



The SOL Genomics Network model: making community annotation work

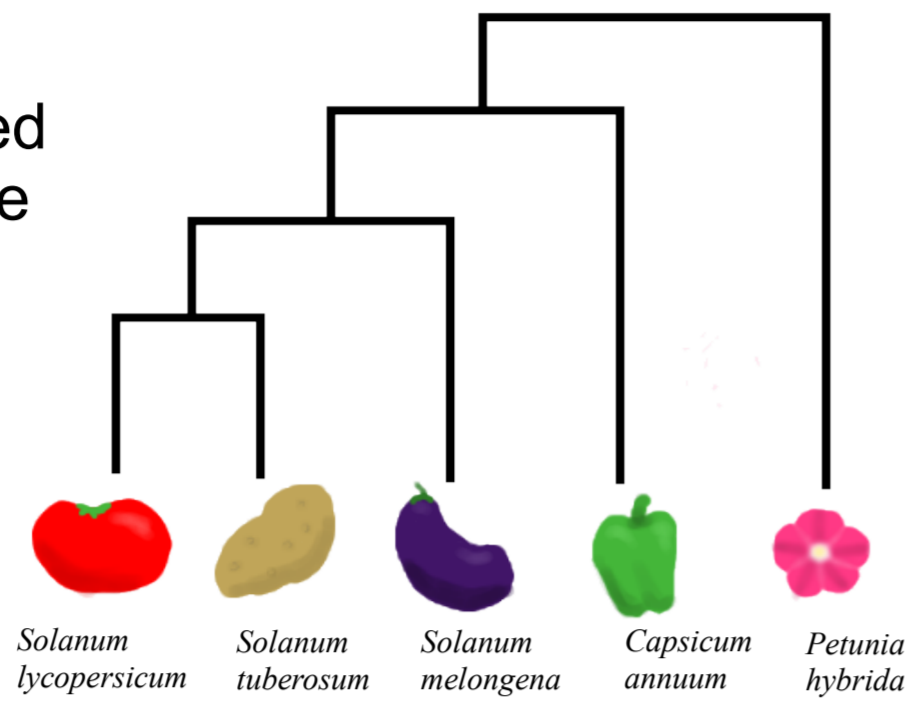


Naama Menda, Isaak Tecle, Anuradha Pujar, Aureliano Bombarely, Tom York, Adri Mills, Joseph Gosselin, Robert Buels and Lukas Mueller

What is SGN?

<http://sgn.cornell.edu/>

SGN is a community centric, clade-oriented database and comparative resource for the plants of the Solanaceae family. SGN houses maps, markers, sequence, locus annotation, and phenotypic data.

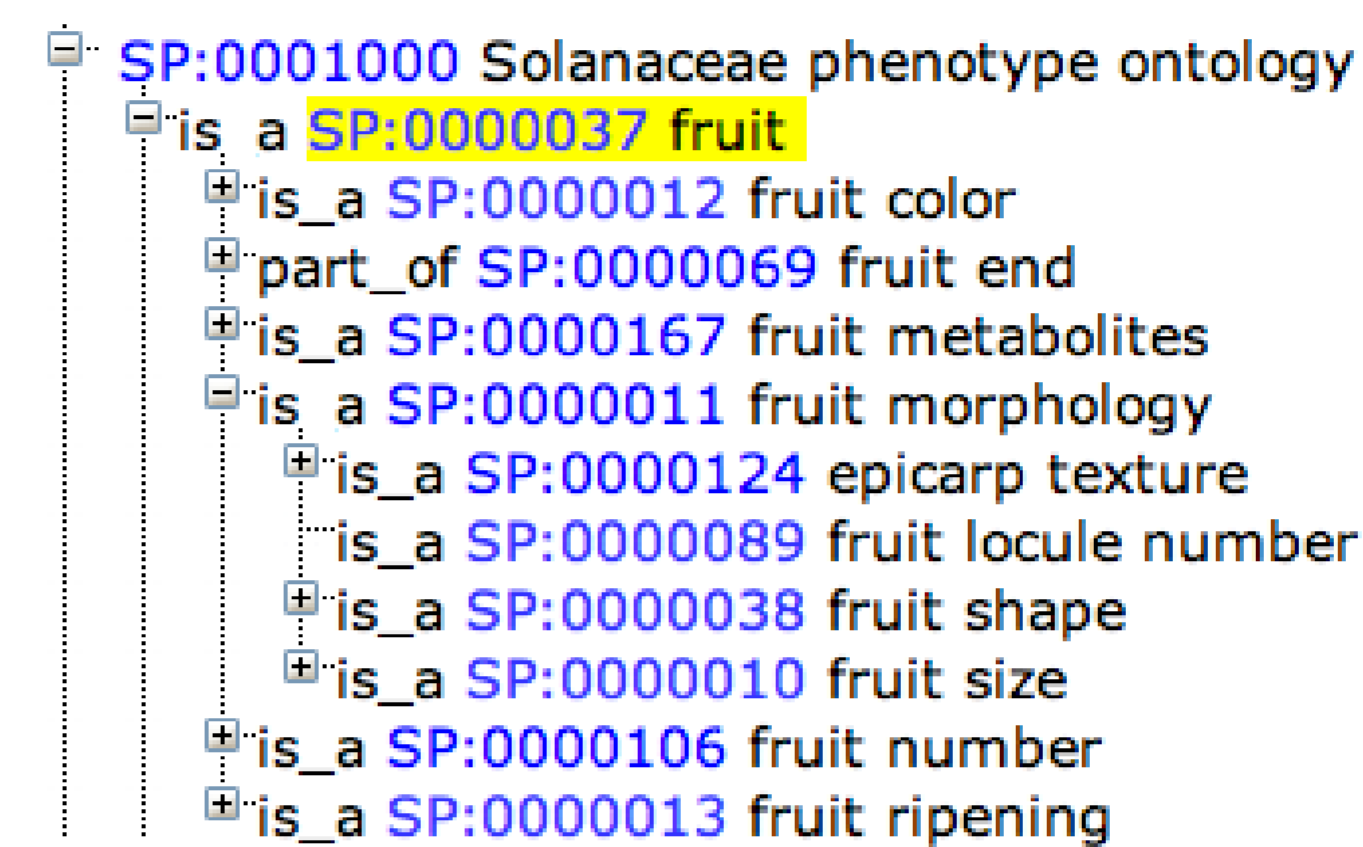


SGN is also part of the bioinformatics platform for the International Solanaceae Project (SOL), which is currently sequencing the euchromatic portion (about 25%) of the 850 Mb tomato genome.

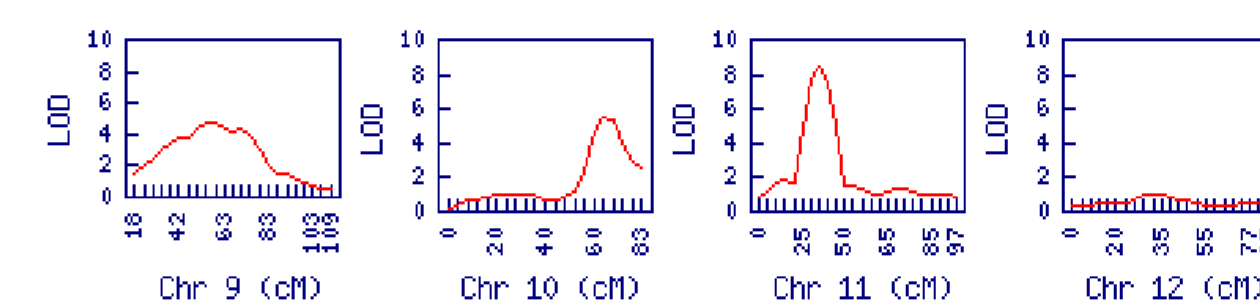
With an SGN submitter account you can contribute your data to SGN easily on the web!

Phenotype Ontology

Solanaceae-specific ontology terms, used mostly by plant biologists and breeders, for describing quantitative and qualitative traits. These terms are mapped to PO, TO, and PATO whenever applicable, allowing cross-database queries.



SGN: fruit shape index values in population QTL Tomato Yellow Stuffer x LA1589 F2



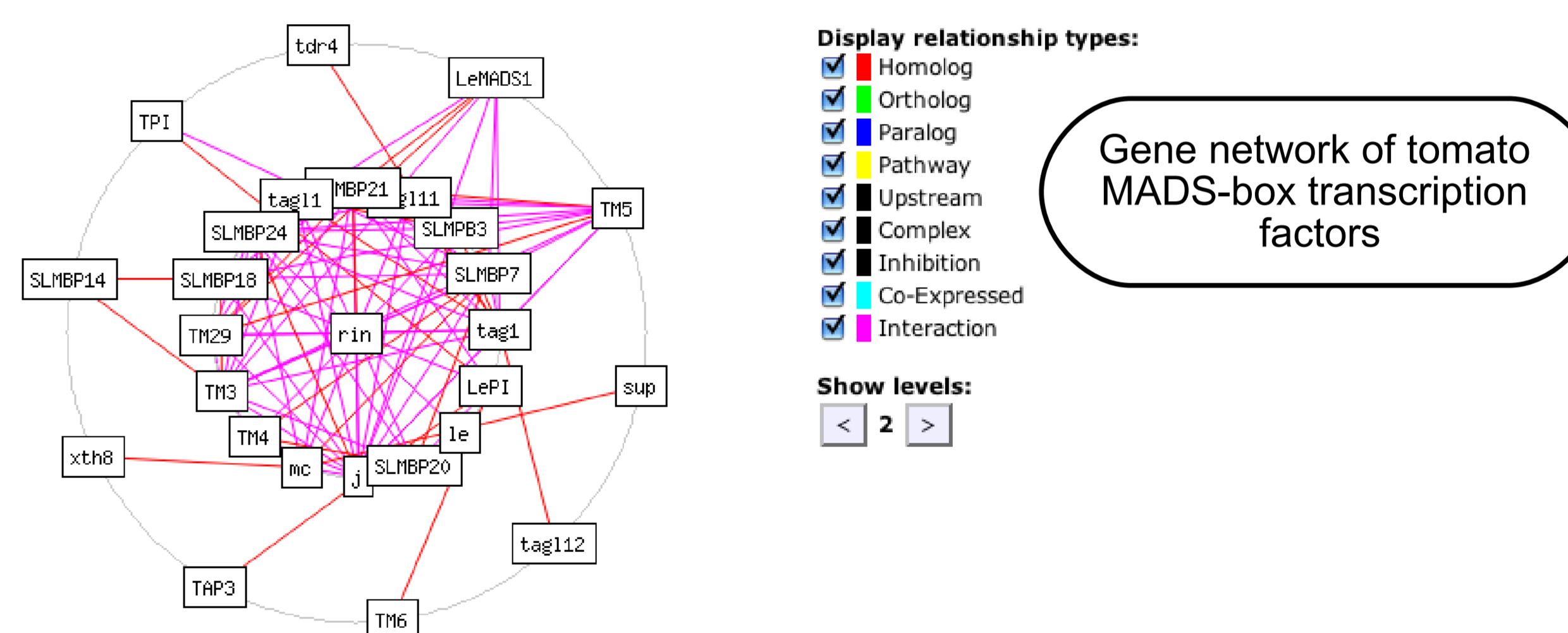
QTLs are calculated by trait. Users can upload their dataset and run R/QTL statistics and map online their loci.

From Genomes to Phenomes

Solanaceae share conserved genomes, yet show rich phenotypic diversity. These characteristics allow us to address one of today's greatest challenges in biology: how do we reveal the phenotypes encoded by genome sequences? The clade-oriented approach is utilized for studying related species in one database.

Gene networks

Since genes occur in networks, we can learn about putative function and phenotypes based on the relationship of one gene with its orthologs, paralogs, interacting genes, and pathway members. We have developed a 'network-browser' tool for visualising the complexity of gene-to-gene relationships.



Literature mining

Publications on Solanaceae genes are indexed and matched against loci descriptors and controlled-vocabulary terms. The result is a ranked list of matching publications and a suggested list of ontology terms associated with each locus.

Community annotation

SGN has a web-interface shared by curators and community-users for annotating genes and phenotypes. This system keeps the database up-to-date, and creates active interactions among researchers.

Conclusions and future development

We have experienced positive responses for our efforts in recruiting the knowledge and expertise of the research community for gene and phenotype annotations. SGN has more than 70 active submitters who curate their favourite genes, submit phenotyping experiments, and assist in development of the phenotype ontology. We are currently seeking ways of increasing interest and participation of researchers and breeders in the task of community-driven annotation. We would also like to collaborate with Journals, and have authors submit gene data to SGN as part of the peer-reviewed publication process.

The emerging tomato genome sequence will serve as the reference genome for other Solanaceae species. The resulting increase in data and usage will be a test for SGN's ability to scale-up while continuing to closely interact with the community.

Funding sources

The following funding agencies are gratefully acknowledged:



<http://sgn.cornell.edu>

Contact SGN at sgn-feedback@sgn.cornell.edu

SGN code and software

Open Source and free to download!
The SGN community platform is portable for other organisms.
<ftp://ftp.sgn.cornell.edu>