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)

BTI BOYCE THOMPSON INSTITUTE

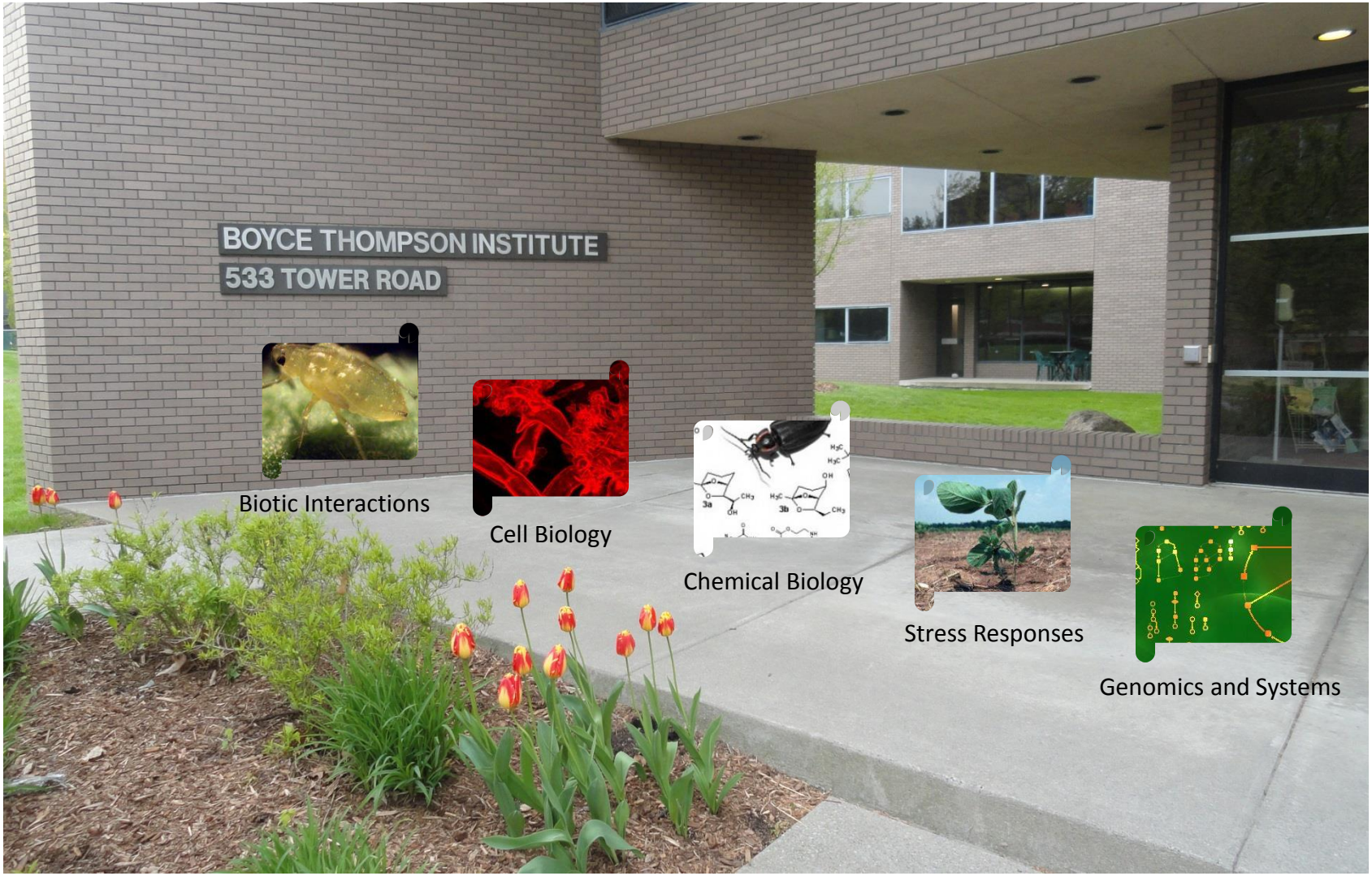
sol genomics network

SRI International

PHILIP MORRIS INTERNATIONAL

METACYC
A member of the BioCyc database collection

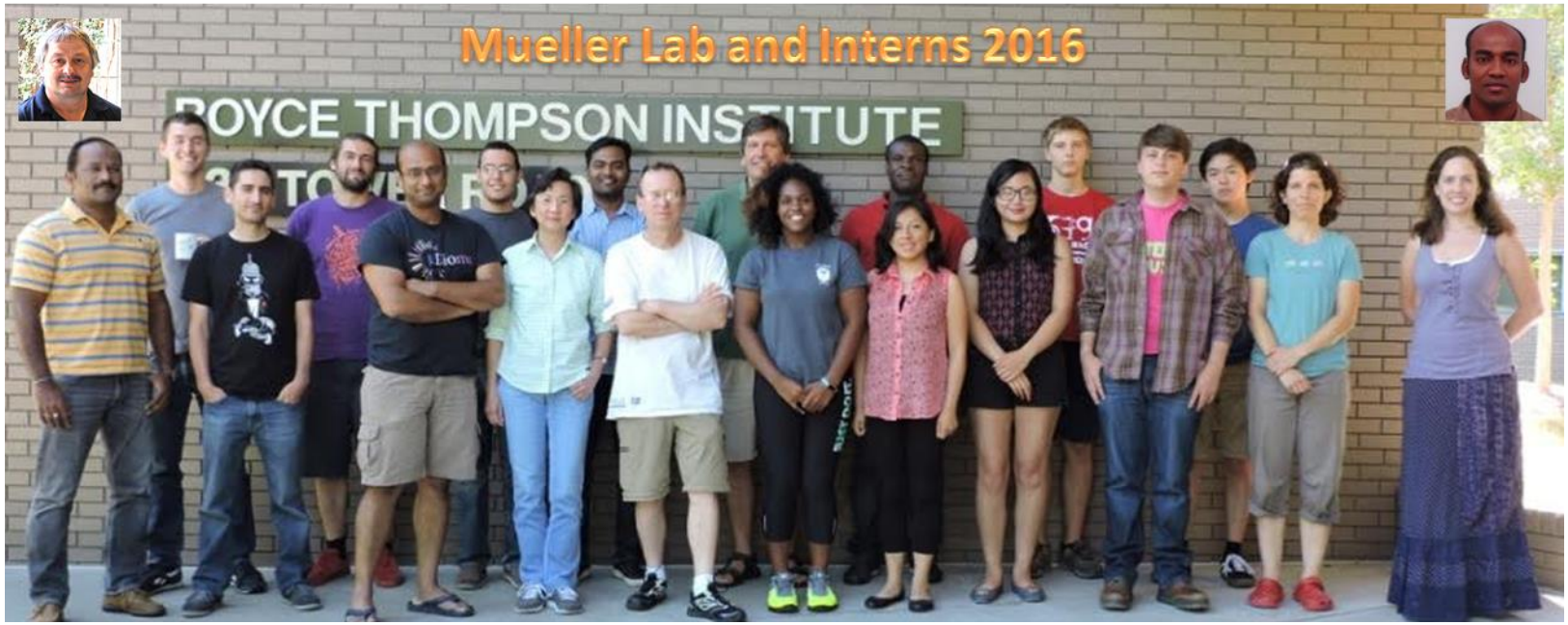
Boyce Thompson Institute (BTI)



BTI BOYCE THOMPSON INSTITUTE

https://btiscience.org/our-research/research-areas/

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)

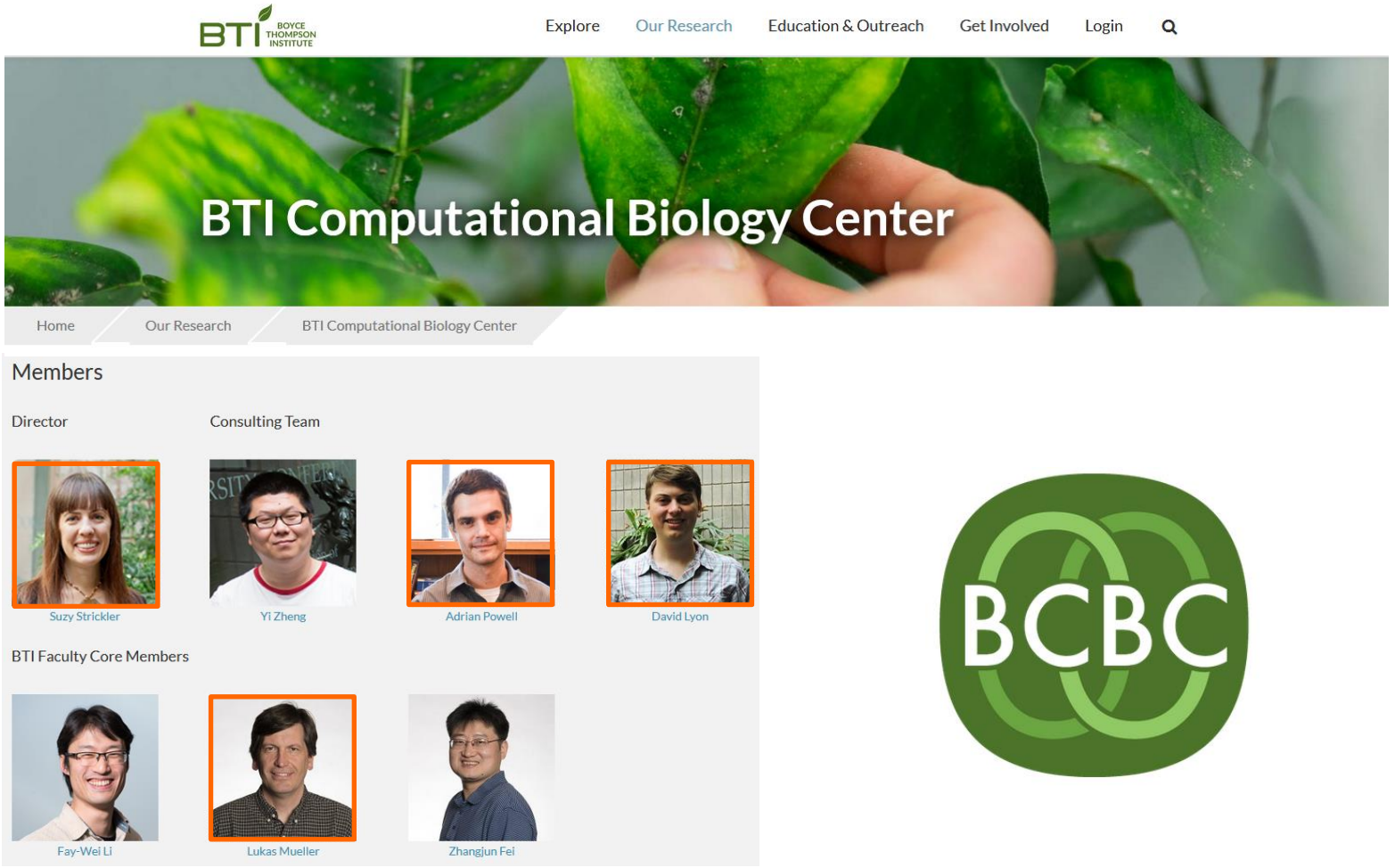
Genome Sequencing Projects (*Solanacea* and related *Asterids*)

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

sol genomics network

BTI BOYCE THOMPSON INSTITUTE

BTI Computational Biology Center (BCBC)

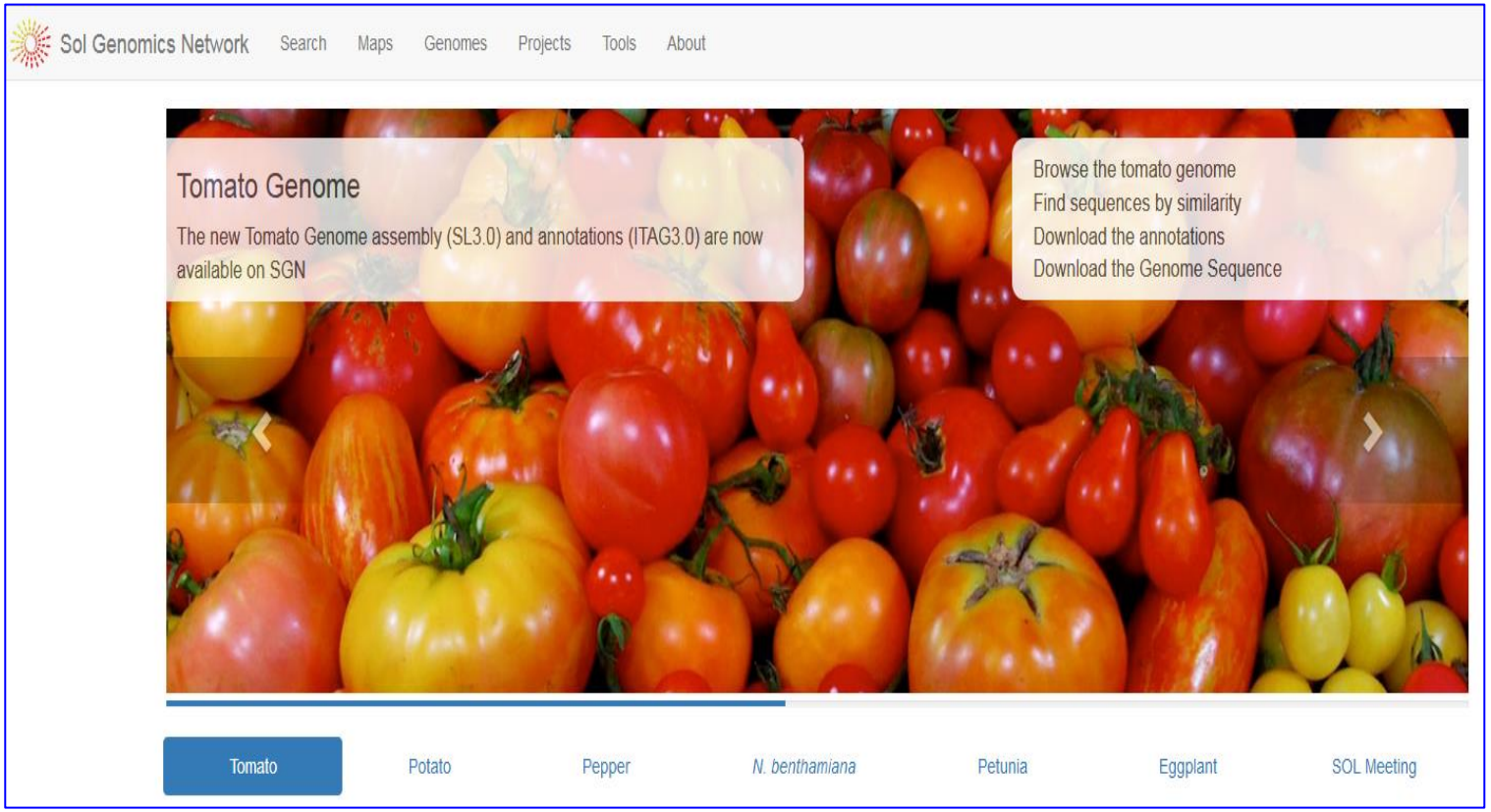


BCBC

BTI BOYCE THOMPSON INSTITUTE

https://btiscience.org/our-research/computational-biology/

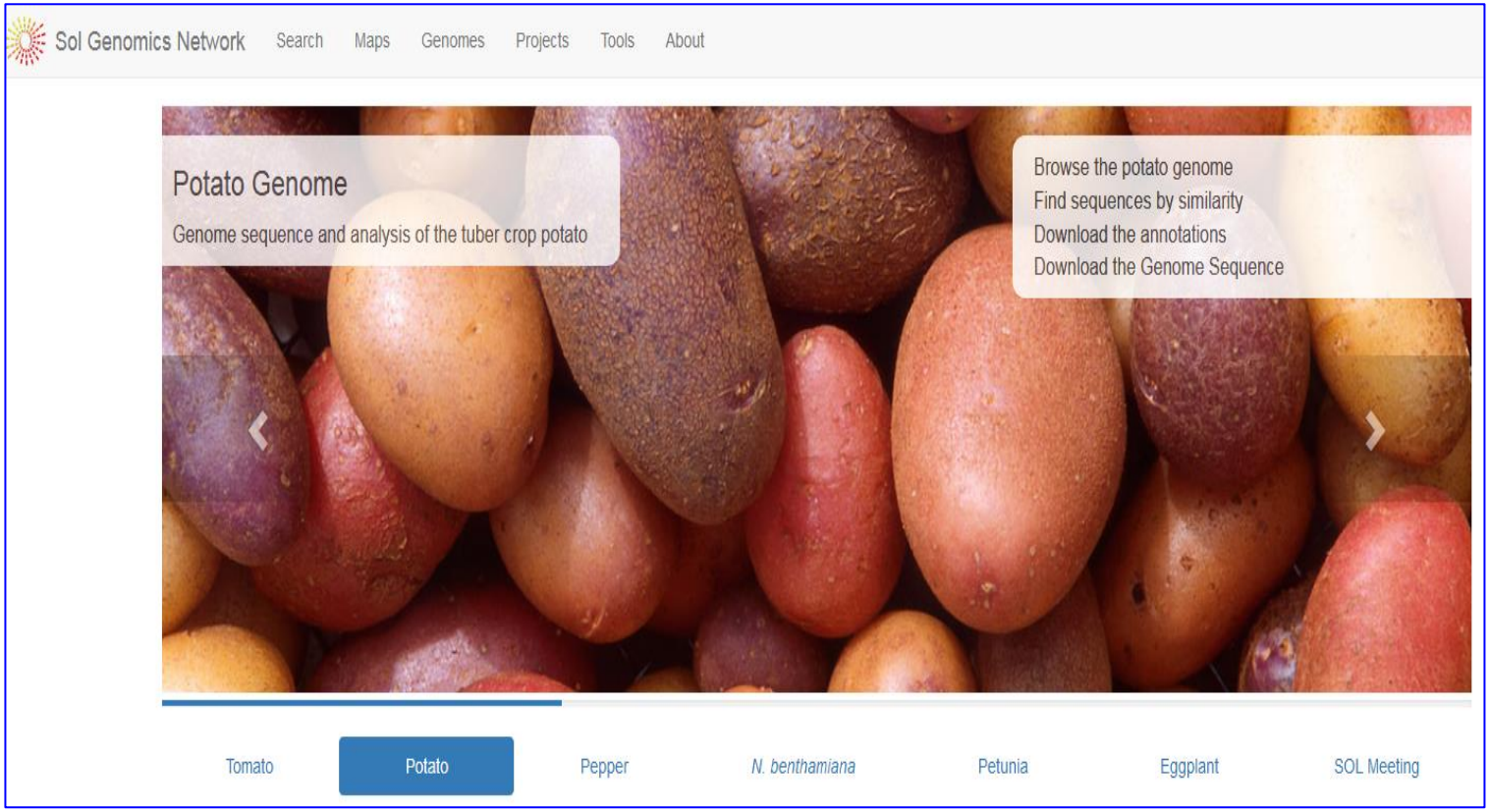
The Sol Genomics Network (https://solgenomics.net/)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

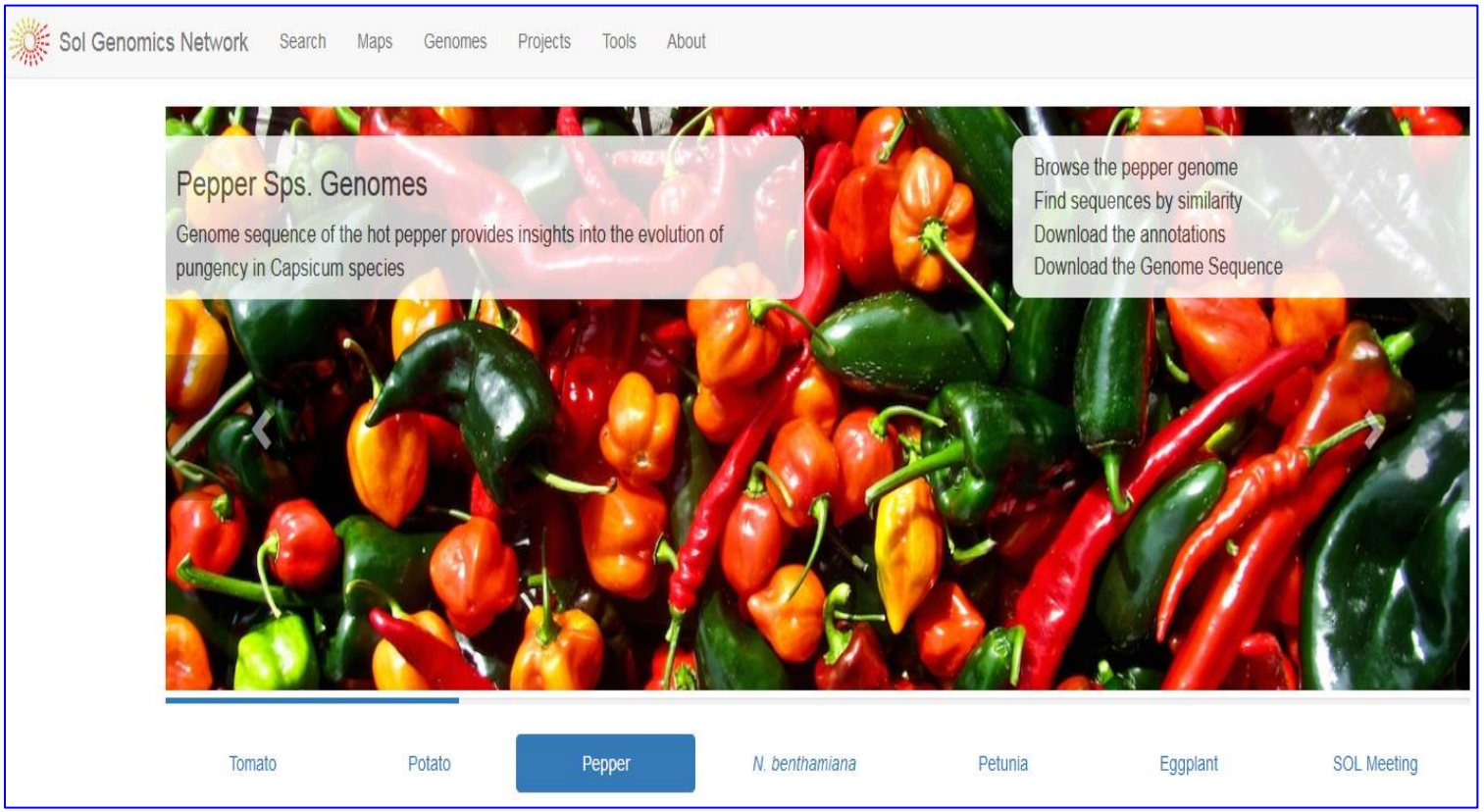
The Sol Genomics Network (https://solgenomics.net/)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

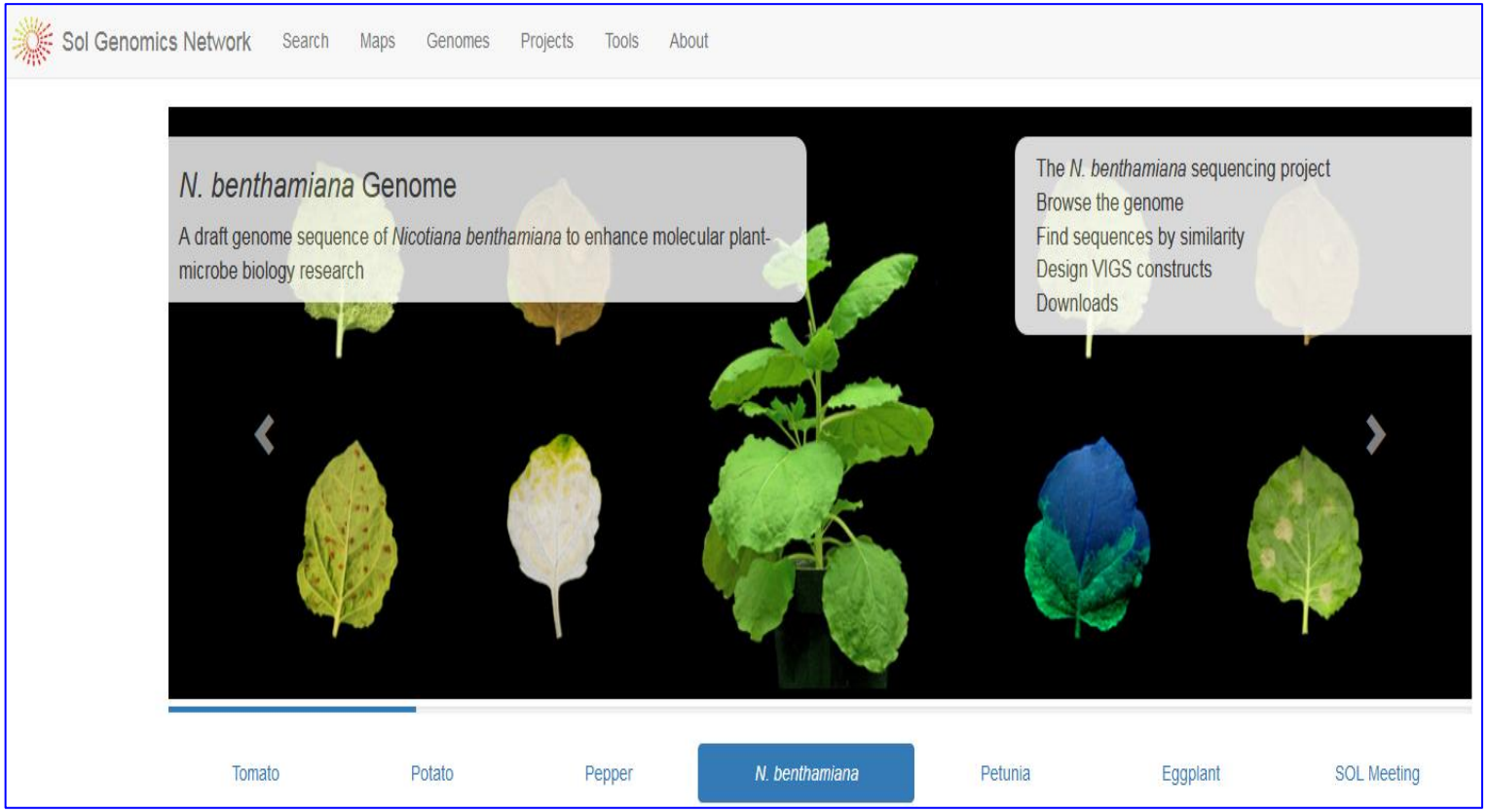
The Sol Genomics Network
(https://solgenomics.net/)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

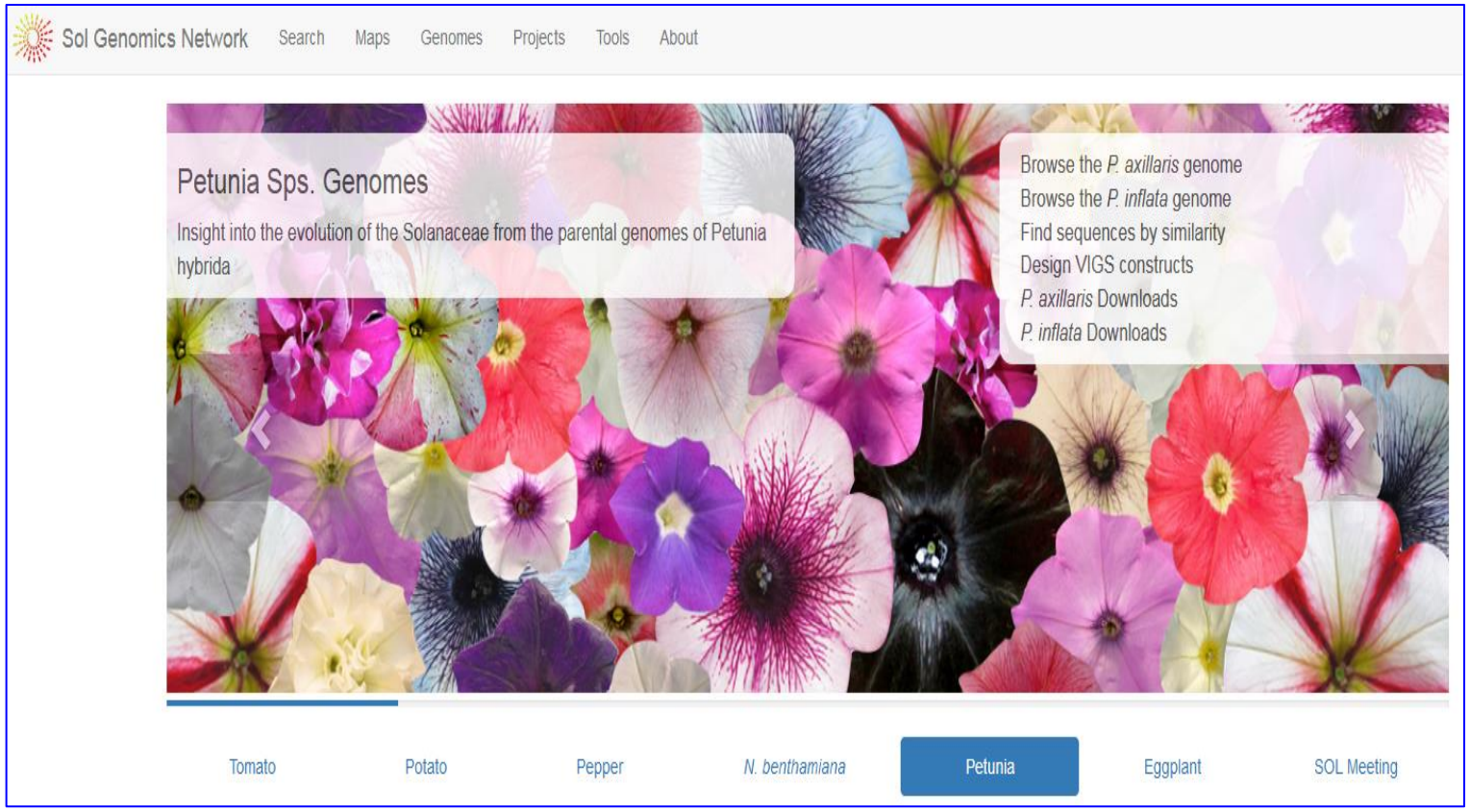
The Sol Genomics Network
(https://solgenomics.net/)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

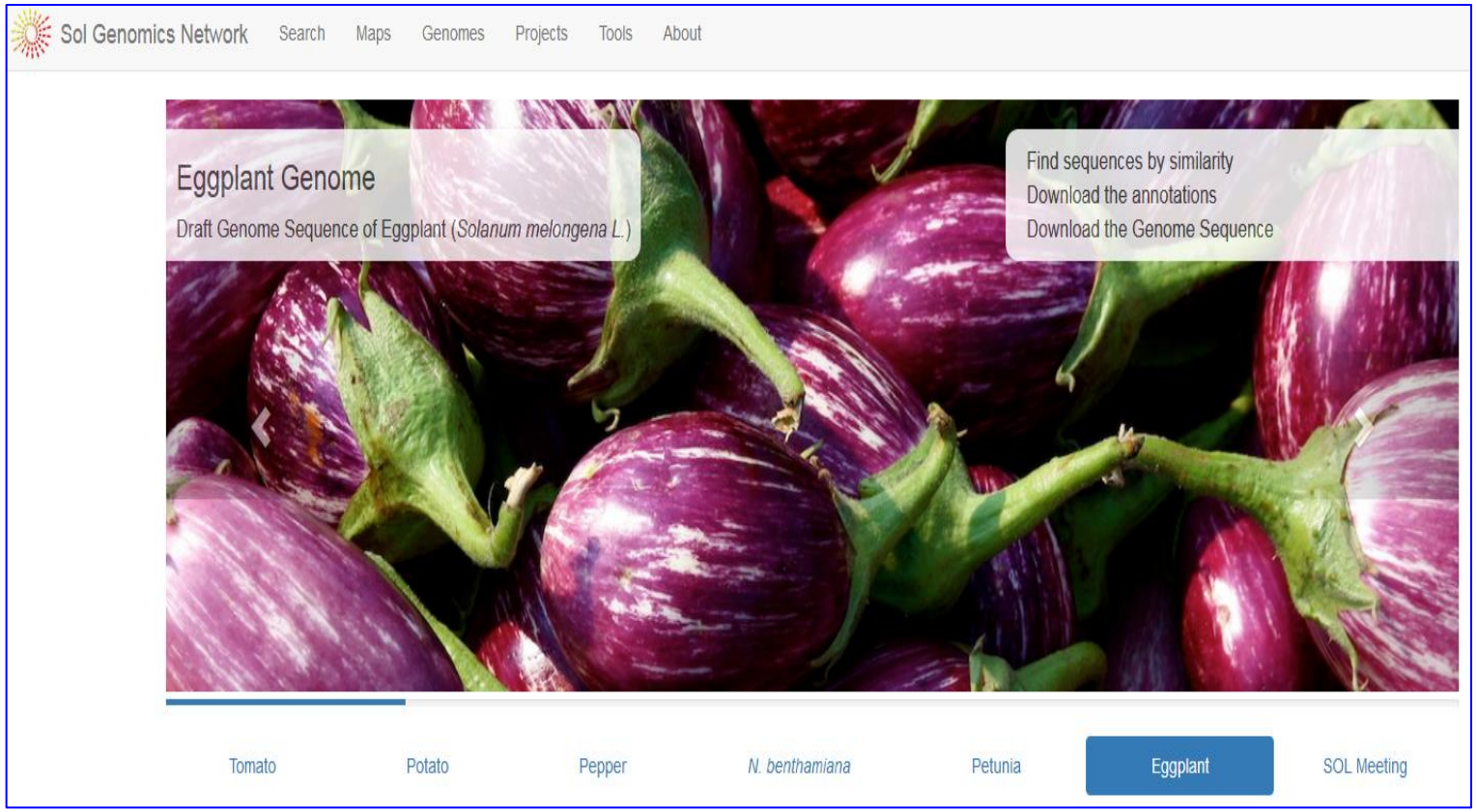
The Sol Genomics Network
(https://solgenomics.net/)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

The Sol Genomics Network
(https://solgenomics.net/)



sol genomics network

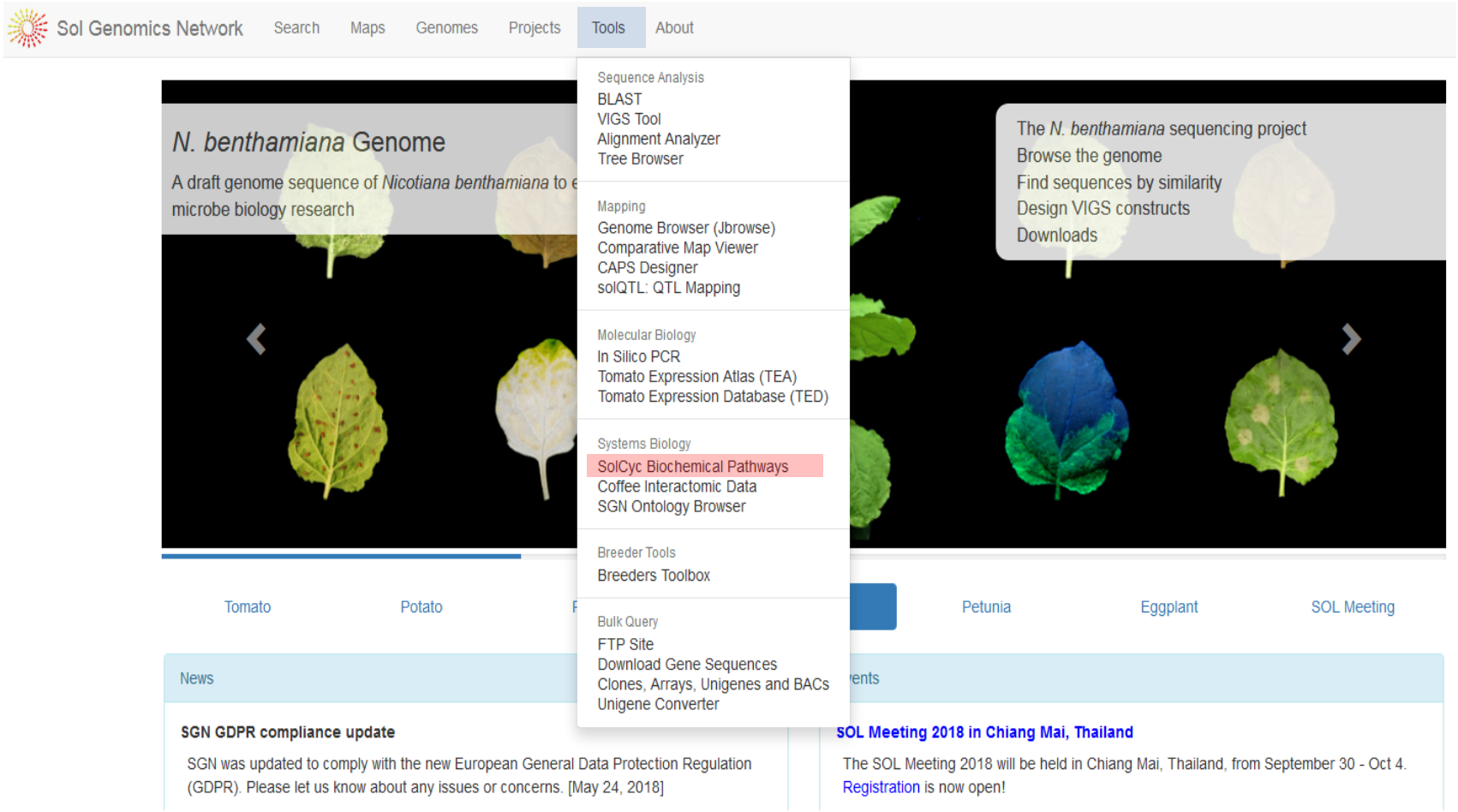
BTI BOYCE THOMPSON INSTITUTE

The Sol Genomics Network
(https://solgenomics.net/)

sol genomics network

BTI BOYCE THOMPSON INSTITUTE

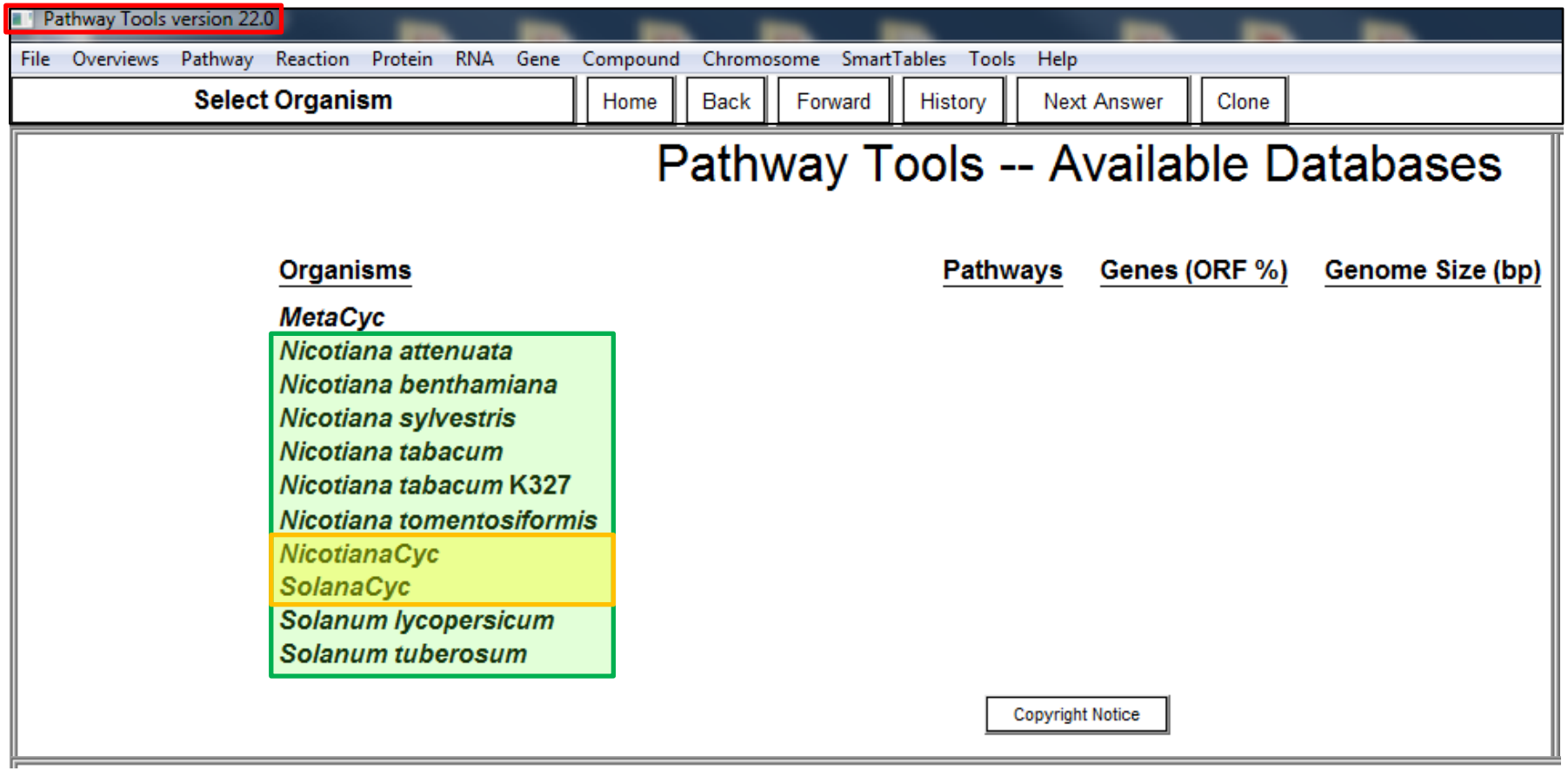
Accessing SolCyc (https://solgenomics.net/tools/solcyc/index.pl)



sol genomics network

BTI BOYCE THOMPSON INSTITUTE

Actively curated *Solanaceae* Cyc's at SGN



sol genomics network

MetaCyc
A member of the BioCyc database collection

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

SolanaCyc	Pathways	Reactions	Enzymes	Transporter	Compounds
version					
1.0	169	835	257	1	1441
2.0	246	1093	491	72	1769
5168 citations					
NicotianaCyc	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

MetaCyc
A member of the BioCyc database collection

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

DATABASE
The Journal of Biological Databases and Curation

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

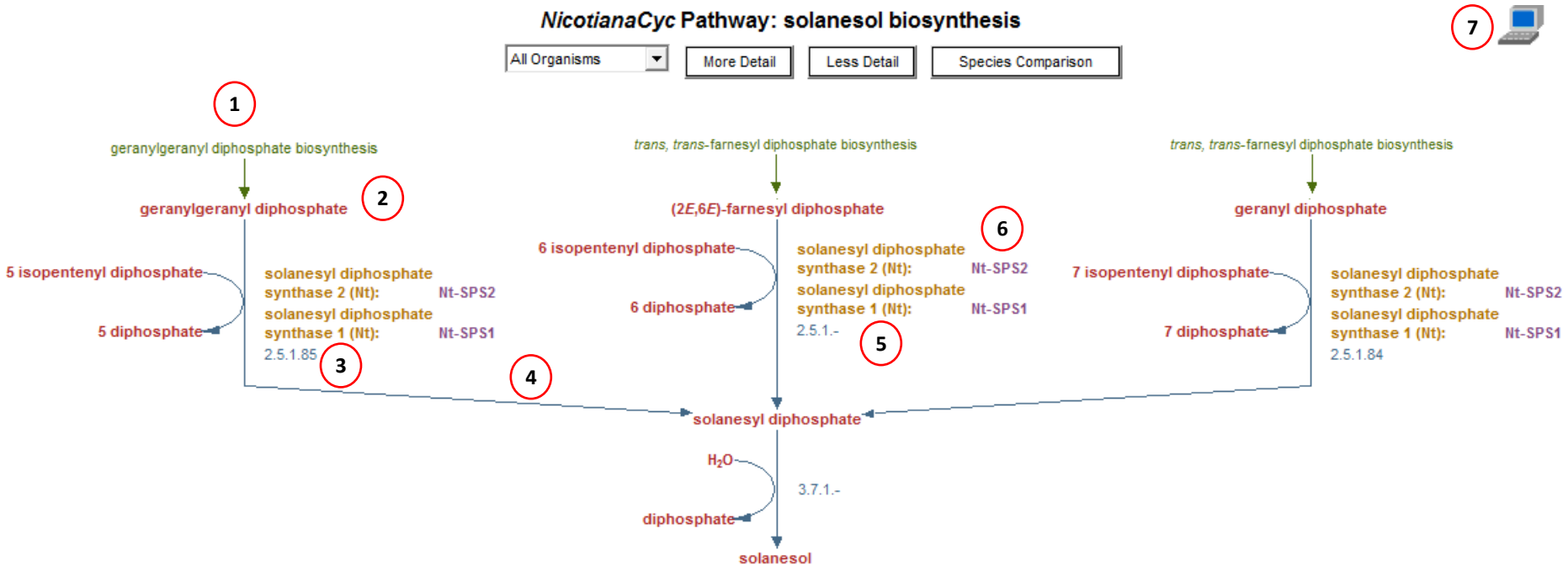
Hartmut Foerster[1], Aureliano Bombarely[2], James N.D. Battey[3], Nicolas Sierro[3], Nikolai V. Ivanov[3] and Lukas A. Mueller[1,*]

[1]Boyce Thompson Institute, 533 Tower Road, Ithaca, New York, 14853-1801, USA, [2]Department of Horticulture, Virginia Polytechnic Institute and State University, 220 Ag Quad Lane, Blacksburg, VA 24061, USA and [3]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

METACYC
A member of the BioCyc database collection

Information ‘at a glance’ of curated pathways



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 **methylerythritol 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(

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). **"Fu 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**

Fukusaki04: Fukusaki E, Takeno S, Bamba T, Okumoto H, Katto H, Kajiyama S, Kobayashi A (2004). **"Biosynthetic pathway for the C45 polyprenol, solanesol,** 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 synth**

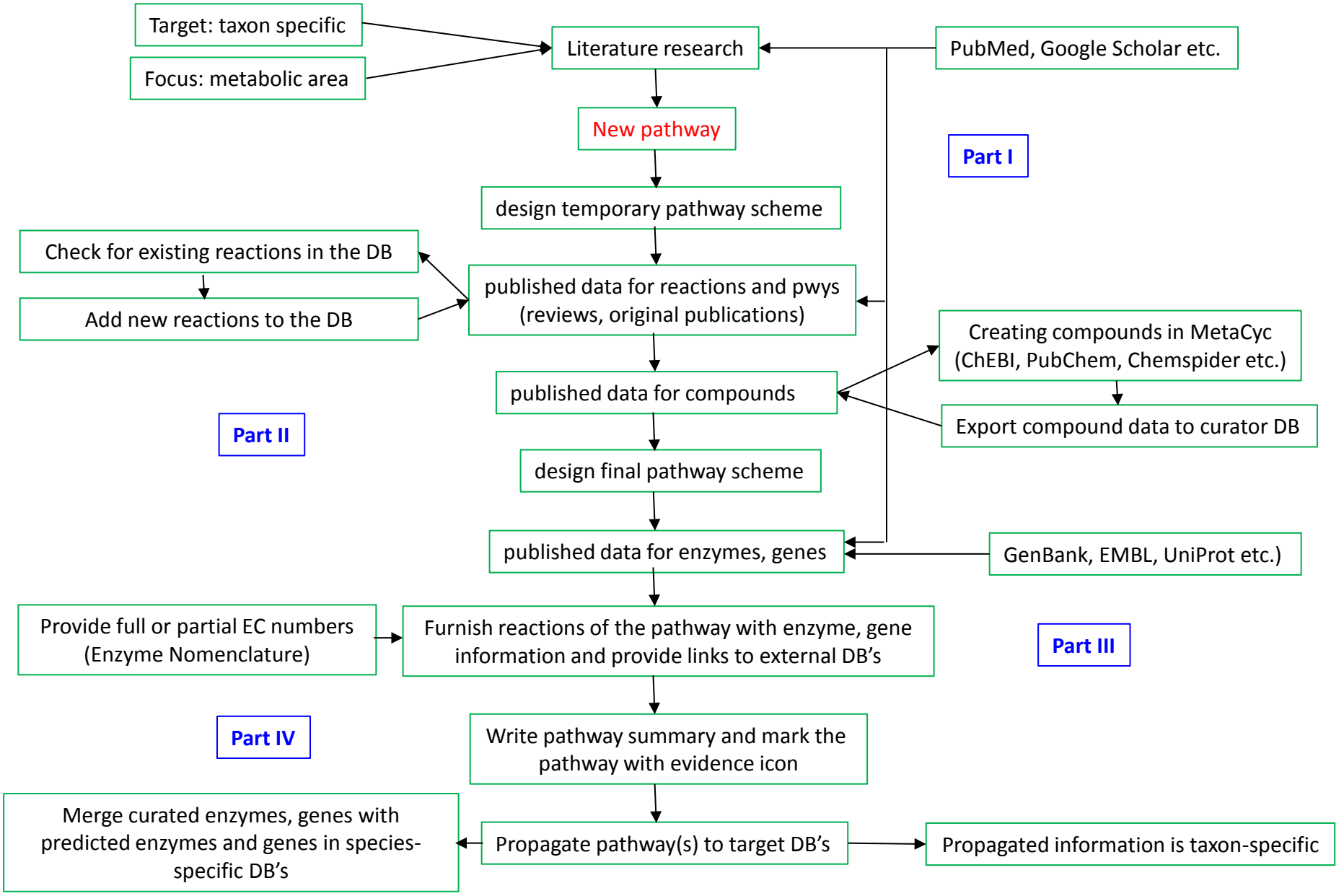
Hirooka05: Hirooka K, Izumi Y, An CI, Nakazawa Y, Fukusaki E, Kobayashi A (2005). **"Functional analysis of two solanesyl diphosphate synthases from Arabi** PMID: 15784989

Ishikawa16: Ishikawa S, Kanemaru Y, Nara H, Erami K, Nagata Y (2016). **"Assessing the mutagenic activities of smoke from different cigarettes in direct exp 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 f**

Jun04: Jun L, Saiki R, Tatsumi K, Nakagawa T, Kawamukai M (2004). **"Identification and subcellular localization of two solanesyl diphosphate synthases fro** 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

NIH NSF USDA PM PHILIP MORRIS INTERNATIONAL

BTI BOYCE THOMPSON INSTITUTE sol genomics network METACYC A member of the BioCyc database collection SRI International