Of tomatoes, genomes and databases (http://solgenomics.net/)

Lukas A. Mueller, Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, New York 14853. lam87@cornell.edu

As genome data accumulates at a rapid pace for many different species, systems are needed that allow the use of this wealth of information for the improvement of germplasms. Many Solanaceae genomes are currently being sequenced; both potato and tomato have high quality draft sequences available, with pepper and *Nicotiana benthamiana* available soon. In addition, a number of wild tomatoes are being sequenced, such as *Solanum pimpinellifolium* and *Solanum pennellii*. Despite the rich sequence information, a persistent problem is to link it to phenotypes, which will ultimately help breeders generate better varieties.

To bridge genotypes and phenotypes, the Sol Genomics Network (SGN, http://solgenomics.net/) has created databases for storing both genomes and phenomes. SGN is a Clade Oriented Database (COD) containing available genomic, genetic, phenotypic and taxonomic information for plant genomes of the Solanaceae, selected species in the Euasterid clade, and important model species such as Arabidopsis and rice, in a comparative fashion. The function of sequences is described in the locus database, which contains information on loci including enzymatic activities, sub-cellular localization of gene products, tissue specific expression, and publications describing the gene. Another database of phenotypes describes plant germplasm, populations, and phenotyping and genotyping experiments. For each accession, trait measurements and genotypic data can be stored, as well as metadata, including images, ontology-based annotations, and literature references.

New links between sequences and phenotypes can be either edited manually, for example based on specific laboratory experiments, or newly discovered using the solQTL tool. The solQTL tool operates on F2 and backcross populations for which both genotypic and phenotypic information is available in the database, and calculates on the fly the QTL positions on the genetic map. Users may also upload populations of their own and run the QTL analysis using user-defined statistical parameters.

The description of traits is based on ontologies that have been developed at SGN to describe general and Solanaceae-specific biology, with mappings to standard ontologies such as the Plant Ontology (http://plantontology.org/). The Solanaceae Phenotype Ontology (SP) currently contains more than 340 terms and is expanded to cover new traits of interest to the community as needed. Currently, most data in the phenotypic database come from the SolCAP project, a large USDA funded project with the goal of genotyping and phenotypic large panels (400+ accessions) for potato and tomato (http://solcap.msu.edu/). Independent researchers also have contributed their data from natural and induced variation.

SGN is unique in being a community curated database. This means that researchers around the world can contribute annotations to genes and phenotypes simply by obtaining an SGN account. Once a user has been approved with editor privileges for a locus or an accession, he or she has control over its details. Community curation provides a more scalable approach to curating biological data.

In the future, SGN will focus on adapting the database to interface more easily with association genetics approaches, and supporting newer breeding strategies such as Genomic Selection. SGN can be accessed at http://solgenomics.net/.