

Sol Genomics Network

Associating genes with phenotypes

Designer: Naama Menda
Presenter: Robert Buels



Cornell University



sol genomics network



What is the Sol Genomics Network?

www.sgn.cornell.edu

- Clade-oriented database for solanaceous species (tomato, potato, pepper, ...), and coffee.
- Bioinformatics hub for tomato genome sequencing

Overall Stats

17% of sequencing is complete

12% of BACs are reported finished

9% of BACs are available for download





Solanaceae genes and phenotypes

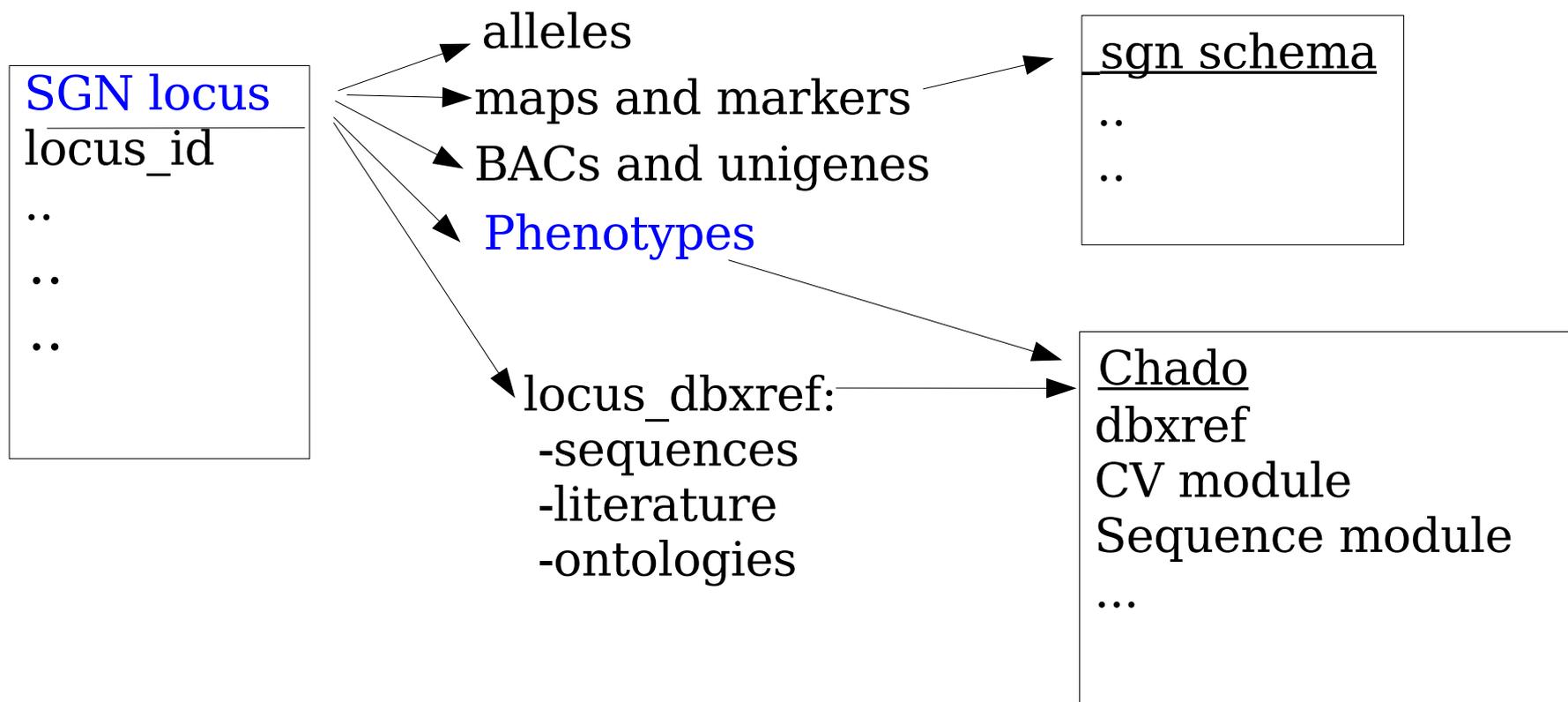
- Database schema
- User interface:
 - Gene search and detail pages
 - Phenotype data
 - User-editable database
- Summary: Gene–phenotype associations





Database schema

- Loci may have multiple phenotypes
- Phenotypes are individuals of populations (mutants, mapping, introgression lines..) May be associated with multiple loci.





SGN gene search page

Genes

Phenotypes

Unigenes

Markers

BACs

Annotations

ESTs and Libraries

People

Gene search

Search for any locus or allele

contains

[\[Submit new locus\]](#)

Advanced search options

Organism

Chromosome / Linkage Group

Locus editor

Show only genes with

- sequence
- markers
- GO/PO annotation

Phenotype

[\[Advanced SGN phenotype search\]](#)

Ontology term

(Term name or ID: e.g. 'carotenoid' or 'PO:0007010')

Genbank ID

(Accession or GI: e.g. 'EF091820' or '118185006')

Search

Search for name, symbol synonym, description, gene-activity.

Search for genes with a recorded phenotype. Advanced search (link) includes phenotypes not associate yet with loci.

Tomato
Potato
(eggplant, pepper, petunia, coffee..)
Locus submitter (has edit privileges)
Not all genes are associated with a sequence, map position, or ontology annotation

PO/GO
id/term/synonym/definition



SGN gene display page

You are logged in as **Naama Menda**
[Home](#) | [My SGN](#) | [Logout](#) | [Forum](#) | [Contact](#) | [Help](#)

sol genomics network

search maps sequencing tools

Tomato locus name: 'plastid terminal oxidase'

Locus details

[\[New\]](#) [\[Edit\]](#) [\[Delete\]](#)

Locus name	plastid terminal oxidase
Locus symbol	ptox
Gene activity	terminal oxidase
Description	
Chromosome	11
Arm	short

Locus synonyms 2: gh ghost [\[add/remove\]](#)

plastid terminal oxidase is a [TGRC gene](#)

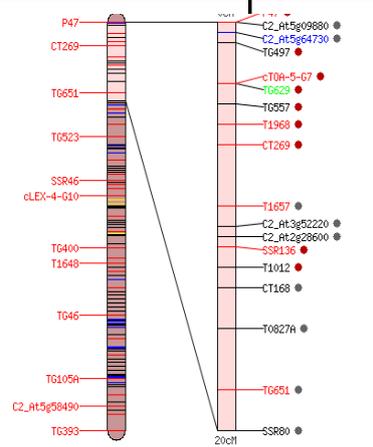
Locus editor: [Naama Menda](#) Last modified on: 2006-06-01



11
short arm

Tomato-EXPEN 2000

link to chr. viewer



Accessions and images

[LA0295](#)      (8 images)

When logged in, edit links appear

Any user can add synonyms

Only the object owner can edit

Images and phenotype data

Allele data- with links to phenotype data (user editable)

Known alleles						
[Add new allele]						
Allele symbol	Allele name	Synonyms	Mode of inheritance	Phenotype	Accessions	
1	ghost	ab	recessive	Ghost phenotype. Incomplete chlorophyll deficiency, starting green, later breaking to white.	3	[Edit]

Sequence annotation:

- SGN unigenes
- GenBank

Sequence annotations	
SGN Unigenes	SGN-U.318588
GenBank accessions	GI:9937100 Lycopersicon esculentum plastid terminal oxidase (PTOX) mRNA, complete cds; nuclear gene encoding plastid protein. GI:10505365 Lycopersicon esculentum plastid quinol oxidase mRNA, complete cds.

Literature annotation

Literature annotation	
PubMed	PMID:10938359

Ontology annotation:
➤ GO and PO
➤ Solanaceae phenotype ontology (under development)

Ontology annotations	
Biological process:	GO:0009266 response to temperature stimulus GO:0009644 response to high light intensity GO:0009657 plastid organization and biogenesis GO:0009658 chloroplast organization and biogenesis GO:0016117 carotenoid biosynthesis
Molecular function:	GO:0009916 alternative oxidase activity
Cellular component:	GO:0009579 thylakoid
Plant structure:	PO:0009001 fruit PO:0009005 root PO:0009046 flower

Any SGN user can add comments

User comments	
No user comments.	
Add comment	



SGN phenotype database

Phenotypes are recorded for 'individuals': accessions from a population

SGN individual: LA0295

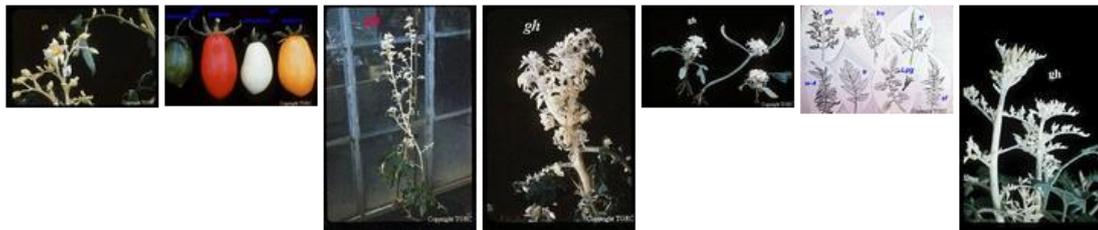
Individual details

[\[New\]](#) [\[Edit\]](#) [\[Delete\]](#)

Individual Name: LA0295
Individual Description:
Population: TGRC monogenic mutant population
Uploaded by: [Roger Chetelat](#)

Only the owner
can edit the
object

Associated loci: [plastid terminal oxidase](#)



Link back to the
locus page

Logged-in users
can submit
images

Mapping data

No map available for this individual.

Known alleles

Locus name	Allele symbol	Phenotype
plastid terminal oxidase	1	Ghost phenotype. Incomplete chlorophyll deficiency, starting green, later breaking to white.

Available germplasms: None

Ontology annotations

Solanaceae phenotype:
SP:0000057 leaf variegation
SP:0000052 white leaf
SP:0000017 chlorophyll content

Mutant
phenotypes are
usually
annotated with
Solanaceae-
specific subset
of descriptors

User comments

No user comments.

[Add comment](#)



Edit Permissions

Permissions are object and user-type dependent:

- Curator
 - Submitter
 - User
- Some objects can be edited only by the owner (**submitter**)
(Locus, allele, accession, image objects have owners.)
- Submitters** can add new objects and annotations to an existing object (load a new image to a locus owned by another submitter).
- **Submitters** can also request ownership of an existing locus object to obtain edit permission (manual email).
 - Other objects can be edited by **users** (synonyms, tags, comments)

Who can become submitter/user?

- Anyone can create an SGN **user** account
- Any **user** willing to load or update data may obtain a **submitter** user-type





User-editable database: The form framework

The locus details are printed in an editable form when clicking the 'Edit' button

Locus details

[New] [Cancel Edit] [Delete]

Locus name *

Locus symbol *

Gene activity

Description

Chromosome

Arm

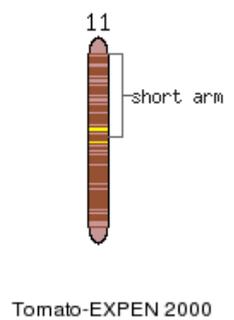
(* denotes required field.)

Locus synonyms 2: gh ghost [\[add/remove\]](#)

plastid terminal oxidase is a [TGRC gene](#)

Locus editor: [Naama Menda](#) Last modified on: 2006-06-01

Accessions and images





User-editable database: Community-based curation

Users can add, edit, and obsolete:

- Genes
- Alleles
- Phenotypes ('individuals')
- Images

Soon: Form framework for ontology annotation (requires ontology browser, literature associations and evidence codes)

SGN Image

[\[Cancel Edit\]](#) [\[Delete\]](#)

Image Name:

Image Description:

Uploaded by: [Naama Menda](#)

Show image size: | [thumbnail](#) | [small](#) | [medium](#) | [large](#) |



Associated tags [\[add/remove\]](#): (0)





Summary: Gene–phenotype associations

- Database of SOL genes:
 - Tomato: ~1,800
 - Potato: 1,000
 - Pepper: 600
 - Petunia: 400
 - Eggplant: 200
- SOL phenotypes:
 - 6 populations with 6,000 individual accessions.
 - 8,000 images associated to both genes and phenotypes.
- Data can be added by users and associated with a gene and/or phenotype (alleles, images, maps, sequences, literature).
- Data differs in what users have permissions to edit it:
 - curators (SGN staff), submitters and object owners (SGN verified users)
 - users (SOL community members).





Summary: Gene–phenotype associations

How?

- Coming soon: more user-editable data!
 - ontology annotations on genes
 - phenotype ontology
 - quantitative data + statistical analysis of QTLs

Why?

Solanaceae share a highly conserved genome, yet extremely diverse phenotypic variation.

The small number of different phenotypes with a known gene sequence calls for a comparative approach for revealing gene function and their associated phenotypes.





Acknowledgments

SGN is:

Senior Staff



Lukas Mueller
Director



Naama Menda
Postdoctoral Fellow

Bioinformatics Analysts



Beth Skwarecki



Robert Buels



Richard "Marty" Kreuter



Adri Mills



Chris Carpita

Interns

- ◆ Bob Albright
- ◆ Johnathon Schultz

