Proposal Title

Sequencing *Solanum pennellii* - the wild parent of the tomato introgression lines (ILs) that reveal a molecular and a system view of complex phenotypes

Sequencing Description

The genome of *S. lycopersicum* (the cultivated tomato) is being sequenced (BAC by BAC) by the International SOL Genome Project involving groups from 10 countries (~17% done; Expected completion 2009; http://sgn.cornell.edu/). The related wild tomato species *S. pennellii*, which is the founding parent of the tomato ILs (see below), originates from the deserts of the Peruvian Andes. Its genome is estimated to be $1x10^9$ bp that is highly syntenic with the sequenced edible tomato, potato, eggplant and other Solanaceae.

Sequencing Justification

Population development and the collection of multiple phenotypic data over years take longer to accomplish than a genome sequence. Immortalized segregating populations, represented by ILs, allow the integration of numerous phenotypic analyses carried out in different environments, platforms and years into a common genetic platform. Over the past 15 years, tomato researchers worldwide have utilized a complete genome coverage introgression line population (ILs, also called congenics) derived from a cross between related species (*Solanum lycopersicum* x *S. pennellii*). This population is most suitable for resolving the molecular basis of complex traits as demonstrated by the cloning of the first quantitative trait loci (QTL) *FW2.2* & *Brix9-2-5* in the year 2000. The population structure and the multiple phenotypic platforms used to characterize the tomato ILs facilitated the most advanced explorations of quantitative traits, and allowed the study of genotype by phenotype associations unparalleled by more widely used model species. Strikingly, the interspecific F1 hybrid plant that founded the population displays dramatic heterosis for biomass production, outperforming the parental lines by at least 3 folds, both in irrigated and dry conditions. This phenotypic diversity facilitated the identification and mapping of tomato overdominant QTLs for yield and biomass. To expose the genes controlling simple and quantitative phenotypes in the biodiversity IL population we need the sequence of both parental genomes.

Sequence Utilization

The public domain IL population, in a homozygous and heterozygous condition, is now the subject of several international projects aiming at integrating plant yield, biomass production, morphology and metabolism into a single map-based “system” framework. Thus, thousands of mapped QTLs that differentiate *S. lycopersicon* and *S. pennellii* have already been identified and verified. The transformation of sequence information into a “single gene” and subsequently, a system level understanding of quantitative variation would therefore become a trivial task, and a general model for genetic-sequence-physiology integration.

Genomics Community Interest

I am co-Chair the International Solanaceae Genome Project (SOL) which functions as a 'virtual umbrella' for the initiative by promoting, coordinating and actively seeking additional scientists and funding agencies to participate in an expedition into developing, understanding and utilizing natural biodiversity (<http://sgn.cornell.edu/solanaceae-project/>). In the SOL project there are more than 30 laboratories that will benefit directly from the sequence of *S. pennellii* as a bridge for linking genes and their favorite phenotypes and QTLs. Moreover, there are hundreds of Solanaceae researchers who will benefit from the availability of the sequence comparisons of the tomato species. The sequencing of *S. pennellii* will establish the technological and conceptual foundation for the longer-term vision of SOL-100, which aims to sequence and phenotype 100 diverse Solanaceae species.

DOE Mission Relevance

The development of bioremediation and biofuel systems is linked at its base to plant breeding. The tomato ILs demonstrated the powers of the “exotic library” approach in resolving basic principles in plant breeding as well as in achieving market leadership (AB2 is a *S. pennellii* based processing tomato hybrid that leads the market in California). The ILs also shed a renewed light on hererosis, or hybrid vigor, which is a major genetic force that contributes to world food and biomass production. The reasons for the superiority of the F1 hybrid over its parents are not known, and thus, our best clues to the molecular basis of these “heterosis genes” are their phenotypes, mapped in a quantitative trait loci (QTL) framework. Further understanding of the genetic basis of hybrid vigor will introduce a higher level of predictability to the domestication and breeding of biofuel or bioremediation crops.