Notes from Tomato Genome Sequencing Meeting at PAG 2007

*These notes represent the highlights from the meeting. Additional notes were posted by Rob Buels at http://sgn.cornell.edu/documents/solanaceae-project/sol-bioinformatics/pag2007_docs/notes_rb.txt. More detail can be found in the presentation files that are posted on SGN in a folder labeled " tomato seq meeting_PAG_07 ".

A meeting was held at the Plant and Animal Genome Conference on Sunday, January 14, 2007. The agenda for the meeting can be found on the last page of this document. Presentations from the meeting can be downloaded from a link on SGN.

Project participants in attendance represented efforts by the following countries: US, Korea, UK, India, Netherlands, Spain, and Italy. Representatives from each of those groups gave brief updates on progress because more detailed updates were provided in the January issue of the SOL Newsletter prior to the conference. Additional update information can be found in the presentation files from each of the groups.

Rene Klein Lankhorst gave an overview of the EU-SOL sequencing activities. They have $34.5M in funding for 5 years. There are three work packages, and efforts include tomato genome sequencing, mitochondrial genome sequencing by Argentina, and de bottlenecking of European tomato projects, which is being managed by Heiko Schoof. René's slides can be found in the file labeled Presentatie Tomato sequencing meeting PAG 2007.

Jim Giovannoni led a discussion on tools available for BAC extension and completion criteria that were developed for the National Science Foundation (NSF) based on feedback from proposals that were submitted by the US team to the Plant Genome Research Program. The new tools for BAC extension include items such as 2M sequence reads provided by Japan and a BAC overgo screen of the MboI library that will be done by the US group. As for completion criteria, Jim reviewed the approach used to estimate the amount of euchromatin in the tomato genome and the gene space that will be missed in this approach. It is anticipated that 85% of the total tomato gene space will be recovered when the sequencing is completed. He gave three points used to define when a BAC is considered finished and criteria that will be followed to determine when sequencing of the euchromatin is completed. Jim's slides are in a file labeled ToolKit_Criteria07.

An announcement of additional resources was made by representatives from Syngenta and Keygene. Erik Legg from Syngenta informed the group that they have mapped approximately 26% of the tomato genome in a total of 763 anchored BAC contigs. The bins are based on the tomato EXPEN 2000 map. Molly Dunn from Syngenta will work with Lukas Mueller to make this data available as soon as possible on SGN. Mark van Haaren from Keygene announced that they would make available markers to find new seed BACs that are based on AFLP markers. They plan to start with chromosome 6 as a pilot and then make them available to the sequencing group. There are no financial conditions for these resources from Syngenta and Keygene. Erik's slides can be found in the file labeled Tom_FPC_Map_PAG. As for Mark's slides, his file could not be transferred due to a problem with his USB flash drive.

There was a brief discussion about possible strategies to sequence the tomato heterochromatin. These included a shotgun approach, BAC-by-BAC along with 454 technology, but it was recognized that a necessary prerequisite for this work would be a physical map. Any sequencing of the heterochromatin would also require the groups to seek additional funding.

Rob Buels from SGN spoke about recent efforts at SGN and the International Tomato Annotation Group (ITAG) and the plan they have put in place. During Rob's presentation, the UK group brought up points related to Tiling Path Format (TPF) and AGP files. It was suggested that an additional column be added onto the TPF page for marker information. As for AGP files, it was noted that this is only for Phase 3 clones. It was also discussed that the groups in sequencing need to remember to upload their TPF and AGP files onto SGN. Rob's slides are in the file labeled pag2007.

Nevin Young from the University of Minnesota gave a presentation about the Medicago genome sequencing project. He is the lead primary investigator on this project. He pointed out that Medicago serves only as a model and is not a crop. They are sequencing the euchromatin (~300 Mbp) in a BAC-by-BAC approach. To date, 60% of the BACs are sequenced, and they have 407 gaps of which they predict that 2/3 will be easy to fill and for 1/3 of the gaps there is not an obvious way to fill the gaps. They are seeing about 20 genes/100 kb in the euchromatin. Chromosome 6 is very different from the other Medicago chromosomes. Nevin's slides are in a file labeled ndy_present_pag07.

Giovanni Giuliano and Matt Lorenz from Affymetrix spoke about tomato microarray platforms and the proposal for the new tomato Affy chip, which includes 2.6M probes representing more than 35,000 transcripts. They asked that there be a grace period of publishing data before they make a public announcement. The time frame for a publicly available tomato microarray will be September 2007. There will be three different types of expression strategies. The current array contains 222,000 probes and is 11 um. The planned new array will have 30 or more probes per transcript and will be 5 um. As for ordering, they want a minimum of 540 microarrays ordered. This order does not have to all come from one institution, different institutions can join together but again the minimum ordered has to be 540. A meeting following the tomato sequencing meeting was held for everyone interested in the Affy chip. Giovanni's slides are in a file labeled Giovanni Affy chip and Matt's are labeled AdvantagesForGeneLevelTomatoArray.

Christian Bachem gave an update on the Potato Genome Sequencing Consortium. Currently, there are fourteen participating labs from around the world and their goal is to complete the sequencing of the potato genome by 2010. The 840 Mb of the potato genome is being sequenced in a BAC-by-BAC strategy, with an average BAC length of 130 kb. There is one BAC library and it has been sent to all the sequencing partners. Christian's slides can be found in the file labeled PGSC Jan 2007.

Agenda for Tomato Sequencing Meeting at PAG 2007

January 14, 2007
10:00 am – 5:00 pm

10:00 - Progress updates – ~ 8 min per group

11:00 – Overview of sequencing activities within EU-SOL – René Klein Lankhorst

11:30 – Discussion on expanding the tool kit for BAC extension – Jim Giovannoni

12:00 – Lunch

1:15 - New tools and procedures – Rob Buels
  -BAC sequencing
  -BAC distributed annotation
  -Coordination of efforts to map BACs to the IL material

1:45 - Discussion

2:15 - A *Medicago truncatula* sequencing update – Nevin Young

2:40 – Coffee break

3:00 - Tomato Affy chip – Giovanni Giuliano and Matthew Lorence

3:30 – The Potato Genome Initiative – Christian Bachem

4:00 – 5:00 Wrap up and discussion