

# 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 (*Solanacea* and related *Asterids*)

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



sol genomics network




# BTI Computational Biology Center (BCBC)




Home Our Research BTI Computational Biology Center

## Members


Director Consulting Team




Suzy Strickler



Yi Zheng




Adrian Powell




David Lyon

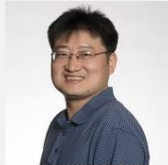
BTI Faculty Core Members



Fay-Wei Li



Lukas Mueller

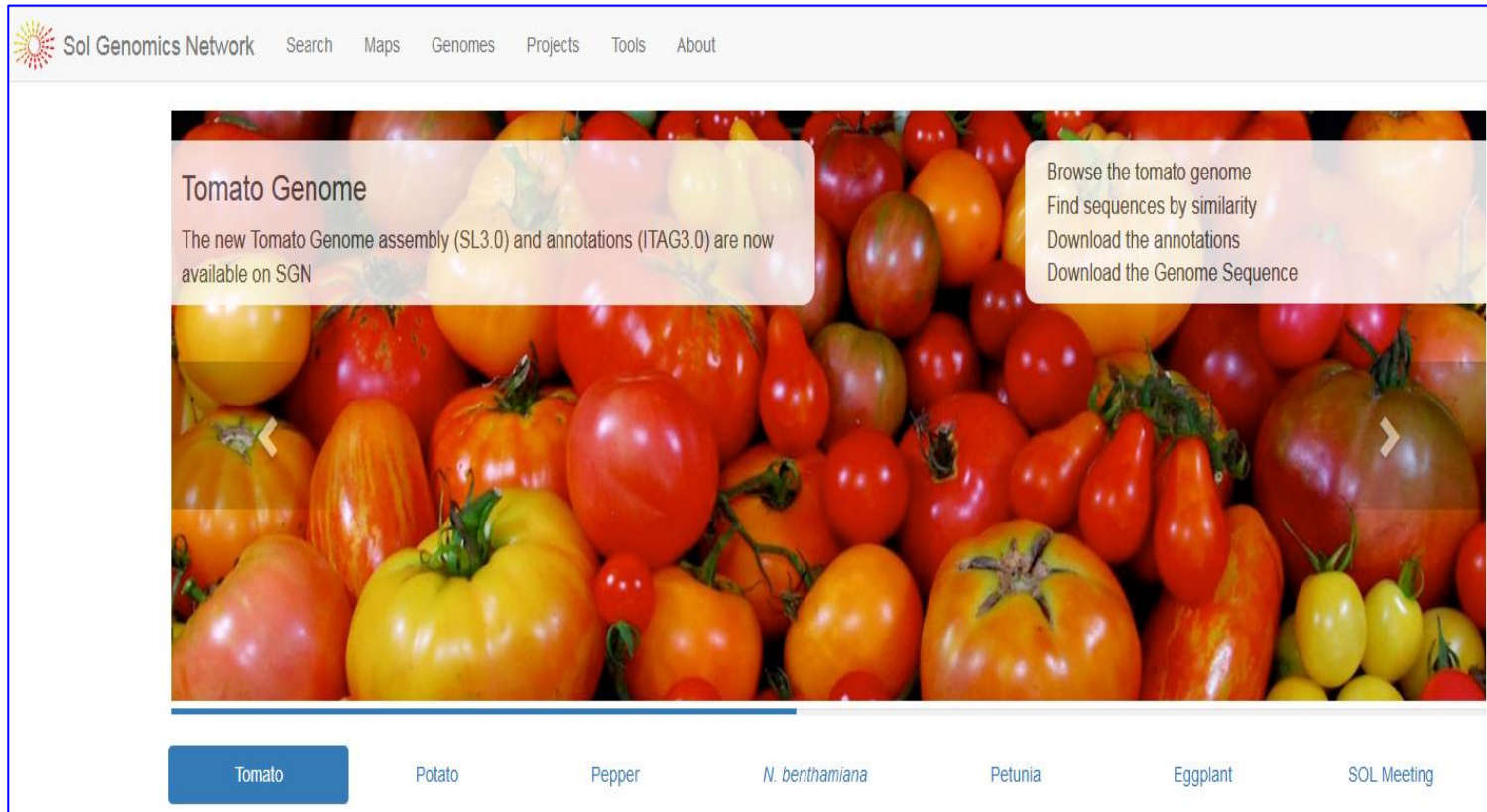


Zhangjun Fei



# The Sol Genomics Network

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



The screenshot displays the Sol Genomics Network website. The header includes the logo and navigation links: Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of various tomatoes. Overlaid on this image is a 'Tomato Genome' section with the text: 'The new Tomato Genome assembly (SL3.0) and annotations (ITAG3.0) are now available on SGN'. To the right of this text is a list of links: 'Browse the tomato genome', 'Find sequences by similarity', 'Download the annotations', and 'Download the Genome Sequence'. At the bottom of the page, there is a horizontal menu with buttons for 'Tomato', 'Potato', 'Pepper', '*N. benthamiana*', 'Petunia', 'Eggplant', and 'SOL Meeting'. The 'Tomato' button is highlighted with a blue background.



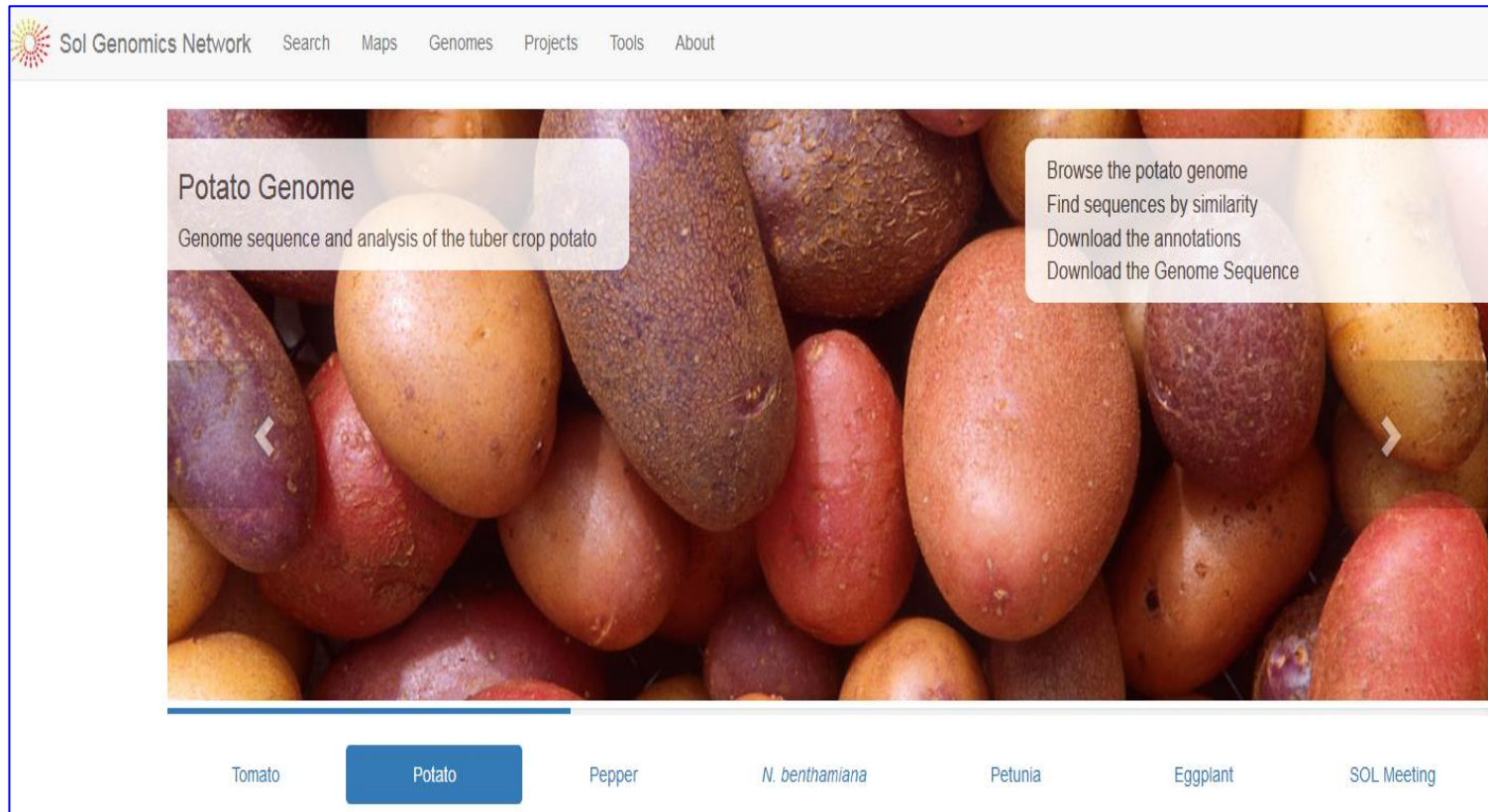
sol genomics network





# The Sol Genomics Network

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



The screenshot displays the Sol Genomics Network website. At the top, a navigation bar includes the logo and links for Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of various potato tubers. Overlaid on this image is a 'Potato Genome' section with the subtitle 'Genome sequence and analysis of the tuber crop potato'. To the right of this section is a list of actions: 'Browse the potato genome', 'Find sequences by similarity', 'Download the annotations', and 'Download the Genome Sequence'. Below the main image, a horizontal menu contains buttons for Tomato, Potato (which is highlighted), Pepper, *N. benthamiana*, Petunia, Eggplant, and SOL Meeting.

Sol Genomics Network Search Maps Genomes Projects Tools About

## 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

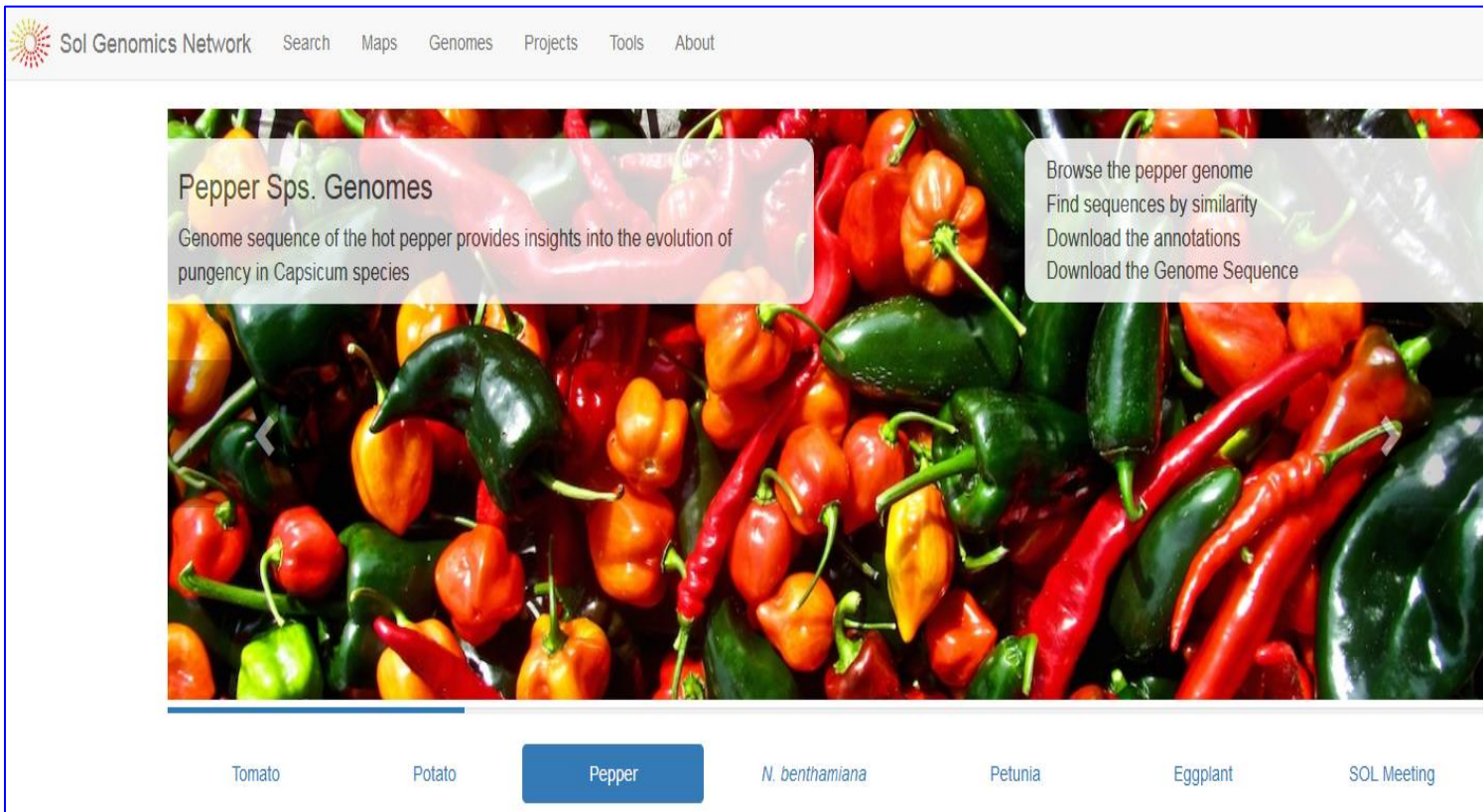


sol genomics network



# The Sol Genomics Network

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



The screenshot shows the Sol Genomics Network website. The header includes the logo and navigation links: Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of various peppers. Overlaid on this image is a semi-transparent box with the text "Pepper Sps. Genomes" and a description: "Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species". To the right of this box is another semi-transparent box with a list of links: "Browse the pepper genome", "Find sequences by similarity", "Download the annotations", and "Download the Genome Sequence". At the bottom of the page, there is a horizontal navigation bar with buttons for "Tomato", "Potato", "Pepper" (which is highlighted), "*N. benthamiana*", "Petunia", "Eggplant", and "SOL Meeting".

Sol Genomics Network Search Maps Genomes Projects Tools About

**Pepper Sps. Genomes**  
Genome sequence of the hot pepper provides insights into the evolution of pungency in Capsicum species

- Browse the pepper genome
- Find sequences by similarity
- Download the annotations
- Download the Genome Sequence

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

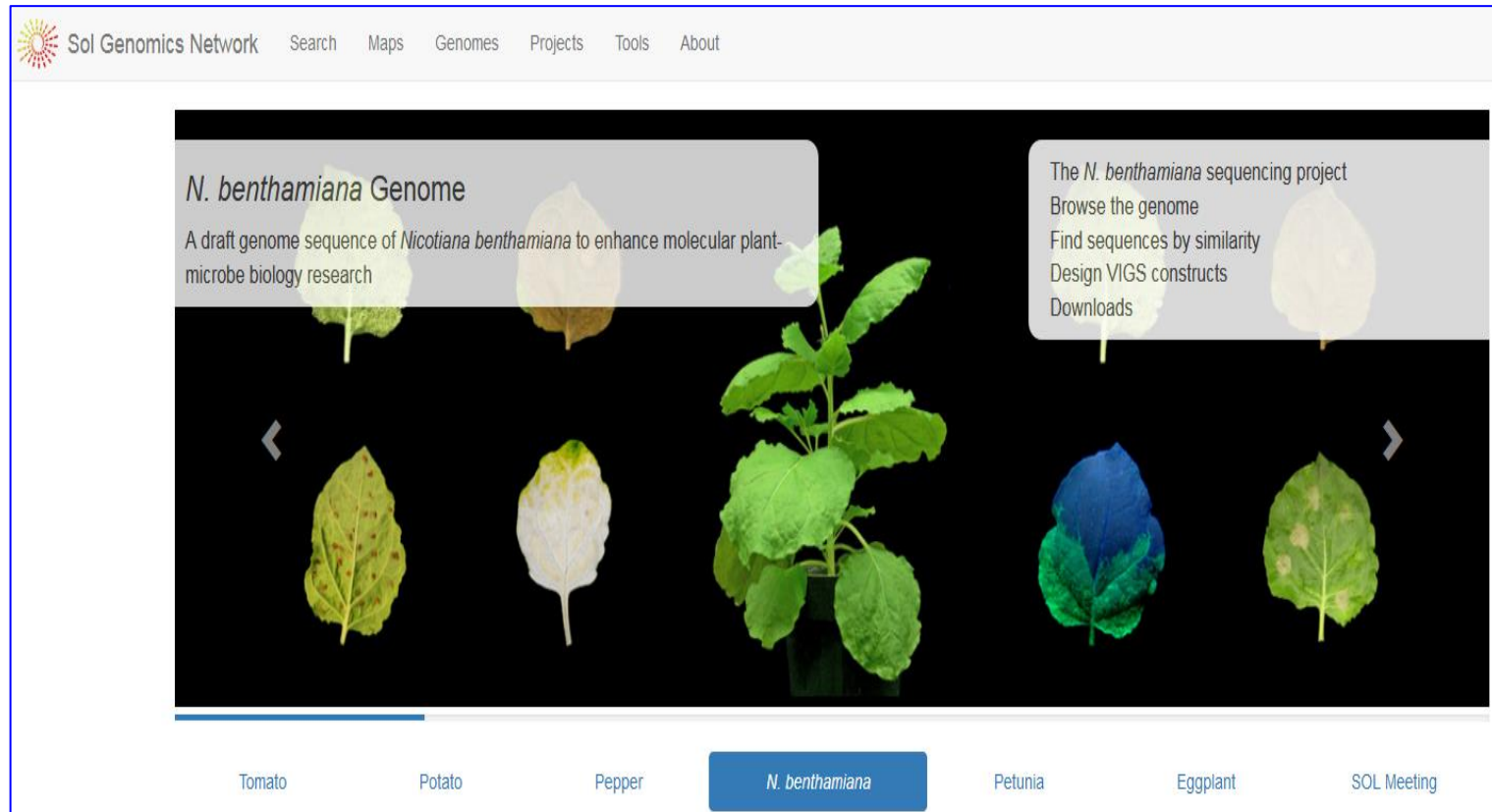


sol genomics network



# The Sol Genomics Network

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



The screenshot displays the Sol Genomics Network website. At the top is a navigation bar with the logo and links for Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of a tobacco plant with several callout boxes. On the left, a box titled "*N. benthamiana* Genome" describes it as a draft genome sequence for plant-microbe biology research. On the right, a box titled "The *N. benthamiana* sequencing project" lists options to browse the genome, find sequences by similarity, design VIGS constructs, and download. Below the main image is a horizontal menu with tabs for Tomato, Potato, Pepper, *N. benthamiana* (which is highlighted in blue), Petunia, Eggplant, and SOL Meeting. Navigation arrows are visible on the left and right sides of the main image area.



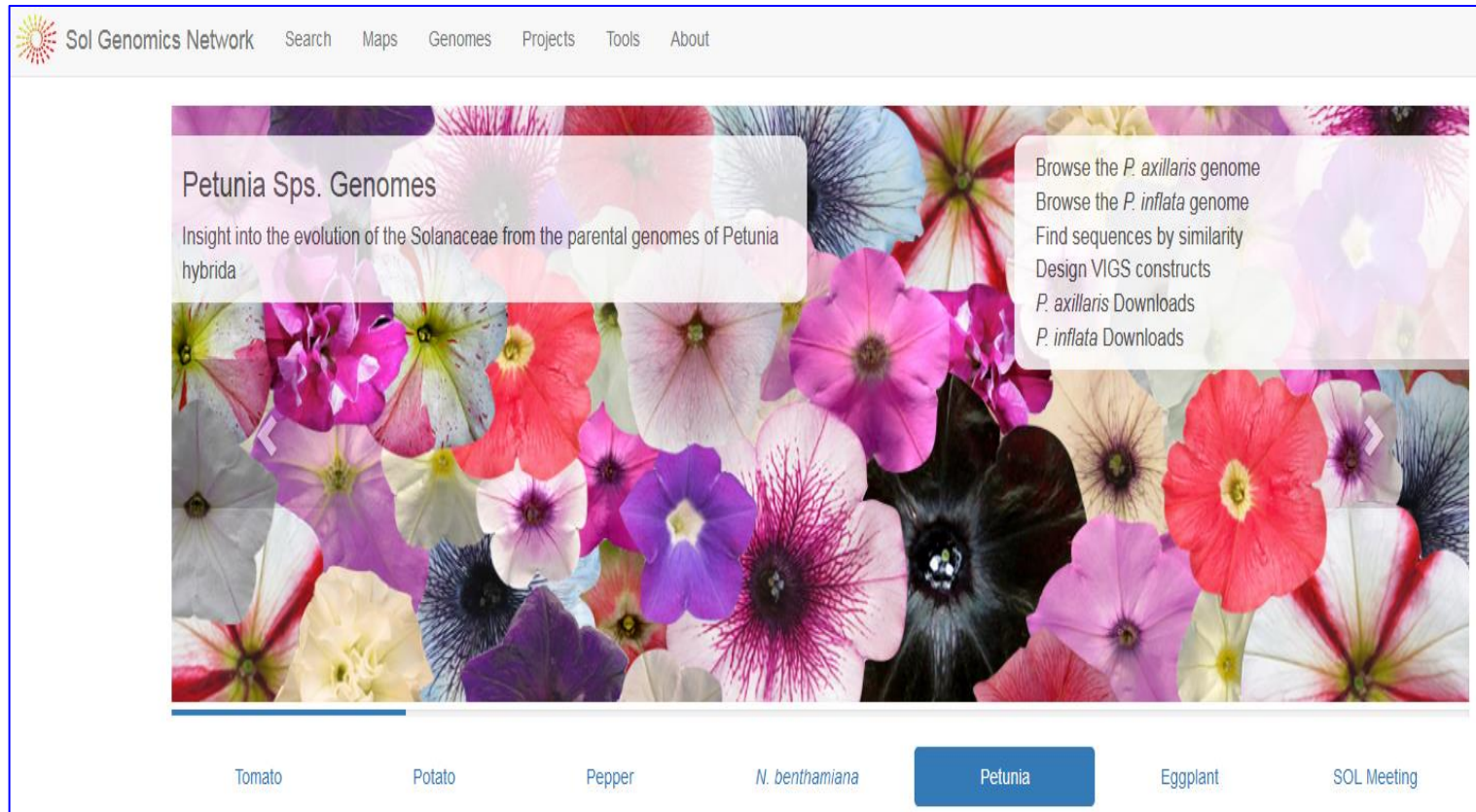
sol genomics network





# The Sol Genomics Network

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



The screenshot shows the Sol Genomics Network website. The header includes the logo and navigation links: Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of various Petunia flowers. Overlaid on this image is a white box with the text "Petunia Sps. Genomes" and a description: "Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida". To the right of this box is a list of 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 is a navigation bar with buttons for Tomato, Potato, Pepper, *N. benthamiana*, Petunia (which is highlighted in blue), Eggplant, and SOL Meeting.

Sol Genomics Network Search Maps Genomes Projects Tools About

**Petunia Sps. Genomes**  
Insight into the evolution of the Solanaceae from the parental genomes of Petunia hybrida

- Browse the *P. axillaris* genome
- Browse the *P. inflata* genome
- Find sequences by similarity
- Design VIGS constructs
- P. axillaris* Downloads
- P. inflata* Downloads

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

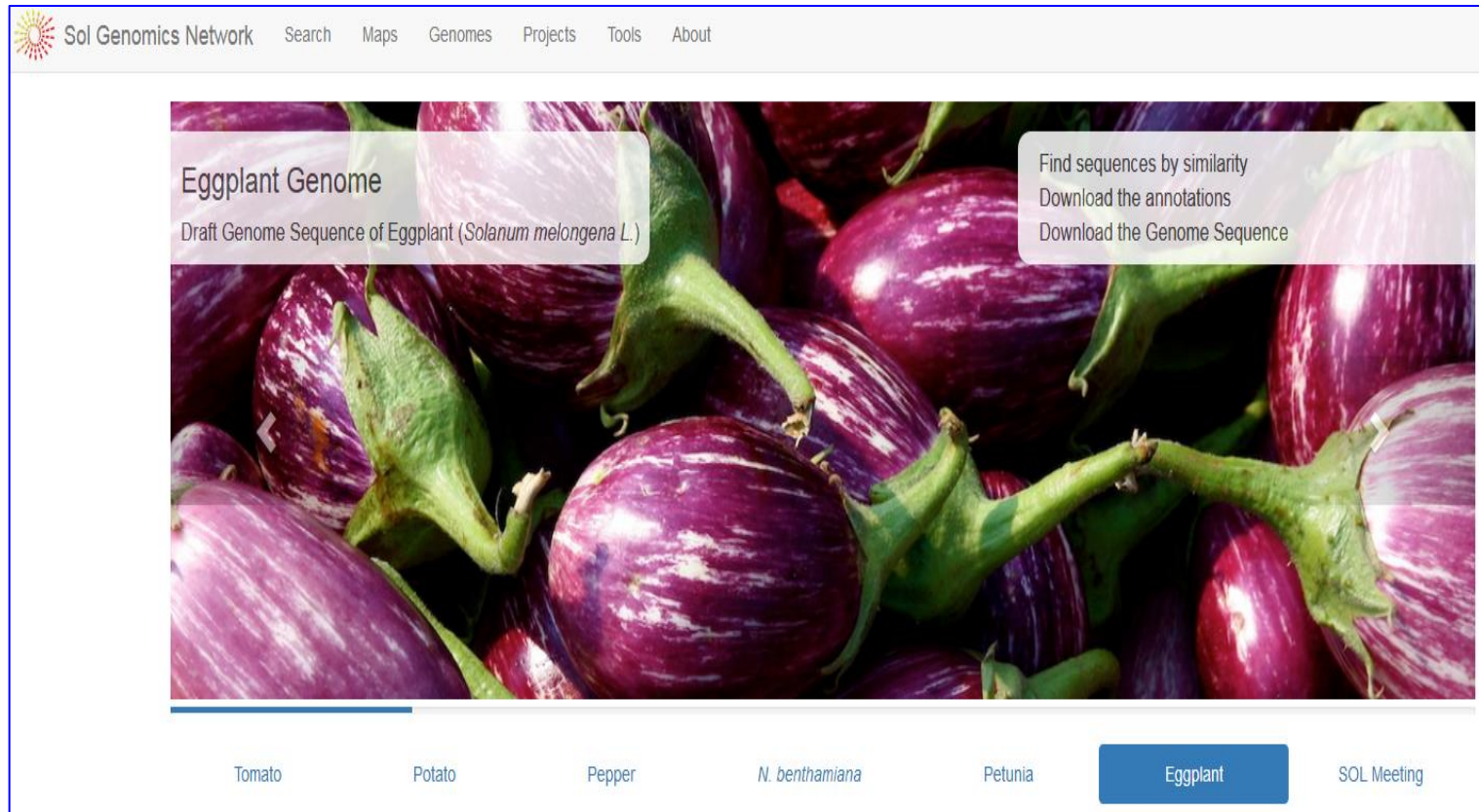


sol genomics network



# The Sol Genomics Network

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



The screenshot shows the Sol Genomics Network website. The header includes the logo and navigation links: Search, Maps, Genomes, Projects, Tools, and About. The main content area features a large image of eggplants. Overlaid on the image is a box titled "Eggplant Genome" with the subtitle "Draft Genome Sequence of Eggplant (*Solanum melongena* L.)". To the right of the image, there are three links: "Find sequences by similarity", "Download the annotations", and "Download the Genome Sequence". At the bottom, there is a navigation bar with links 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



sol genomics network



# 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](#)





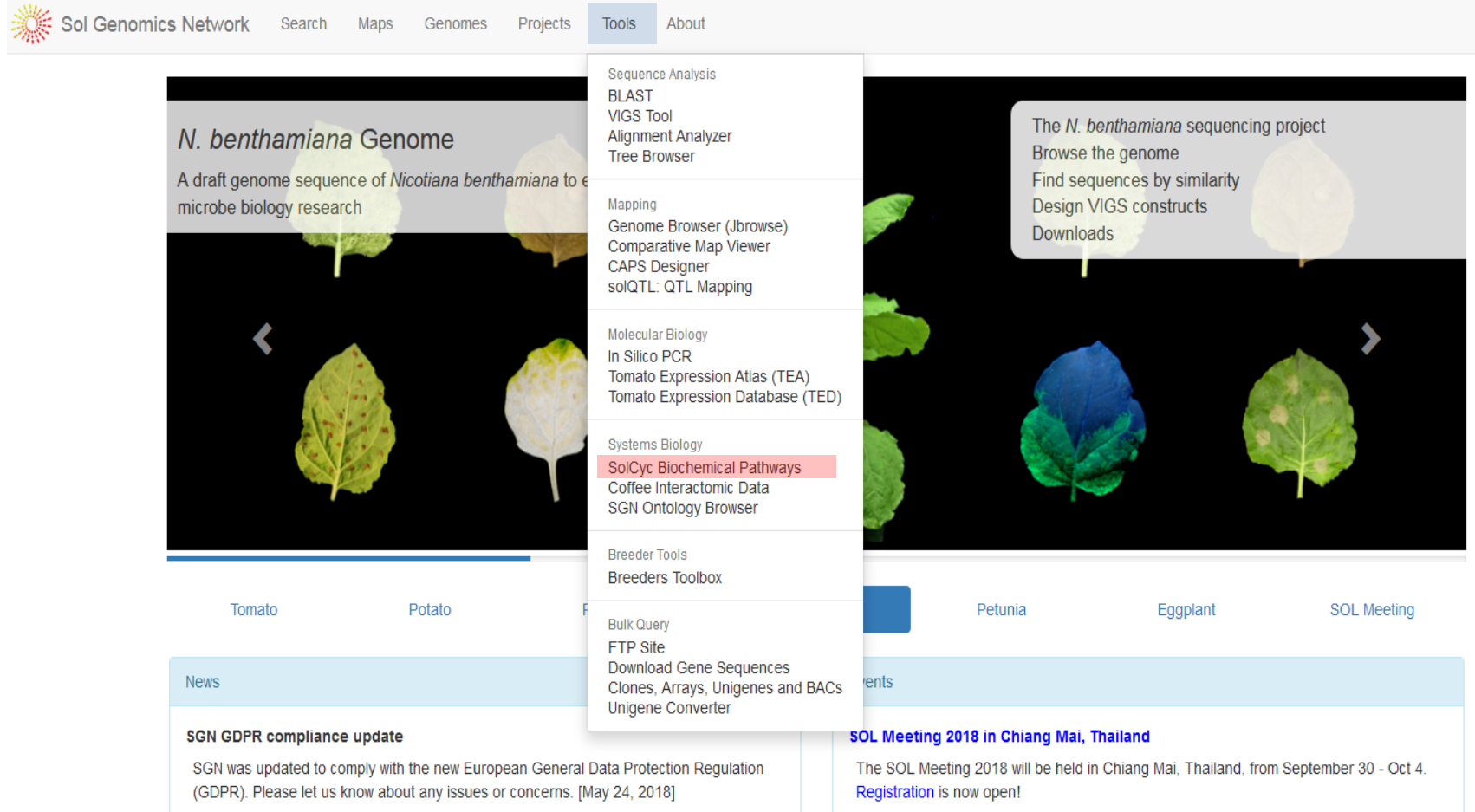
sol genomics network





# Accessing SolCyc

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



The screenshot shows the Sol Genomics Network homepage. The 'Tools' menu is open, displaying a list of tools categorized into Sequence Analysis, Mapping, Molecular Biology, Systems Biology, Breeder Tools, Bulk Query, and FTP Site. The 'SolCyc Biochemical Pathways' option is highlighted in red. The background features a large image of a tomato leaf with a text overlay about the *N. benthamiana* genome. Below the main image are navigation tabs for Tomato, Potato, Petunia, Eggplant, and SOL Meeting. A news section at the bottom left contains a 'SGN GDPR compliance update'.

**Sol Genomics Network** Search Maps Genomes Projects **Tools** About

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

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  
Breeder's Toolbox

Bulk Query  
FTP Site  
Download Gene Sequences  
Clones, Arrays, Unigenes and BACs  
Unigene Converter

The *N. benthamiana* sequencing project  
Browse the genome  
Find sequences by similarity  
Design VIGS constructs  
Downloads

Tomato Potato Petunia Eggplant SOL Meeting

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!





# 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



sol genomics network





# 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 Foerster<sup>1</sup>, Aureliano Bombarely<sup>2</sup>, James N.D. Battey<sup>3</sup>,  
Nicolas Sierro<sup>3</sup>, Nikolai V. Ivanov<sup>3</sup> and Lukas A. Mueller<sup>1,\*</sup>**

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



sol genomics network



# Information ‘at a glance’ of curated pathways

## NicotianaCyc Pathway: solanesol biosynthesis

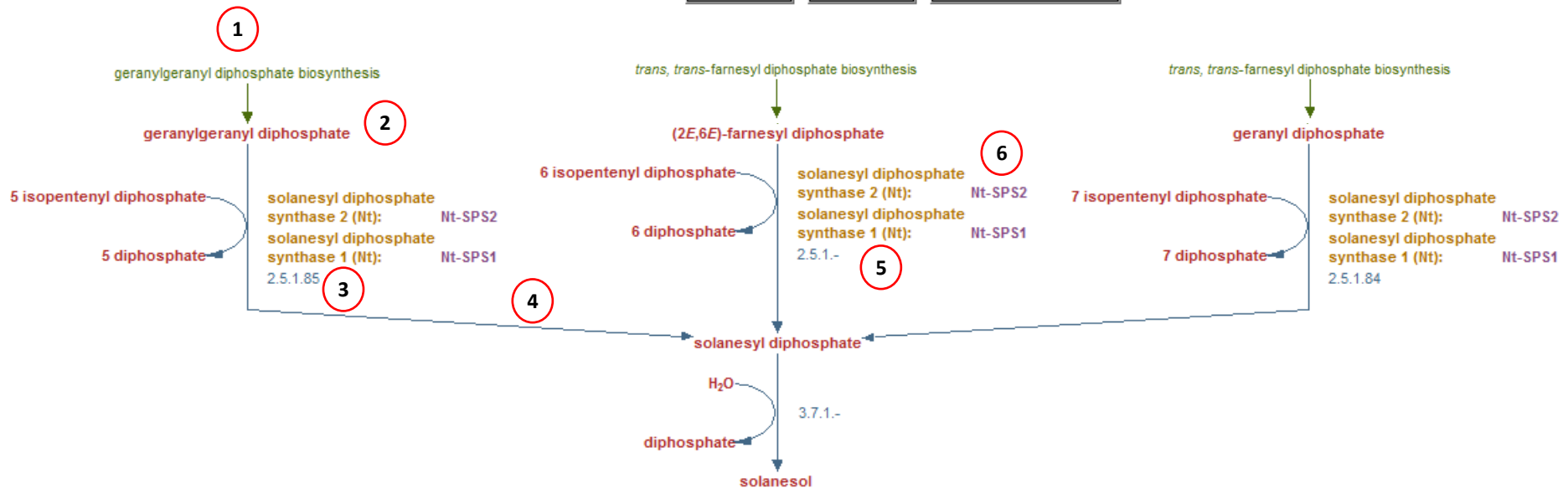
All Organisms

More Detail

Less Detail

Species Comparison

7



sol genomics network



# 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 **methylethylerythritol 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

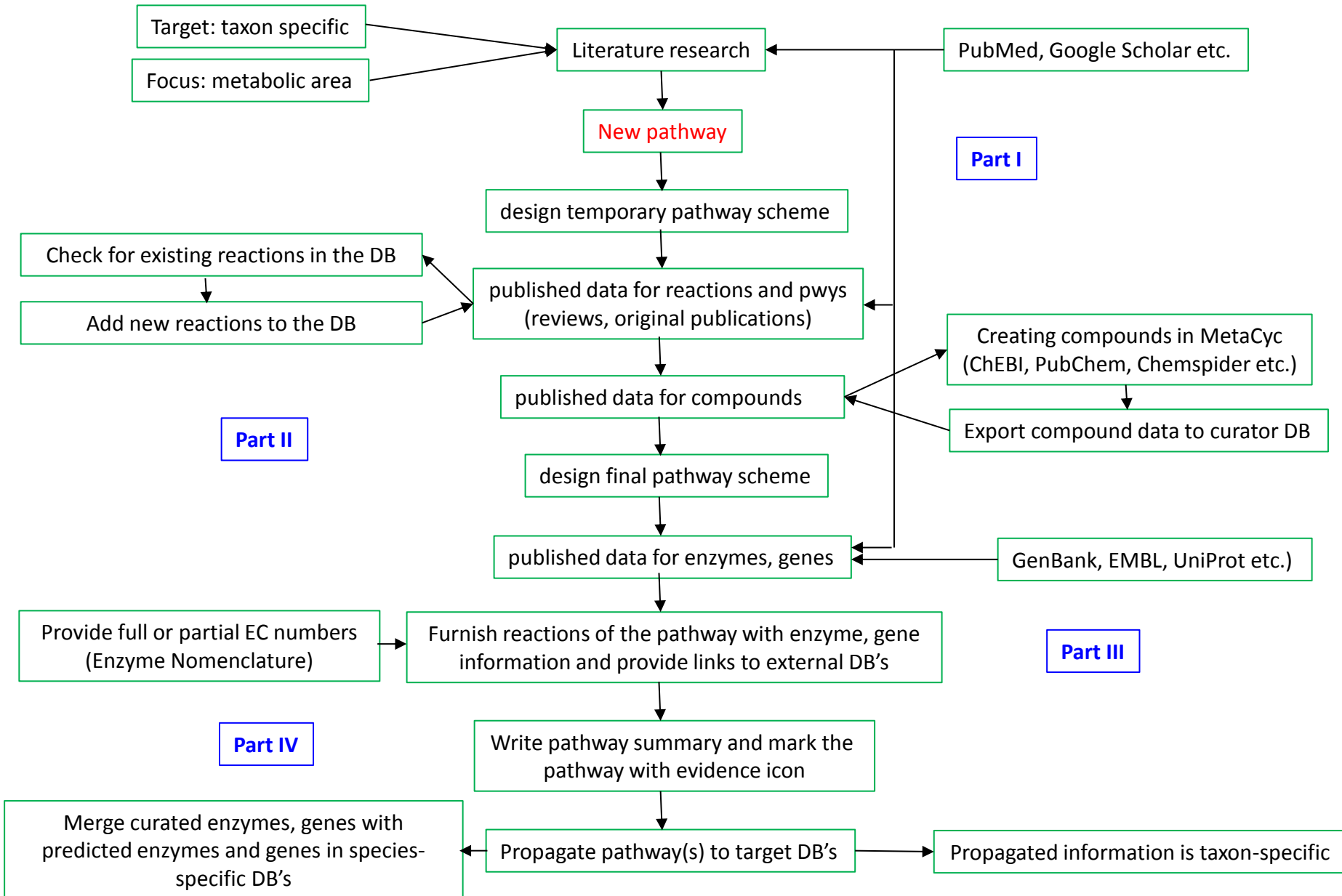
## References

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, Chojnacki 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(4):545-554. PMID: 19444444
- 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 analysis of two 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* 170(4):1455-1465. PMID: 27044444
- Fukusaki04: Fukusaki E, Takeno S, Bamba T, Okumoto H, Katto H, Kajiyama S, Kobayashi A (2004). "**Biosynthetic pathway for the C45 polyprenol, solanesol, in Arabidopsis thaliana.**" *Plant Physiol* 135(4):1455-1465. 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* 133(4):1455-1465. PMID: 12744444
- 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(4):1455-1465. 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* 161(4):1455-1465. PMID: 23444444
- 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* 135(4):1455-1465. PMID: 15653808



# 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



sol genomics network



**METACYC**

A member of the BioCyc database collection

**SRI International**