SOL Newsletter

Issue No. 6: September, 2005

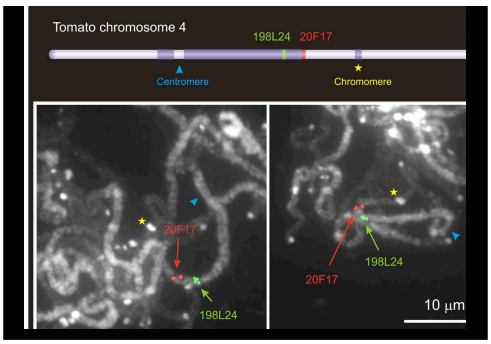
Tomato Sequencing Updates

Chromosomes 1, 10, 11 (US)

Sequencing of 12 BACs is in progress as part of the US effort. It is estimated that six of the BACs wll be completed sometime in early September. Full tomato BAC sequences from 40 BACs are available at SGN with preliminary annotations. These sequences were contributed from all the tomato genome sequencing project participants. For more details, see the "What's New on SGN?" section of this newsletter.

Since our last report, we have identified sites of hybridization for four additional BACs by FISH experiments using tomato SC spreads. Previous experiments had shown that two of these BACs (216M19 and 251G05) contained repeated sequences that resulted in FISH labeling at numerous locations on most or all of the chromosomes. In the absence of CISS (chromosomal in situ suppression) hybridization, 216M19 labeled multiple sites near the telomeres of many of the chromosomes, whereas 251G05 showed heavy labeling of the pericentromeric heterochromatin of all of the chromosomes. With CISS hybridization using Cot100 tomato genomic DNA, labeling with 216M19 was restricted to a single site near the telomere at the end of the short arm of chromosome 8. Similarly, in the presence of Cot100 DNA, 251G05 hybridized only to a site near the euchromatin-heterochromatin border on the short arm of chromosome 6. CISS hybridization was not required for localization of two additional BACs. One of these, BAC19, hybridizes to a single site on the long arm of chromosome 2. The other, 250I21, produced a unique locus of labeling near the euchromatin-heterochromatin border on the short arm of chromosome 6, at a location slightly nearer to the centromere than 251G05. Finally, at the request of the Mapping Core Group at the Wellcome Trust Sanger Institute in the UK, we investigated the location of BAC 198L24, which was believed to be located on the long arm of chromosome 4. Using FISH with tomato chromosome squash preparations, we confirmed that this BAC localizes to a site within the heterochromatin on the long arm of chromosome 4, as shown by the green signals in Figure 1. BAC 20F17 (red signals) is a marker for chromosome 4.

Figure 1: FISH image of BAC 198L24 on the long arm of chr 4 (green signals). BAC 20F17 (red signals) is a marker for chr 4.



Chromosome 2 (Korea)

Since our last update, we completed the sequencing of 12 additional BAC clones, and one BAC clone is in progress. To date, we have finished sequencing a total of 31 seed BACs and the sequence information was deposited into the SGN database. Currently, BAC extension is being applied.

Chromosome 3 (China)

A total of 20 BACs (18 euchromatic BACs close to telomeres and euchromatin-heterochromatin borders, and 2 BACs from the pericentric heterochromatin) have been selected based on overgo hybridization data from SGN and FISH to represent various regions of chromosome 3. Five completed BACs are being annotated, and 15 have sequences in stage 1 and 2.

Future plan:

To facilitate the BAC-by-BAC sequencing, we plan to construct a complete physical map. The minimum tilling path will be developed based on available BAC-fingerprinting, overgo hybridization, and tomato genomic sequences (BAC ends, unmethylated sequences, BAC sequences, etc.), as well as manual editing of FPC results, more BAC anchoring and mapping data being developed in China. First, we did extensive manual editing of the FPC (Fingerprinted Contigs) physical map, and reduced the contig number from 6,794 to 3,000. The current FPC contains 88,650 BAC clones from the HindIII library, representing a 15X coverage of the tomato genome which is about 788 Mb of physical regions. After integrating the current overgo hybridization data into the FPC contigs, 443 contigs were anchored to the euchromatic regions in the F2.2000 genetic map, which covers about 192 Mb of the tomato genome. The WebFPC can be viewed at http://159.226.24.153/webfpc/WebAGCoL/Demo/index.html. Second, we will integrate the genetic map of F2.2000 and the fingerprinting contigs using three approaches: 1) associating BAC clones with genetic markers without overgo results; 2) genetic mapping the unanchored contigs; and 3) FISH. A total of 950 genetic markers without overgo hybridization data were selected for PCR screening of BACs containing the markers. The remaining unanchored physical contigs will be mapped to F2.2000 using the BAC end sequences. In addition, we will provide a website (http://159.226.24.153/webfpc/WebAGCoL/Demo/index.html) to display the physical map information, periodically make simulated digests of the sequenced BAC clones available and put them back to the FPC. The purpose is to ensure the integrity of the physical map, make additional anchoring, and provide a framework for sequencing.

Chromosome 4 (UK)

As reported in the last update, we have conducted *in silico* analysis of the FPC contigs that have potential assignment to chromosome 4 on the basis of the overgo probe analysis. All possible electronic merges at this stage have now been made and the number of contigs potentially located on chromosome 4 remains at 58. Therefore, we are currently looking into the possibilities and practicalities of augmenting the FPC database with further fingerprints from the MboI library. This would increase the existing genomic coverage represented by the FPC database. We feel the increase in fingerprint data would have a positive, valuable impact to assist mapping and sequencing across all tomato chromosomes. Our mapping strategy continues to aim at reduction of the contig number and generation of larger contigs across which minimal tilepaths may be selected and sequenced. Currently, we are conducting tests of the fingerprinting technique on several MboI library plates and also some existing fingerprints of the LE_HBa library.

Presently, we have two clones (31H05 and 198L24) from the LE_HBa library in Shotgun sequencing. These clones are predicted to overlap according to the FPC map, which formed the basis of their selection for sequencing. The clones underwent PCR verification prior to sequencing after single colonies were isolated from our library copy. An additional three clones were selected to be sequenced and further verification is in progress.

Chromosome 5 (India)

Members of the Indian Initiative on Tomato Genome Sequencing have confirmed 6 BAC clones from tomato chromosome 5 with the help of markers (T1252, C2-At1g60200, cLET-8-B23, T0876, cLED-8-G3, BS4) by sequencing with marker-specific custom primers, end sequencing, and fingerprinting. Shotgun libraries have been made and high throughput sequencing has started.

Chromosome 6 (The Netherlands)

To date, 16 BACs have been completed to Phase 1, and 2 BACs have been fully closed. An additional 17 BACs are currently in the sequencing pipeline and 11 BACs are in the FISH pipeline. Of the completed BACs, 10 have been uploaded to SGN for public release.

The results are very encouraging from the BAC-end approach to radiate out of the seed BACs. Using the latest dataset from SGN (approx. 220.000 BAC ends both from the HindIII and MboI libraries) multiple flanking BACs were identified for 15 out of 18 seed BACs. Currently, these flanking BACs are being fingerprinted to identify the minimal overlapping ones. Sequencing two such BACs confirmed their predicted minimal overlap with the corresponding seed BACs as determined by AFLP fingerprinting. Using these two BACs in a second round of BAC-end screening, yielded a number of flanking BACs so that a second step away form the seed-BACs probably has been made.

This shows that the combination of BAC-end screening and AFLP mapping forms a powerfull tool to quickly radiate out of the seed BACs. The EcoRI library is also now in use in the chromosome 6 project, and yielding extended possibilities to identify flanking BACs with minimal overlaps.

Chromosome 7 (France)

Update pending.

Chromosome 8 (Japan)

A total of 67 BAC clones corresponding to 33 DNA markers on chromosome 8 have been shipped up to present. Sequencing of five BAC clones associated with five markers (CT64, CT68, CT148, T1123, and TG176) has been completed. Six clones are in the production phase, and nine are in the shotgun phase. Processing of the clones for the remaining 13 markers are suspended for several reasons: no PCR amplification with the marker-primers, significant disagreement of sequences between the original markers and the PCR products, etc.

We have just received all three BAC libraries from the Boyce Thompson Institute for Plant Research (Dr. Joyce Van Eck). We are hoping these libraries, combined with the EST information, will help efficient selection of the clones to analyze and accelerate sequencing.

Chromosome 9 (Spain)

Update pending.

Chromosome 12 (Italy)

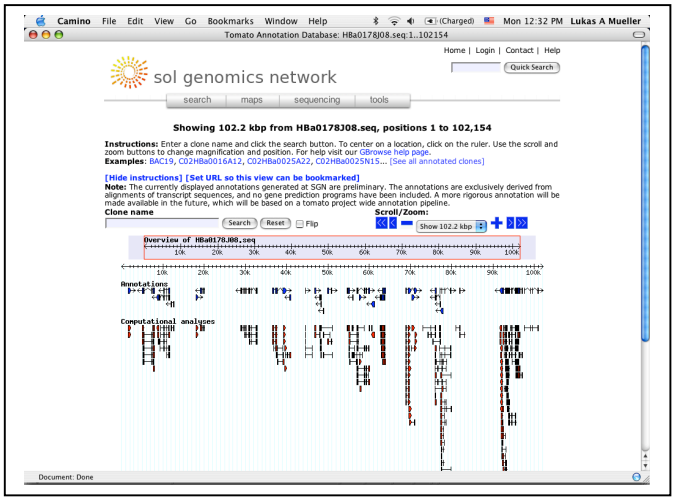
To date, we have received the set of seed BAC clones anchored to 50 markers on chromosome 12. As potential starting points for sequencing, we chose 20 seed BACs identified by 16 overgo markers. The first two selected BACs have been assembled with few gaps and additional sequencing is underway in order to achieve the required quality standards. The identity of five other seed BACs was confirmed by PCR and IL mapping, and these BACs are now in the sequencing pipeline. Verification of 13 additional seed BACs is in progress. However, the BAC-end sequences of a few clones show some mismatches with the BAC-end sequence data posted on the SGN website.

Completed BAC Sequences

Sequences from 40 BACs contributed by all project participants have been completed and added to the SGN website. The sequences can be downloaded from the FTP site and the annotation of most of the sequences can be viewed on the SGN website (Figure 1). A BLAST dataset of the complete BAC sequences is available in the SGN BLAST tool (found under the "tools" menu).

Links:
SGN FTP site: ftp://ftp.sgn.cornell.edu/tomato_genome/bacs/
SGN BAC sequence browser: http://sgn.cornell.edu/cgi-bin/gbrowse/
SGN BLAST: http://sgn.cornell.edu/cgi-bin/tools/blast/simple.pl
SGN BAC Search: http://sgn.cornell.edu/cgi-bin/search/direct_search.pl?search=bacs

Figure 1: The SGN BAC annotation



BAC Ends

Currently, there are 354,029 BAC end reads from 3 libraries on the SGN website. 301,602 reads yielded high quality sequence (85.2%), and 274,117 sequences contained high quality, contamination-screened inserts. The BAC ends can be downloaded in bulk from the FTP site, and can be queried on the SGN BAC search. A BLAST dataset is also available on the SGN blast page (http://sgn.cornell.edu/cgi-bin/tools/blast/simple.pl).

Read statistics broken down by library:
152,819 HindIII BAC Library
101,755 MboI BAC Library
99,455 EcoRI BAC Library
354,029 Total

Links:
FTP download of all BAC end sequences:
ftp://ftp.sgn.cornell.edu/tomato_genome/bac_ends/seq_dir/

FISH Map

A FISH map has been added to SGN. It displays Fluorescence In-Situ Hybridization (FISH) experiments. The map can be viewed on SGN by selecting "Tomato FISH map" from the maps menu (Figure 2). The map shows an idiogrammic representation of the chromosomes. The actual micrographs from the experiments can be viewed through links on the map. Currently, the database contains 6 experiments. The FISH map can be compared to the genetic map in the comparative viewer.

Figure 2: The FISH map, displaying FISHed BACs on pachytene chromosome glyphs. The physical position can be compared to the genetic map position in the SGN comparative viewer.

The BAC end data can be downloaded in batch from our FTP site and be queried from the SGN database (go to the "search" menu and choose "BACs"). The sequences have also been submitted to Genbank.

Improved Unigene Annotations

-Predicted protein and cds sequences

Predicted protein and cds sequences have been added to the unigene detail pages. ESTScan with a tomato specific matrix was used for the predictions. The data can be viewed on the web and downloaded from the FTP site. To see an example, go to: http://sgn.cornell.edu/cgi-bin/search/unigene.pl?unigene_id=222226

-GO annotations and interpro domains

Using the predicted protein sequences as a basis, Interpro domains were identified in the SGN unigene sequences using software from the Interpro consortium. The annotations appear on the unigene detail pages and can be downloaded from the FTP site. To see an example, point your browser to http://sgn.cornell.edu/cgi-bin/search/unigene.pl?unigene_id=222226.

-Gene Family annotations

The predicted protein sequences derived from the unigene sets and the Arabidopsis proteome were clustered using tribeMCL. The results can be viewed on the unigene detail pages. To see an example, go to: http://sgn.cornell.edu/cgi-

New Potato Build

A new potato unigene build is available on the SGN website. The build incorporates almost 50,000 new sequences obtained from the Canadian Potato Genome project, and also integrates transcripts sequences deposited in Genbank. We would like to thank Prof Barry Flinn and his associates for submitting the data.

User Comments and SOL Forum

The ability to add user comments has been added to certain pages, such as detail pages for markers and maps. To make a comment about, for example, a marker, go to the corresponding marker detail page using the marker search and click on 'Add comment...'. You need to be logged in to make comments, and the comment will be automatically tagged with your name, date, and time of submission. Comments are then displayed at the bottom of the screen.

SOL Forum

The SOL Forum allows you to enter a discussion topic and post comments in the topic sections (http://sgn.cornell.edu/forum/). A 'Job Posting' section has already been created (http://sgn.cornell.edu/cgi-bin/solpeople/posts.pl?topic_id=2) that allows registered users to add job announcements.

Recent Solanaceae Publications

Contributed by Chris Bowler

Davuluri, G.R., van Tuinen, A., Fraser, P. D., Manfredonia, A., Newman, R., Burgess, D., Brummell, D.A., King, S.R., Palys, J., Uhlig, J., Bramley, P.M., Pennings, H.M.J., and Bowler, C. 2005. Fruit-specific RNAi-mediated suppression of *DET1* enhances carotenoid and flavonoid content in tomatoes. Nature Biotechnology, 23:890-895.

Dixon, R.A. 2005. A two-for-one in tomato nutritional enhancement. Nature Biotechnology, 23:825-826.

Conference and Workshop Announcements

Just a few more weeks until the 2nd Solanaceae Genome Workshop!

September 25 – 29, 2005
Hotel Continental Terme – Ischia – Italy
http://www.solanaceae2005.org

Plant and Animal Genome XIV Conference
January 14 – 18, 2006
Town and Country Convention Center
San Diego, California
http://www.intl-pag.org

A meeting will be scheduled during the conference for all tomato genome sequencing project participants. Details will be posted in a future newsletter and on SGN.

Solanaceae Recipes

Recipe from www.marthastewart.com

Harvest Vegetable Tian

Serves 4 to 6

A tian is a French dish whose name—like that of the casserole and the terrine—refers to the food itself as well as the vessel in which it is cooked.

- 4 to 5 tablespoons extra-virgin olive oil
- 2 medium baking potatoes, peeled and thinly sliced
- 1 small Japanese eggplant, trimmed and thinly sliced on the diagonal
- 1 small zucchini, trimmed and thinly sliced on the diagonal
- 4 plum tomatoes, thinly sliced
- Coarse salt and freshly ground pepper
- 4 sprigs fresh rosemary
- 1 head garlic
- Crusty bread, sliced 1/2 inch thick and toasted, for serving (optional)

1. Preheat the oven to 400° with a rack in the center. Coat the bottom of a 9-by-13-inch baking dish with 1 tablespoon oil.

2. Arrange the vegetables in the baking dish in slightly overlapping rows, alternating potato, eggplant, zucchini, and tomato. Season with salt and pepper and top with rosemary sprigs. Place garlic in center of dish. Drizzle all over with remaining 3 to 4 tablespoons oil.

3. Bake, swirling baking dish every 10 minutes to distribute oil, until vegetables are tender and garlic is soft, 40 to 50 minutes. This will keep the vegetable moist and tender.

4. Remove baking dish from oven. Remove garlic and let stand until cool enough to handle. Squeeze garlic pulp into a small bowl and mash with a fork. Spread over vegetables, or serve on toast, if desired.