2009 Report: Natural Variation and Comparative Genomics

Prepared by Julin Maloof (Co-chair, jnmaloof@ucdavis.edu) and J. Chris Pires (Co-chair, piresjc@missouri.edu)

Arabidopsis thaliana serves not only as a model system for understanding the genetic, molecular and biochemical functions underlying plant life, but also for determining the mechanisms by which these functions (and variation in them) contribute to ecological and evolutionary success. The ease of genetic manipulation, abundant natural variation, and rich understanding of genetic and biochemical pathways all point to the suitability of *Arabidopsis* and its relatives for ecological, quantitative genetic, and evolutionary studies. Indeed *Arabidopsis* and its relatives represent an ideal system for understanding environmental adaptation, quantitative genetic variation, and microevolution at the mechanistic level. Natural variation and comparative genomics studies are required for true understanding of how genes function. For example, understanding how genes are used to build an *A. thaliana* plant requires knowledge not only about molecular functions in *A. thaliana*, but also an understanding of why *A. thaliana* genes don't make a plant that looks more like *Capsella*, or *Brassica*, or *Cleome*, or cotton. Thus, understanding the genetic basis of developmental, metabolic, or physiological differences between species is at the very crux of plant biology. Finally, diverse species with different structures, life histories, and environmental adaptations provide tools for exploring gene function (in the molecular sense), that complement those traditionally deployed in *A. thaliana*. More generally, *A. thaliana* is second only to humans when it comes to knowledge and ability to exploit sequence variation. *A. thaliana* surpasses humans when it comes to tools available for understanding how sequence variation affects biological processes. *A. thaliana* indeed is serving as a useful model for developing methods that will be applicable in medical genetics.

Notable Advances, Publications, and New Resources

The most notable advances continue to be in the area of 'omics technology. Whole genome sequence of *Arabidopsis lyrata* has been assembled, annotated, and released (http://genome.jgi-psf.org/Araly1/Araly1.home.html). A draft sequence of papaya, a basal *Brassicales* is in the same order as Arabidopsis, has been produced [1]. The utility of these sequences can be seen in studies where comparison of the papaya, *Arabidopsis* and grape genomes led to the conclusion that gene transposition is much more frequent than commonly thought [2]. These three genomes along with poplar and rice were used to define ancient hexaploidy in the angiosperm lineage [3].

Short-read sequencing is being used to sequence many Arabidopsis accessions. Three have been published already [4], more than one hundred are in progress, and 1001 are planned in total (http://1001genomes.org/). Characterization of variation in 'omics traits, and determining loci responsible for that variation, is moving beyond the transcriptome analysis that we reported last year. A whole-genome tiling array is available and is being used for analysis of variation not only in transcript levels but also splicing, allele-specific expression, and methylation [5,6]. Similarly variation in metabolite, ion, and protein abundance is being characterized and mapped [7-10]. A SNP chip that queries 250,000 polymorphisms has been developed and is being used to genotype hundreds of natural accessions (http://walnut.usc.edu/). This data will be a tremendous resource for genome-wide association mapping. For QTL mapping, at least seven additional Recombinant Inbred Line (RIL) populations have been developed and released [11,12].

Needs and recommendations

- Infrastructure for archiving, organizing, analyzing, and displaying the huge amount of sequence data that will be generated in the next few years is needed.
- Longer funding cycles (4-5 years minimum) are needed to allow QTL mapping and identification.
- Better access to mapping populations and data is needed. Some have been deposited in the stock centers whereas others are available from individual researchers. We encourage the deposition of all mapping populations in the stock centers; some of us have had problems obtaining lines from individual researchers because of institutional Material Transfer Agreements (MTAs). The community would benefit from an organized effort to collect and organize NILs and HIFs for each RIL population. Finally, all genotyping data should be provided in a common and easily transformable format.
- Last year we discussed the need for a "fingerprinting" method for identifying *A. thaliana* stocks. SNP chip genotyping will provide the reference data, but we need to develop an inexpensive way for individual labs to fingerprint their own stocks.
- An integrated database for storing and retrieving QTL data and results, especially for 'omics traits is needed. Ideally this would use a common mapping framework to facilitate comparison among experiments and populations. This is a non-trivial task and is best carried out at the community level. It would be ideal if this could be incorporated into TAIR.

<u>References</u>

1 Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, et al (2008) The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452: 991-996.

2 Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. Genome Res 18: 1924-1937.

3 Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Res 18: 1944-1954.

4 Ossowski S, Schneeberger K, Clark RM, Lanz C, Warthmann N, Weigel D (2008) Sequencing of natural strains of Arabidopsis thaliana with short reads. Genome Res 18: 2024-2033.

5 Zhang X, Byrnes JK, Gal TS, Li WH, Borevitz JO (2008) Whole genome transcriptome polymorphisms in Arabidopsis thaliana. Genome Biol 9: R165.

6 Zhang X, Shiu S, Cal A, Borevitz JO (2008) Global Analysis of Genetic, Epigenetic and Transcriptional Polymorphisms in Arabidopsis thaliana Using Whole Genome Tiling Arrays. PLoS Genet 4: e1000032.

7 Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ (2008) Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell 20: 1199-1216.

8 Fu J, Keurentjes JJ, Bouwmeester H, America T, Verstappen FW, et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet 41: 166-167.

9 Baxter I, Muthukumar B, Park HC, Buchner P, Lahner B, et al (2008) Variation in molybdenum content across broadly distributed populations of Arabidopsis thaliana is controlled by a mitochondrial molybdenum transporter (MOT1). PLoS Genet 4: e1000004.

10 Baxter IR, Vitek O, Lahner B, Muthukumar B, Borghi M, et al (2008) The leaf ionome as a multivariable system to detect a plant's physiological status. Proc Natl Acad Sci U S A .

11 Simon M, Loudet O, Durand S, Bérard A, Brunel D, et al (2008) Quantitative Trait Loci Mapping in Five New Large Recombinant Inbred Line Populations of Arabidopsis thaliana Genotyped With Consensus Single-Nucleotide Polymorphism Markers. Genetics 178: 2253-2264.

12 Balasubramanian S, Schwartz C, Singh A, Warthmann N, Kim MC, et al (2009) QTL Mapping in New Arabidopsis thaliana Advanced Intercross-Recombinant Inbred Lines. PLoS ONE 4: e4318.