



In this Issue

In this issue you will find brief updates from participants in the tomato genome sequencing effort, eggplant related news in the Community News section, the latest on SGN and bioinformatics, plus information for the 2nd Solanaceae Genome Workshop.

Tomato Sequencing Updates



Chromosomes 1, 10, 11 (US)

The sheared library has been completed with a total of 4800 sequences. The EcoRI library will be finished in early April making it the third library available to the community. 75,000 BAC sequences are now available from SGN (see the "What's New on SGN?" section of this newsletter for more information). The overgo project has been completed and all marker-BAC associations and short lists of candidate seed BACs for each chromosome have been placed on the ftp site on SGN. Telomeric and centromeric BACs were sent to Steve Stack for FISH analysis.

Chromosome 2 (Korea)

Five BACs have been completed and the group is currently doing the annotation. Draft sequences from one BAC with 16 gaps were generated. Ten libraries have been made and are ready to be sequenced. Finally, a total of 30 BACs were confirmed by FISH.

Chromosome 3 (China)

As a first step towards sequencing chromosome 3, we are focusing on constructing a fine physical map. The first phase of our effort is to integrate all available resources, including the FPC physical map, overgo hybridization data, BAC end sequence data, and FISH data. We FISHed 47 BACs onto chromosome 3, with each BAC connected to one genetic marker.

Based on these FISHed BACs, we identified 36 contigs anchored to the chromosome based on FPC. By merging contigs with FPC, our 36 contigs extend to about 8 MB. The length of each contig ranges between 100 kb to 1.4 Mb. Almost all contigs are distributed on paracentromeric regions and the long arm of chromosome 3 including the so-called "heterochromatic" regions. No contig was identified on the short arm because no overgo hybridization data is available for genetic markers located on this arm.

The next phase of our effort will be to FISH additional BACs to confirm the identity of merged contigs, to do more overgo hybridization to anchor additional contigs on the short arm, and to fill gaps between contigs. Therefore, the availability of the BAC library or filters is critical to our effort in developing a fine physical map.

Chromosome 4 (UK)

All three libraries (HindIII, Mbol, EcoRI) have been requested and collaborators at the Wellcome Trust Sanger Institute are reviewing various databases to select appropriate BACs.



Chromosome 5 (India)

The Department of Biotechnology, Government of India, has approved an Indian Initiative on Tomato Genome Sequencing with a view to sequence chromosome 5 as part of the international effort. The work will be carried out at three institutions in New Delhi, namely Department of Plant Molecular Biology, University of Delhi South Campus; National Research Centre for Plant Biotechnology, Indian Agricultural Research Institute; and National Centre for Plant Genome Research. Approval for other programmes in the area of functional genomics is in the pipeline.

Chromosome 6 (The Netherlands)

To date, a total of 48 BACs has been mapped on chromosome 6 using AFLP markers and also some available RFLP markers. After FISH analyses, a total of 25 BACs was found suitable for sequencing. The remaining BACs were either derived from heterochromatic regions on chromosome 6, or were derived from different chromosomes. Of the 25 selected BACs, 14 BACs have been completed to Fase 1 and 1 BAC has been completed to Fase III. A further 9 BACs are currently in the sequencing pipeline.

Preliminary results have been obtained for screening of the BAC-end library, which at present holds 40,000 BAC-ends from the HindIII library. When 6 finished tomato BACs were used for screening the BAC-end library by BLASTN, 5 out of the 6 BACs gave multiple hits with various BAC ends. Currently, the overlap of the corresponding BACs with the 6 selected tomato BACs is confirmed by AFLP fingerprinting.

Chromosome 7 (France)

Funding has been awarded and the project is under development.



Chromosome 8 (Japan)

We have received 42 BAC clones corresponding to 8 DNA markers on chromosome 8. Most of the clones generated PCR products of appropriate sizes with the primers designed based on the marker sequences. However, we had difficulty in confirming integrity of these clones mainly due to limited amount of available data on end sequences and fingerprints at the moment. The nucleotide sequences of three BAC clones associated with three markers, CT64, CT68 and CT148, of which the integrity was confirmed, have been determined, whereas sequencing of the clones for the remaining five markers was suspended. The obtained sequences were subjected to gene modeling and annotation using our automatic pipelines, and a total of 56 protein-encoding genes were assigned.

Fingerprinting analysis of the BAC clones for the remaining five markers indicated that the majority of the clones originated from the respective single locus, and four additional clones were newly selected for sequencing. Five candidate clones for the marker SSR63 showed diverse fingerprinting patterns suggesting that these clones represent multiple loci.

Chromosome 9 (Spain)

Awaiting a decision on funding.

Chromosome 12 (Italy)

To date, we have received 250 seed BACs anchored to 30 markers on chromosome 12. We selected five seed BACs located on the short arm. The map position was confirmed for three of these seed BACs by PCR-based mapping, and they are currently in the sequencing pipeline at the CRIBI sequencing service. For the remaining two seed BACs, PCR-based mapping confirmation is in progress. As for the long arm, the selection and verification of two additional seed BACs has been started. The physical localization by FISH of the first seed BACs will be done in Dr. Stack's lab. FISH facilities are being set up at the University of Naples and at CNR-IGV of Naples after a training period in Hans de Jong's laboratory.

Community News



Eggplant

Contributed by: Anne Frary and Sami Doganlar

Here at the Izmir Institute of Technology we propose to develop eggplant genomics so that researchers have the knowledge and resources not only to improve eggplant but also to address broad questions about evolution and adaptation in the Solanaceae. To this end, we are developing a higher density genetic map by the addition of AFLP, COSII and microsatellite markers to the existing RFLP map. In addition, we plan to develop maps for the lesser-known cultivated forms of eggplant, *S. macrocarpon* and *S. aethiopicum*. These species along with *S. melongena* and its putative progenitor, *S. incanum*, are being used to produce advanced backcross QTL populations. In parallel with SOL, we are also seeking funding for partial sequencing of the eggplant genome. It is hoped that the information generated by this project will allow a greater understanding of evolution and adaptation in the family and also allow better application of tomato resources to eggplant.

If you have news from your community that you would like to contribute to the next SOL newsletter, send it to Joyce Van Eck at jv27@cornell.edu before April 15.



SGN

What's New on SGN?

- (1). **75,000 BAC sequences are now available from SGN.** The sequences can be searched using the SGN search function (<http://sgn.cornell.edu/>, click on the Search menu and choose BACs), and downloaded as a total dataset from the ftp site at ftp.sgn.cornell.edu/tomato_genome/bac_ends/
- (2). **Overgo project:** all marker-BAC associations and short lists of candidate seed BACs for each chromosome have been put on the ftp site. To access: Select "Download" from the SGN home page, go to the ftp site, select the tomato genome ftp level, select the seed BACs folder, then click on the chromosome of interest to download the information.

Bioinformatics



Summary from the meeting held during the International Plant and Animal Genome XIII Conference

Contributed by: Lukas Mueller

The Tomato Genome Sequencing project met at the Plant and Animal Genome conference in San Diego, CA, on January 15 to discuss project progress and future directions. All groups were represented except China, however, Eileen Wang from Cornell gave their update. There was a presentation for each chromosome (PowerPoint files at: <http://sgn.cornell.edu/solanaceae-project/sol-bioinformatics/>), a presentation about seed BAC selection and quality control, and a discussion on bioinformatics issues. It was exciting to see the sequencing progress that has already been made for several chromosomes.

In the meantime, SGN has implemented a BAC registry, in which project members can update the status of their BACs by logging into SGN and using an interactive webpage. The sequencing status overview page is directly generated from that data (http://sgn.cornell.edu/help/about/tomato_sequencing.html). Upload accounts for uploading sequence, chromatogram, assembly, etc. for each BAC were also generated.

In the bioinformatics discussion, the need for more stringent nomenclature was raised, which will be discussed in more detail on the mailing lists and then added to the SOL guidelines. The Genbank submissions were discussed; a consortium is currently being defined with Genbank that includes all sequencing and annotation centers to allow everybody to update annotations on the BACs; the sequencing centers will be responsible for the actual submissions. A distributed annotation pipeline was outlined. The idea is to be able to write annotation pipelines that can tie into different tools available on the web using MOBY or other means, and that return their results in the common format GAME XML.

Additional information is available at <http://sgn.cornell.edu/solanaceae-project/sol-bioinformatics/>.

Upcoming Meetings

[2nd Solanaceae Genome Workshop](#)
September 25 - 29, 2005
Hotel Continental Terme - Ischia - Italy

The website for information related to the meeting is
<http://www.solanaceae2005.org>.

