

Update on the International Tomato Sequencing Project

Joyce Van Eck^{1*}, *Jim Giovannoni*^{1,2}, *Lukas Mueller*³, *Steve Tanksley*³, *Steve Stack*⁴

¹ The Boyce Thompson Institute for Plant Research, Cornell University, Ithaca, NY 14853, USA

² Department of Agriculture-Agricultural Research Service, Plant, Soil, and Nutrition Lab, Cornell University, Ithaca, NY 14853, USA

³ Department of Plant Breeding and Genetics, Cornell University, Ithaca, NY 14853, USA

⁴ Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

The International Tomato Genome Sequencing Project is a consortium of research groups representing ten different countries committed to sequencing the gene-rich euchromatin for all twelve chromosomes, which accounts for roughly 25% of the DNA and approximately 90% of the estimated 35,000 genes. We are following a BAC-by-BAC sequencing strategy that allowed assignment of chromosomes to each participating group, often based on their interest in particular mapped loci. To date, Korea, the UK, and the Netherlands have sequenced more than 30% of the euchromatin in chromosomes 2, 4, and 6, respectively. In addition, the euchromatin of the short arm of chromosome 6 is almost completely sequenced with the exception of two gaps that will be filled by the end of 2007. The US team continues to develop resources and provide support infrastructure for the consortium. The most recent resource made available by the US team is a fosmid library of small inserts with an average size of 40 kb, which will play an important role in filling gaps along the minimum tiling path of each chromosome. To date, end sequences have been generated for 1,152 fosmid clones, and this preliminary sequence data has already been proven useful for closing a gap on chromosome 4. The repository for all data generated through this consortium is the SOL Genomics Network (SGN; www.sgn.cornell.edu). SGN continually upgrades the database with new tools for end sequence queries and repeat sequence identification, in addition to developing preliminary sequence annotation pipelines and visualization tools. We are particularly interested in feedback from genome sequence users regarding the types of tools and information that might be developed within SGN.