erami K, Nagata Y (2016). "Assessing the mutagenic activities of smoke from different cigarettes in direct exposure assay." *Mutat Res Genet Toxicol Environ Mutagen* 803-804:13-21. PMID: 27265375

Jones13: Jones MO, Perez-Fons L, Robertson FP, Bramley PM, Fraser PD (2013). "Functional characterization of long-chain prenyl diphosphate synthases from *Arabidopsis thaliana*." *Plant Physiol* 162(3):1320-1333. PMID: 23545500

Jun04: Jun L, Saiki R, Tatsumi K, Nakagawa T, Kawamukai M (2004). "Identification and subcellular localization of two solanesyl diphosphate synthases from *Arabidopsis thaliana*." *Plant Physiol* 134(3):1320-1333. PMID: 15388978

Pathway curation flow



Conclusions and Outlook

- Metabolic databases are valuable tools for evaluating, transforming and visualizing the ever-increasing amount of data into the biological context of species
- Manual curated databases provide a high level of accuracy by incorporating verifiable data from the published literature
- Datasets from highly curated reference databases significantly increase the probability for predicting the correct network of pathways, enzymes, genes and compounds of a species
- Manually curated taxon-specific reference databases will improve metabolic network predictions even more so by significantly reducing the rate of false positives
- Taxon-specific database containing only manually validated pathways have the potential to develop into knowledge pools for the biochemistry and molecular biology of taxonomic ranks

Acknowledgements

□ **BTI**

Lukas A Mueller (*PI*)

Hartmut Foerster (*Research Associate*)

□ **Virginia Tech**

Aureliano Bombarely (Assistant Professor)

□ **PMN**

Seung Rhee (*PI*)

Peifen Zhang (*PMN Director*)

□ **PMI (R&D)**

James ND Battey

Nicolas Sierro

Nikolai V Ivanov

□ **SRI**

Peter Karp (*PI*)

Ron Caspi (*Head Curator*)

Carol Fulcher (*Curator*)

Anamika Kothari (*Curator*)

Suzanne Paley (*Software Developer*)

Quang Ong (*Scientific Programmer*)

Pallavi Subhraveti (*Programmer*)

Markus Krummenacker (*Bioinformatician*)

Mario Latendresse (*Computer Scientist*)

Sources of funding and sponsors

