

MASC Natural Variation and Comparative Genomics- 2007 Subcommittee Reports

Prepared by Tom Mitchell-Olds (Chair, tmo1@duke.edu) *Note: expanded reports are available online at http://www.arabidopsis.org/portals/masc/masc_docs/masc_sub_rep.jsp*

Natural Variation

Arabidopsis offers a resource for studying not only gene function but also ecological processes in modern contexts. One important benefit of studying natural variation in *Arabidopsis* is the extension of gene function and effects from the individual to the species level. Additionally, plants provide the best system for studying environmental responses. Since plants have adapted to different environments, they are particularly good system for studying environmental adaptation. These analyses have clear relevance for evolutionary studies in general, as well as providing benefits for agriculture, understanding of biological responses to global climate changes, and other economically important problems. However, there are a number of challenges facing researchers studying Natural Variation in *Arabidopsis*. Even with the ever-growing set of resources available, the ability to study variation between *Arabidopsis* accessions requires more resources than most areas of *Arabidopsis* research. Seed collections of *Arabidopsis* accessions maintained by individual labs should be deposited in public stock centers. Furthermore, recent expansion and internationalization of this research community creates increasing needs for electronic and institutional facilitation of communication among research groups. Funding for *Arabidopsis* research has been declining in some regions making the competition for resources greater each year. These issues can be partly addressed through greater coordination of resources and effective communication.

Major obstacles to Natural Variation studies

- (1) Seed contamination is a consistent problem. A genetic fingerprint of relevant lines and a fast assay would be an excellent development. Several labs are working towards solutions to this problem. This approach also should be deployed at the stock centers.
- (2) Limited availability of viable seeds from related species restricts comparative studies.

Key needs in the field

- (1) Accession Resources: Recombinant inbred lines (RILs), near isogenic lines (NILs), and other pedigreed, genotyped populations are of enormous value as are genetic and molecular resources for efficient mapping and genotyping. Widespread collections of natural accessions, collections suitable for association studies and comparative sequence information from related species are needed. It would be extremely valuable to have RIL sets bulked and densely genotyped, and common sets of densely genotyped accessions are needed for genome-wide association mapping. Efforts are needed to collect and characterize populations in ancestral habitats. In the longer term, complete genomic sequences from many accessions will be extremely valuable.
- (2) Genetic tools appropriate for ecological studies are needed: The lack of ecological information on the public seed stocks is a severe impediment to our understanding of natural variation. More seed collections are needed from refugial locations and from recently introduced populations, a thorough characterization of the geographic and ecological range of *A. thaliana* is needed, and ecological information from sites of collection is essential. In addition, a series of maintained field sites that span latitude, longitude, and habitat would be extremely valuable. It would be extremely useful to have a dedicated resource for *Arabidopsis* field research in Europe and North America—European sites representing both ancestral and introduced sites, and North American sites representing unambiguously introduced sites.

- (3) Genetic resources using an appropriate genetic background for ecological studies need to be developed: Currently, many mutant lines and transgenic materials have been constructed in the Landsberg *erecta* (Ler) mutant background, however, Ler is not the most appropriate ecotype for natural variation studies and a suite of ecotypes is needed.
- (4) Discussion regarding appropriate growth conditions and phenotyping protocols, especially relating to environmental variation among natural populations, is needed: It is important to emphasize the environment-dependence of phenotypes, and the utility of standardizing phenotypic measurements across different environments.
- (5) Community databases for sharing of phenotype and QTL information are needed to allow direct comparison among studies: WebQTL/The Gene Network (<http://www.genenetwork.org/>) provides an early example, including data from *Arabidopsis* RILs. Likewise, a centralized web resource listing current and past grants on *Arabidopsis* research would be very useful.

Relevant Natural Variation databases

In addition to TAIR, most relevant databases belong to individual labs (e.g., <http://naturalvariation.org/>, <http://www.dpw.wau.nl/natural/>, <http://www.inra.fr/internet/Produits/vast/>, <http://www.genenetwork.org/>, <http://www.mpiz-koeln.mpg.de/masc/index.html>); it would be desirable to have all information on genotyping and phenotypes of accessions and mapping populations coordinated at TAIR.

Comparative Genomics

Comparative genomic data have been used to improve annotation of genes in *A. thaliana* and will allow development of more appropriate models for traits important in agricultural or natural habitats, such as apomixis, or tolerance to drought, heavy metals, or high salt concentrations. Comparative genomics with *Arabidopsis* relatives will be facilitated by rapid advances in new sequencing technologies and by three high-quality genome sequences that are pending (*A. lyrata*, *C. rubella*, and *Thellungiella halophila*). In *Brassica*, coordinated international efforts are moving forward with BAC by BAC sequencing of *B. rapa*. The genome sequence of papaya will serve as an outgroup for comparative genome analysis of the family *Brassicaceae*. Examples of other developing areas include *Thellungiella* (*T. halophila* has been approved for sequencing at JGI), heavy metal-resistant *Thlaspi* genus and *Arabidopsis halleri*, and *Boechera*, a model for ecological studies among *Arabidopsis* relatives.

Major Obstacle to Comparative Genomics

Limited financial resources for *Brassicaceae* comparative genomics is a fundamental problem. Among the close *Arabidopsis* relatives, only *Brassica* is an economically important crop plant. Consequently, funding sources which focus on *A. thaliana* or on crop species provide little support for many aspects of *Brassicaceae* comparative genomics.

Key needs in the field

- (1) Establishment of phylogeny of related species is needed: Although many clusters of closely related species are known in the *Brassicaceae*, the ordering of tribes within the family contains many unresolved nodes. This large-scale framework is essential for planning and interpretation of comparative genomics experiments. In order to resolve phylogenetic relationships across the family, information is required from multiple loci from the major groups in the crucifers. An important unsolved problem is to determine optimal evolutionary distances for genomic comparisons examining a range of evolutionary and functional questions.

- (2) Discussion and agreement on the goals for model systems and the species to develop is needed: Genomic models should be selected carefully and democratically. Important factors include phylogenetic position, interactive user communities, unusual physiological and morphological traits, and groups for which ecology is well studied. Experimental advantages include conserved orthologous markers, integrated linkage and physical maps which can be anchored to genome sequences, efficient transformation, rapid cycling lines, physiological and anatomical studies, multi-species and multi-tissue transcriptome datasets, informatics, inbred lines for sequencing, suitability for interspecific crosses and genetics, and well-documented and accessible germplasm collections.
- (3) Integrated bioinformatics tools that can move between *Arabidopsis* and other *Brassicaceae* genomes are desperately needed for the crucifer genomics community. There is also need for more taxonomic and morphological information for the evolutionary and genomics communities.

Relevant Comparative Genomics databases

Existing databases in Britain and Korea focus on *Brassica* resources (accessible via www.brassica.info) TAIR (and others) focus on *Arabidopsis* (www.arabidopsis.org). A papaya genome database is being developed by Andrew Paterson's lab.