QTL Analysis, Visualization and the Link to Genomes at the SGN Database

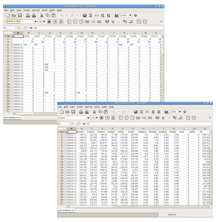
Isaak Y Tecle, Naama Menda, Robert Buels and Lukas Mueller

Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, Tower Rd, NY 14853.

One of the major challenges of the post-genomic world is linking genomic variation to phenotypic variation of complex traits. Quantitative trait loci (QTL) analysis is used to dissect the genetic basis underlying complex traits. At the Sol Genomics Network (SGN) (http://solgenomics.net), we are developing web technologies and a database to store raw phenotype and genotype data from QTL studies, perform on-the-fly QTL analysis using R/QTL statistical software (http://www.rqtl.org), and visualize QTL map locations. Users can identify peaks and flanking markers for QTLs of traits of interest. The QTL software fully integrates with other analysis tools at SGN such as the Comparative Map Viewer (http://solgenomics.net/cview/) and tomato genome browser (http://solgenomics.net/gbrowse/) and is cross-referenced with relevant datasets at SGN and external databases. For example, using the Comparative Map Viewer, users can compare predicted QTL regions to genetic maps of interest from the same or different Solanaceae species. Using the tomato genome browser, users can also identify corresponding sequences, which can yield candidate genes for a trait of interest. To allow QTL comparisons across studies, traits are mapped to the Solanaceae Phenotypes Ontology (http://solgenomics.net/tools/onto/).

Currently, QTL data from three tomato F2 population QTL studies on fruit morphology traits (up to 46 traits per population) is available at the SGN website for viewing.

QTL Data Submission



Step 1: Submit Population Details

Select Organism [Select Organism]

Population Details

Cross type: F2

Population name: QTL Mapping in Tomato Howard German x LA1589 for fruit shape

Population description: Tomato F2 population constructed from Solanum lycopersicum cv. Howard German crossed with S. pimpinellifolium accession LA1589 was analysed for fruit shape by using a software program Tomato Analyzer. The population (130 individuals) was grown in a greenhouse in the summer of 2003.

Female parent: Howard German

Male parent: LA1589

Recurrent parent:

Donor parent:

Submit

Step 2: Submit the List of Traits

Traits list

Traits list: /home/tecle/Desktop/QTL Sample data/traits_list.txt Browse... Submit

Step 3: Submit the Phenotype Data

Phenotype dataset

Phenotype dataset: /home/tecle/Desktop/QTL Sample data/phenotype_dat Browse... Submit

Step 4: Submit the Genotype Data

Genotype dataset

Genotype dataset: /home/tecle/Desktop/QTL Sample data/genotype_data Browse... Submit

Step 5: Set the Statistical Parameters for the QTL Analysis

Statistical Parameters

QTL model: Single-QTL Scan

QTL mapping method: Maximum Likelihood

QTL genotype probability method: Calculate

Genotype probability significance level: 0.05

Genome scan size (cM): 10

No. of draws:

No. of permutations: 1000

Significance level of permutation test: 0.05

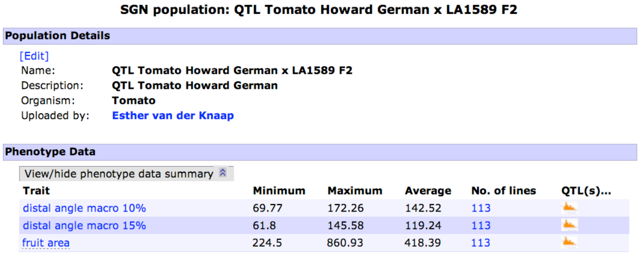
Do you want to make the data public? ● Yes ○ No

Submit Reset

Currently, SGN curators upload phenotype and genotype raw data. In the near future, researchers will be able to use the web interfaces to the right to upload their data and set their statistical parameters for the QTL analysis.

Once data is uploaded to the database, a dynamic population page with its description, list of the traits and corresponding descriptive statistics and links to the QTL page is generated. Hereby, we demonstrate the SGN QTL analysis tool and the cross-references to relevant datasets using a QTL identified for fruit area in the *Solanum lycopersicum* cv. Howard German crossed with *S. pimpinellifolium* accession LA1589.

QTL Analysis



SGN population: QTL Tomato Howard German x LA1589 F2

Population Details

[Edit]

Name: **QTL Tomato Howard German x LA1589 F2**

Description: **QTL Tomato Howard German**

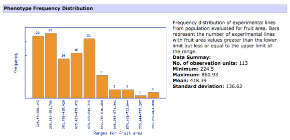
Organism: **Tomato**

Uploaded by: **Esther van der Knaap**

Phenotype Data

View/hide phenotype data summary

Trait	Minimum	Maximum	Average	No. of lines	QTL(s)...
distal angle macro 10%	69.77	172.26	142.52	113	📈
distal angle macro 15%	61.8	145.58	119.24	113	📈
fruit area	224.5	860.93	418.39	113	📈

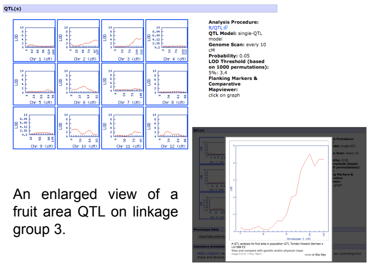


Phenotype Frequency Distribution

Frequency distribution of experimental lines from population evaluated for fruit area. Bars represent the number of experimental lines with fruit area values greater than the lower limit but less or equal to the upper limit of the range.

Data Summay:
No. of observation units: 113
Minimum: 224.5
Maximum: 860.93
Mean: 418.39
Standard deviation: 136.62

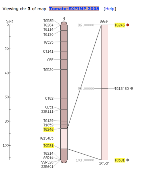
Clicking the graph icon on a trait of interest (above; eg. fruit area) prompts the on-the-fly QTL analysis and produces the phenotypic data frequency distribution and QTL maps to the left.



QTL(s)

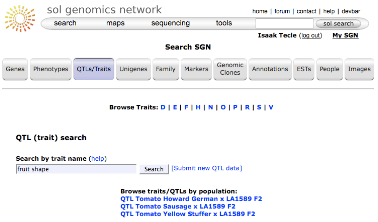
Analysis Procedure: R/QTL
QTL Model: single-QTL model
Genome Scan: every 10 cM
Probability: 0.05
LOD Threshold (based on 1000 permutations): 5%: 3.4
Flanking Markers & Comparative Mapviewer: click on graph

An enlarged view of a fruit area QTL on linkage group 3.



Linkage group 3 with the flanking markers for the fruit area QTL (highlighted).

QTL Accessibility



Browse Traits: D | E | F | H | N | O | P | R | S | V

QTL (trait) search

Search by trait name (help)

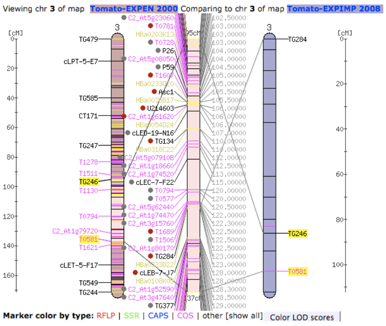
fruit shape Search [Submit new QTL data]

Browse traits/QTLs by population:
QTL Tomato Howard German x LA1589 F2
QTL Tomato Sausage x LA1589 F2
QTL Tomato Yellow Stuffer x LA1589 F2

QTLs can be accessed by: (1) browsing the alphabetically indexed list of traits; (2) searching for a trait using relevant keywords; or (3) browsing the list of QTL populations. At the populations level, all traits evaluated for QTL in the respective population are listed.

Genetic and Genomic Comparative Analysis

Comparative map viewer (http://solgenomics.net/cview)

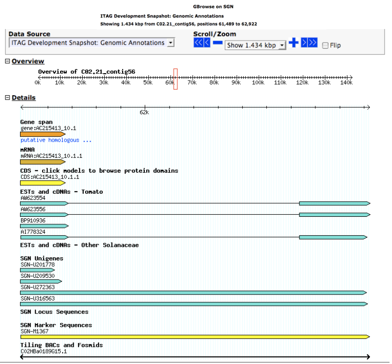


Viewing chr **3** of map **Tomato-EXPEN 2000** Comparing to chr **3** of map **Tomato-EXPIMP 2008**

Marker color by type: RFLP | SSR | CAPS | COS | other [show all] Color LOD scores

A fruit area QTL (flanked by markers TG246 and T0581) on linkage group 3 of Tomato-EXPIMP 2008 is compared to a region flanked by the same markers in the linkage group 3 of Tomato-EXPEN 2000. The corresponding region on Tomato-EXPEN shows more markers and BAC contigs.

Linking QTLs to the tomato genome sequence



GBrowse on SGN

ITAG Development Snapshot: Genomic Annotations

Showing 1.434 kbp from C02.21_contig56, positions 61,489 to 62,922

Data Source: ITAG Development Snapshot: Genomic Annotations

Scroll/Zoom: Show 1.434 kbp ☐ Flip

Overview

Overview of C02.21_contig56

Details

Gene span
gene:AC215413_10.1
putative homologous ...

mRNA
mRNA:AC215413_10.1.1

CDS - click models to browse protein domains
CDS:AC215413_10.1.1

ESTs and cDNAs - Tomato
AW623554
AW623556
BP910936
AI778324

ESTs and cDNAs - Other Solanaceae

SGN Unigenes
SGN-U201778
SGN-U209530
SGN-U272363
SGN-U316563

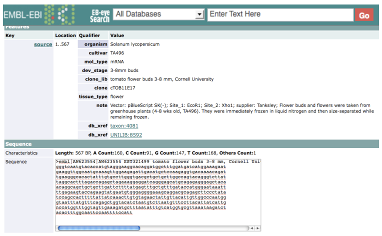
SGN Locus Sequences

SGN Marker Sequences
SGN-M1367

Tiling BACs and Fosmids
C02HBa0189G15.1

A search using the flanking markers (eg. TG246) on the tomato genome sequence (http://solgenomics.net/gbrowse/) annotated by the International Tomato Annotation Group (ITAG) reveals a matching sequence. The matching sequence may be annotated for gene, mRNA, and protein models, ESTs, cDNAs, Unigenes, and experimentally characterized genes.

Cross-references to External Databases



Key	Location	Qualifier	Value
source	1..567	organism	Solanum lycopersicum
		cultivar	TA496
		mol_type	mRNA
		dev_stage	3-8mm buds
		clone_lib	tomato flower buds 3-8 mm, Cornell University
		clone	cTOB11E17
		tissue_type	flower
		note	Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Tanksley; Flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen.
		db_xref	taxon:4081
		db_xref	UNILIB:8592

Sequence

Characteristics: **Length:** 567 BP, **A Count:**160, **C Count:**91, **G Count:**147, **T Count:**168, **Others Count:**1

Sequence:

```
>embl|AW623554|AW623554 EST321499 tomato flower buds 3-8 mm, Cornell Uni
gggtcaatgtacaccatgtagggaaggcacaggatggctttggatgatcatggaaagaat
gaaggttggcaatgcaaagttggaagagattgacatgctccaagaggtgacaaaacagat
tgaagggcacactatttgtgccttgggtgacgctgctgcttggccagtacagggtcttat
taggcactttagaccagagctagaaaggaggatcagggagcatgcagagagggagctaca
acaggcagctgctgcttgattcttttatgagtttgctgtttgataccatgggaataaatt
ttgagaagtaccagaagtatgaatgtgggaggggaaagcaggacgcagagcttccctata
tccagccacttttattatcaaacttgtgtagaactattgttacattgttggcccaatgg
gtaatttatgtttcagagctggtacatctaatgtcttaatgtttccttacattatcattg
nccatggtttggtagttgaaagatgctttaatatttgtcatggtgcgttaaataagatct
acactttggcaattccaattttccatt
```

Detailed data on sequence matches to the tomato genome region of interest can also be found on external databases. For example, in the above tomato genome region of our interest, a matching tomato EST is linked to the EMBL-EBI database (http://www.ebi.ac.uk) where details of the EST can be retrieved.

Conclusion and Future Developments

QTL identification is an important approach in understanding the genetic basis of complex traits and development of marker assisted selection for faster, more reliable and cheaper improvement of polygenic traits in plants. The SGN QTL analysis tool simplifies QTL detection, retrieval of relevant datasets from other sources and sharing data among the Solanaceae community. Through integration of user data with already existing genetic and genomic maps at SGN, users are able to do comparative analysis, and possibly identify candidate genes and more. In the near future, user-friendly web interfaces to upload QTL data and perform analysis at user-determined statistical parameters will be implemented.

References:

Brewer MT, Moyseenko JB, Monforte AJ, van der Knaap E (2007) Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. J Exp Bot. 58(6):1339-49.

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889-890.

Mueller LA, Mills AA, Skwarecki B, Buels RM, Menda N, Tanksley SD (2008) The SGN comparative map viewer. Bioinformatics. 24(3):422-3.

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond. Plant Physiol 138(3):1310-7.

Funding Sources

http://solgenomics.net
Contact SGN at:
sgn-feedback@solgenomics.net

SGN Code and Datasets
ftp://ftp.solgenomics.net