

Meeting Summary

Topic: *An International Workshop to Discuss Sequencing of the Tomato Genome: Feasibility, Benefits and Strategy*

Location: Washington, D.C. (Holiday Inn Dulles International)

Date: November 3, 2003

Objectives

The purpose of this gathering was to determine the feasibility, utility, strategy and level of international interest/commitment for sequencing the tomato genome as a reference for the family Solanaceae and other closely related plant families. The workshop brought together an international group of scientists to discuss: 1) the current status of Solanaceae research – including aspects of plant biology for which Solanaceous species are a preferred model, 2) the impact of sequencing the tomato genome on research in the Solanaceae and plant biology in general. 3) sequencing strategy, 4) mechanisms by which such a sequencing project can be conducted as part of a multinational consortium, and 5) strategies for sequence information management, curation and public dissemination.

Summary of Events

Broad international interest in tomato genome sequencing as a central component to a larger Solanaceae Genome Initiative. Sixty-nine researchers representing academic and government research labs, industry and extension/outreach specialists with interest in the Solanaceae met for a full day to discuss how the tomato genome could be sequenced as a reference for the broader Solanaceae. Representatives of eleven countries summarized national interest and prospects for state-sponsored sequencing support within their respective countries. The Netherlands has already funded sequencing of 20 Mb of the R-gene rich region of tomato chromosome 6 (apx. 8% of the estimated 250 Mb tomato gene space) and the US is nearing completion of an anchored tomato physical map to facilitate targeted sequencing of the gene space. Representatives of France and the UK reported on discussions with administrators of their respective national funding agencies and indicated proposals to sequence one or two tomato chromosomes were encouraged. Similar, though less specific positions were relayed by representatives of Spain, Italy, Canada and S. Korea and all agreed that funding would depend on the initiation of efforts in other countries to insure that the outcome of their investments would be ultimately realized. There was general agreement among those working on other members of the Solanaceae (pepper, potato, eggplant, tobacco) that tomato presented the appropriate reference genome for the Solanaceae. Ongoing and proposed activities in these species including development of comparative maps, EST and low density shotgun sequencing which would facilitate connection of the various Solanaceae genomes to the tomato reference.

Developing the Solanaceae as a model for systems biology in plant research. Systems biology takes advantage of diverse and integrated knowledge to yield novel biological insights which could not be established via more traditional narrow

analyses. Systems biology is an outcome of the "omics" era and the Solanaceae are positioned to be a model for Systems biology. Following the morning breakout discussion session, summaries of the state of comparative Solanaceae genomics, host-pathogen interactions, fruit and tuber development, ecological genetics, quantitative traits, systematics, germplasm resources, metabolism, and proteomics were presented. A major outcome of these discussions was that there are clear and optimal opportunities for system analysis in the Solanaceae. Comparative maps link the major members of this family to tomato and each other and extensive marker resources are publicly available. Diverse germplasm including large collections of mutations, wild species, and multiple fixed and ordered introgression populations provide unsurpassed allelic variation for simple and quantitative trait analysis. In particular, the use of fixed homozygous introgression collections developed in tomato has facilitated extensive analysis of diverse metabolites and QTLs which can now be linked to mapped gene loci, allelic variation, gene expression and protein accumulation. The repertoire of molecular and germplasm resources available in the Solanaceae make this family an ideal model for systems biology. In addition and as pointed out by several of the discussion presenters, the ease of tomato transformation and its relatively short life cycle (apx. 12 weeks seed to seed) has made tomato an increasingly popular model crop for testing hypotheses developed in model systems such as Arabidopsis, further extending the biological insights that can be gained through the Solanaceae.

Strategy for tomato genome sequencing.

There was general agreement to sequence the euchromatic gene space of tomato via a targeted minimal tiling approach in which BACs would be anchored to the genetic map and those from the gene space would serve as initiation points for genome sequencing. BAC fingerprinting and end-sequencing would provide the information needed to make additional steps down the tiling path. This approach would also facilitate the division of labor among participating groups/countries based on whole or partial chromosomes thus providing clear references for group and national contributions. There was strong and broad agreement to development of a single virtual entry point for Solanaceae genome sequence and related genomics and systems information to facilitate utility by the user community. A plan for implementation of a mirrored informatics interface that could be developed in multiple participating countries but with a single external face was also viewed positively by the group as the objective of such a plan would be to maximize input and participation while delivering a user friendly and comprehensive product.

Followup and Next Steps

Given the strong international interest in sequencing the tomato genome and developing the Solanaceae as a model for systems biology, several specific items were agreed upon as next steps.

- 1) Formation of Steering Committee for the International Solanaceae Genome Project. (ISGP) It was decided that the committee will be comprised of representative scientists from each of the countries actively participating in the ISGP. Marc Zabeau (Belgium) and Dani Zamir (Israel) have agreed to serve as initial co-chairs of this committee and will work with the international community to establish membership of this committee no later than Dec 31,

2003. This committee will be charged with facilitating and helping coordinate research and funding for the various components of the ISGP. They will also be charged with drafting a “white paper” for the ISGP (see below). The committee will meet annually at the Solanaceae Genomics meeting which will be held annually in one of the participating countries (also see below).

- 2) Drafting of a “white paper” for the Solanaceae Genome Initiative. A white paper describing the goals, components and (where possible) time tables for the ISGP will be drafted and distributed to the participants of the meeting for comments and suggestions no later than December 1, 2003. The goal is to have a final draft of this white paper no later than December 31, 2003 so that various participating countries can use the document to help with planning and funding of the components in which they will be involved.
- 3) Drafting of a “technical document” for sequencing of the tomato genome. A central component of the ISGP is sequencing of the tomato genome. Based on the November 3, 2003 meeting in Washington DC, a general strategy was decided (see earlier section) which would accomplish a fully contiged sequence of the euchromatic portion of the genome. The sequencing project would be conducted in a manner that would allow multi-country participation. To facilitate this central component of the ISGP, a “technical document” describing the specifics of the sequencing project will be drafted and distributed to the participants of the meeting no later than December 1, 2003. The goal is to have a final “technical document” for the sequencing portion of the ISGP completed no later than December 31, 2003.
- 4) Formation of an International Bioinformatics Steering Committee to oversee bioinformatics for the ISGP and especially sequencing of the tomato genome. Lukas Mueller (curator for SGN – Solanaceae Genome Network database/website) has agreed to organize and serve as the chair of this committee. This committee will be charged with drafting and overseeing both the bioinformatics components of the ISGP as well as the “technical document” for sequencing the tomato genome. This committee will work hand-in-hand with the ISGP Steering Committee. These bioinformatics components will be contained in the drafts of both the “white paper” for the ISGP and the “technical document” for the tomato sequencing project.
- 5) Establishment of an annual International Solanaceae Genomics Conference. It was agreed that an annual conference will be held devoted to the topic of genomics research in the Solanaceae. The first meeting will be held in July 2004 and hosted by Holland. The meeting times and locations for the 2005 and 2006 meetings will be decided within the next 6 months. It was suggested that the 2006 meetings might be held at the University of Wisconsin as part of the Solanaceae systematics meetings. Both the ISGP Steering Committee and the Bioinformatics Steering Committee will hold discussion sections at these gatherings.
- 6) Assimilation and distribution of both the powerpoint presentations and digitally filmed copies of the November 3, 2003 Washington DC meeting.

Currently, Lindsay Welsh (cugenomics@cornell.edu) at Cornell University is collecting the powerpoint files which will be ready for distribution to participants by December 1, 2003. Likewise, digital copies of the meeting proceedings will be ready shortly for distribution. Several of the participants thought these items would be useful in discussing participation and funding in their own countries.

- 7) Compilation of a comprehensive list of Solanaceae researchers worldwide, including institutional affiliations, contact information (e.g. e.mail addresses) and major research focus areas. The first draft of this list should be completed and ready for distribution by December 31, 2003.