### Report of the MASC Natural Variation Working Group-Summer 2007

### What resources are available for the community?

Recombinant inbred lines (RILs), near isogenic lines (NILs), and other pedigreed, genotyped populations are of enormous value. The community of researchers on natural variation also benefits from genetic and molecular resources for efficient mapping and genotyping. Widespread collections of natural accessions, and collections suitable for association studies are also important resources for *Arabidopsis* biology. Comparative sequence from related species will be very important for evolutionary inferences and biological comparison. The entire range of genomic, metabolomic, and proteomic technologies will contribute to identification and interpretation of natural variation in *A. thaliana*.

### Which researchers may be willing to give seeds to the stock centers?

Many researchers have seed collections of *A. thaliana* accessions, and the community has a strong tradition of providing seeds upon request. Public stock centers provide a very important service by distribution of natural accessions and mapping populations. A partial summary of available RIL populations can be found at <a href="http://www.inra.fr/vast/RILs.htm">http://www.inra.fr/vast/RILs.htm</a>. In general researchers are willing to provide their materials once results have been published. This is a rule for most journals, which should be enforced. Community standards regarding background data and protection against contamination need further discussion.

# Is there a good network for interactions among researchers? If not, what can we do to improve communication?

Historically, communication among researchers working on natural variation in *A. thaliana* has been excellent. However, recent expansion and internationalization of this research community creates increasing needs for electronic and institutional facilitation of communication among research groups.

#### How is the funding situation?

Funding for *Arabidopsis* research in the USA is difficult. Researcher perceptions of funding in Europe are mixed. Some labs report a wave of positive consideration from funding agencies. However, at the EU level, Framework 6 and 7 provide little support in this area. Researchers in the USA and EU report difficulties in obtaining funding of sufficient duration to map, clone and characterize QTLs.

### What are the key needs in the field?

Many RIL sets are in development. It would be extremely valuable to have these resources bulked and densely genotyped. Also needed are common sets of densely genotyped accessions for genome-wide association mapping, as well as appropriate methodologies for these tests, taking into account population structure. (One solution is development of an unstructured population of randomly mated genotypes containing the geographically distributed genetic diversity in which we are interested. Such a resource would greatly expand the ecological utility of association mapping.) Efforts are needed to collect and characterize populations in ancestral habitats, which may be locally adapted may represent unique ancestral diversity. Some excellent efforts to characterize this sort of "classical" adaptation are already underway. Finally, in the longer term, complete genomic sequence from many accessions (such as those with increased DNA contents relative to Col-0) will be extremely valuable.

To make use of *Arabidopsis thaliana* as an ecological research system, genetic tools appropriate for ecological studies are needed. First, little information exists on the ecological context of most of the collections in the public seed banks, many of which were collected from disturbed sites in central Europe. More seed collections are required from refugial locations and from recently introduced populations, a more thorough characterization of the actual geographic and ecological range of *A. thaliana* is needed, and ecological information from sites of collection is essential. The lack of ecological information on the public seed stocks is a severe impediment to our understanding of natural variation. In addition, a series of maintained field sites that span latitude, longitude, and habitat would be extremely valuable. Some already exist, but are maintained by individual research groups. It would be extremely useful to have a sort of 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.

In addition, it would be helpful to develop genetic resources using an appropriate genetic background for ecological studies. Currently, many mutant lines and transgenic materials have been constructed in the Landsberg *erecta* (Ler) mutant background, although the Columbia ecotype is now being developed as well. The Ler mutant background is not the most appropriate for studies in the field, and known effects of genetic background argue for development of a suite of ecotypes as genetic resources selected according to their ecological properties, such as geographic distribution and life history. Phenotypic and genetic surveys conducted by members of the *Arabidopsis* community can identify target ecotypes for development.

Discussion is needed regarding appropriate growth conditions and phenotyping protocols, especially relating to environmental variation among natural populations. It is important to emphasize the environment-dependence of phenotypes, and the utility of standardizing phenotypic measurements across different environments as a powerful tool for studying development.

Community databases for sharing of phenotype and QTL information are needed to allow direct comparison among studies. WebQTL / The Gene Network (<a href="http://www.genenetwork.org/">http://www.genenetwork.org/</a>) 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.

# What are the major obstacles in the field and how is the community approaching these problems?

One consistent problem is seed contamination. 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.

A second obstacle for comparative genomics is limited availability of viable seeds from related species.

#### What is the state of the art in this research area?

Studies of natural variation in *Arabidopsis* have been closely linked to development of genomic technologies in the broader *Arabidopsis* research community. Methods for transcription profiling, transformation, resequencing, insertional mutagenesis, positional cloning, association studies, and other approaches are an essential part of studies of natural variation.

# What can we expect from this area of research regarding plant gene function elucidation?

There are many reviews summarizing the main benefits of studying natural variation of *Arabidopsis*, which have placed *Arabidopsis* in the eco-evo context. One important benefit of studying natural variation in *Arabidopsis* is the extension of gene function and effects from the individual to the species level (broad sense gene function). In contrast to standard molecular biology's view (narrow sense gene function or simply, gene effect), understanding genes at a species level is a more complex and dynamic process that depends on time and space, and therefore genetic background. Although currently we know little about effects of genetic background on analysis of gene function, *Arabidopsis* offers an ideal system for investigating this sort of epistasis. This of course pertains to the goal of investigating genetic pathway structure and its variation, but there are still very few tools directed towards these studies.

A. thaliana offers a resource for studying not only gene function, but also ecological processes in modern contexts. We should not neglect the opportunity to study the fundamental process of range expansion using a model genetic organism. In addition, extending the properties of *Arabidopsis* as a model for ecological genetics will facilitate future work in other wild and cultivated species.

# What are the benefits of this research for *Arabidopsis thaliana*, for other plants, as well as broader impacts?

The field has developed approaches to identify the molecular bases of natural phenotypic variation due to large-effect allelic variation; thus the area is providing the first mechanistic understanding of phenotypic natural variation. However, in the long term questions that could be approached from the study of natural variation in the field (e.g., ecological questions related to adaptation) are still at a very early stage.

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.

#### What are the relevant databases?

In addition to TAIR, most relevant databases belong to particular 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) but it would be desirable that all that information on genotyping and phenotypes of accessions and mapping populations be coordinated at TAIR.

# Report of the MASC Comparative Genomics Working Group- Summer 2007 What is the state of the art in this research area?

Recent reviews of comparative genomics of *Arabidopsis* relatives have highlighted progress and possibilities in this area [1-3]. Three high-quality genome sequences are pending (*A. lyrata*, *C. rubella*, and *Thellungiella halophila*). Together with the rapid advance in new sequencing technologies, comparative genomics in Brassicaceae will see explosive growth.

In *Brassica*, coordinated international efforts are moving forward with BAC by BAC sequencing of *B. rapa*, and a number of international groups are involved with many aspects of *Brassica* research. To date, public databases have finished sequence from >500 BACs, and more than one million total sequences. The number of *Brassica* EST sequences should exceed 600,000 by summer 2007, followed by release of an Affymetrix GeneChip. Resources include BAC libraries, linkage and physical maps, TILLING and diversity populations (domesticated and wild), QTL studies, and rapid cycling lines. Physical mapping of *B. oleracea* and its close diploid relative, *Sisymbrium irio*, are in progress.

Caricaceae (papaya family) is closely related to the Brassicaceae in the order Brassicales. The genome sequence of papaya will serve as an outgroup for comparative genome analysis of the family Brassicaceae. Whole genome shotgun sequencing of the papaya genome has been completed to approximately 5x coverage. These sequences will be deposited to GenBank in the second half of 2007. Genetic and physical maps, SSR markers, cDNA and BAC libraries are available.

The *Thellungiella* [*Eutrema salsuginea*] research community is working on multiple aspects of tolerance to abiotic stress [4], especially salt tolerance [5]. Full-length cDNA and BAC libraries are being used in *Arabidopsis* transformation and forward screening for altered phenotypes, and a T-DNA insertion strategy has generated close to 100,000 transformed lines. *T. halophila* has been approved for sequencing at JGI.

Genus *Thlaspi* [6] shows genetic resistance to heavy metals. Analyses of linkage and QTL maps [7,8], microarrays [9,10], and ESTs [11] have been published. A fosmid library is available, and successful transformation [12] has been reported. In addition, studies of heavy metal tolerance have made progress in *Arabidopsis halleri* [13,14], as well as evolutionary studies of genetic variation [15].

Genus *Boechera* has become a model for ecological studies among *Arabidopsis* relatives, showing genetic variation for apomixis, drought tolerance, insect resistance, flowering time, and other traits of ecological importance [16]. A dense linkage map is available [17], three BAC libraries have been constructed, positional cloning efforts are ongoing, and recombinant inbred lines will be completed in 2007. In addition, light shotgun sequencing and bioinformatic analysis of promoter conservation is available [18].

Isatis (dyers woad) (*Sisymbrieae*) is the target of molecular studies and diversity comparisons in the EU, and commercial production of indigo from woad is being carried out in a variety of centers in Europe.

Unpublished work on *Arabis alpina* by Weigel and colleagues includes light shotgun sequencing, a linkage map, and a BAC library.

Cleomacaceae is the sister family to the Brassicaeae [2]. *Cleome spinosa* has a BAC library and some comparative BAC sequencing [19], as well as progress towards a linkage map.

# What are the benefits of this research for *Arabidopsis thaliana*, for other plants, as well as broader impacts?

Comparative genomic data has been used to improve annotation of genes in *A. thaliana* [20], and will enable phylogenetic footprinting on a whole-genome scale in order to identify functionally important conserved regulatory regions. Already, large-scale comparisons of *Brassica*, *Boechera*, and papaya genome sequences are being used to identify conserved non-coding sequences [18,21]. Ultimately, comparative approaches may also enable identification of conserved regulatory networks. In addition, genomic comparisons can identify genes in the Brassicaceae which have been deleted from *A. thaliana* [11], making these loci high-priority candidates for functional analysis.

Comparative genomics also 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. The broad physiological and morphological diversity of the Brassicaceae will enable new understanding of adaptive evolution and species differences [e.g., 22]. By looking systematically at the effects of genes from other species in *Arabidopsis* and *vice versa* we can hope to elucidate developmental processes that might be hard to study using conventional forward and reverse genetic approaches. Greater understanding of the effects of copy number variation will also be possible, by contrasting whole-genome duplication versus simple gene duplication. Recent identification of conserved linkage blocks and the n = 8 ancestral karyotype [3,23,24] also plays a central role in comparison and synteny studies with species in much of the Brassicaceae.

# Which technologies and approaches have been developed?

Development of transformation for a number of related species provides important tools for comparative genomics, although facile transformation by floral dip is still very limited. The use of new high throughput sequencing technologies has allowed for cheaper sequencing of BAC pools, as well as fairly extensive EST sequencing. High throughput SNP scoring is an enormous benefit for development of linkage maps.

#### What resources are available for the community?

Of the resources discussed above, nearly all are available to the research community, or soon will be released publicly.

### Which researchers may be willing to give seeds to the stock centers?

Many researchers have seed collections of species and accessions of Brassicaceae, and the community has a strong tradition of providing seeds upon request. There is a clear need for central repositories to store, maintain, and distribute seeds of *Arabidopsis* relatives to the research community. Possible solutions to this problem are under discussion by the Brassicaceae Phylogeny Working Group. Diversity Fixed Foundation Sets are being developed for several *Brassica* species, with seed available via WarwickHRI Genomics Resource Centre. For *Thlaspi*, NASC (Nottingham Arabidopsis Stock Centre) has offered to distribute seeds. Successful solutions for germplasm distribution will require financial support.

#### What are the relevant databases?

Existing databases in Britain and Korea focus on *Brassica* resources (accessible via www.brassica.info), and TAIR and others on *Arabidopsis*. A papaya genome database is being developed by Andrew Paterson's lab. Integrated bioinformatics tools that can move between *Arabidopsis* and other Brassicaceae genomes [e.g., 25,26] are desperately needed for the crucifer genomics community. There is also need for more taxonomic and morphological information for the evolutionary and genomics communities.

# Is there a good network for interactions among researchers? If not, what can we do to improve communication?

Historically the plant biology community has been divided between systematists and population geneticists on the one hand, and functionally oriented molecular biologists on the other. Recently, however, both groups have begun to recognize the scientific benefits which can come from an integrated approach to comparative plant biology. While there has always been a cropcentric focus in the *Brassica* community, there is a recent trend toward more networking and sharing of resources, with more awareness and interaction between a range of researchers in different countries. The current situation provides a good beginning, but coordinated efforts to increase communication are fundamentally important for future progress.

# How is the funding situation?

DOE-JGI is funding genomic sequencing of *A. lyrata*, *C. rubella*, and *Thellungiella* genomes. At the US National Science Foundation there has been a tendency for Brassicaceae comparative genomics to fall between existing panels dealing with evolution versus functional genomics, or Arabidopsis versus plant genomics. As a consequence, some efforts in this field have been viewed as peripheral to the priorities of existing panels, and therefore not fundable in either forum.

# What are the key needs in the field?

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. It is likely that optimal distances for these comparisons will differ for various goals and questions.

Agreement is needed within the community on goals for model systems in comparative genomics, and on species which support progress towards these goals. These genomic models should be picked 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.

# What are the major obstacles in the field and how is the community approaching these problems?

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.

### What can we expect from this area of research regarding plant gene function elucidation?

Other species can be viewed as a source of mutant alleles that can be used to understand gene function. Genes in other species will often have retained core functions while there has been loss or modification of peripheral functions. For example, genes that have a core function that results in lethality of knockouts may have numerous other functions that cannot easily be detected by reverse genetics, because mutants don't get far enough in development. A gene from another species may however reveal these hidden functions. Also, there is broader interest in how genomes diverge in evolutionary timescales that can only be addressed comparatively.

# How useful are Arabidopsis thaliana and other near relatives for comparative genomics?

Taken together, genomics of *Arabidopsis* and Brassicaceae relatives provides an ideal system to address fundamental questions in plant biology. Partly this is because of available tools and the fact that in *A. thaliana* we have an excellent understanding the way that a genotype yields a phenotype – though we still have a long way to go. Most Brassicaceae genomes are smaller (and therefore more cost effective) than comparable studies of grasses or vertebrates. Furthermore, many Brassicaceae are small and rapid cycling, which offers many practical advantages for experimentation.

#### **Committee Members:**

Thomas Mitchell-Olds (USA), tmo1@duke.edu (Chair)

David Baum (USA), dbaum@wisc.edu

Hans Bohnert (USA), bohnerth@life.uiuc.edu

Sean Cutler (Canada), cutler@botany.utoronto.ca

Graham King (UK), graham.king@bbsrc.ac.uk

Marcus Koch (Germany), marcus.koch@urz.uni-heidelberg.de

Martin Lysak (Czech Republic), lysak@sci.muni.cz

J. Chris Pires (USA), piresjc@missouri.edu

Ray Ming (USA), rming@life.uiuc.edu

Eric Schranz (Netherlands), schranz@science.uva.nl

Detlef Weigel (Germany), weigel@tuebingen.mpg.de

11 March 2007

#### **References:**

- 1. Clauss MJ, Koch MA: Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* 2006, 11:449-459.
- 2. Schranz M, Song B-H, Windsor A, Mitchell-Olds T: Comparative genomics in the Brassicaceae: a family-wide perspective. *Curr Opin Plant Biol* 2007, 10:168-175.
- 3. Schranz M, Lysak M, Mitchell-Olds T: The ABC's of comparative genomics in the Brassicaceae: Building blocks of crucifer genomes. *Trends in Plant Science* 2006, 11:535-542.
- 4. Wong CE, Li Y, Labbe A, Guevara D, Nuin P, Whitty B, Diaz C, Golding GB, Gray GR, Weretilnyk EA, Griffith M, Moffatt BA: Transcriptional Profiling Implicates Novel Interactions between Abiotic Stress and Hormonal Responses in Thellungiella, a Close Relative of Arabidopsis. *Plant Physiol.* 2006, 140:1437-1450.
- 5. Gong QQ, Li PH, Ma SS, Rupassara SI, Bohnert HJ: Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. *Plant Journal* 2005, 44:826-839.
- 6. Peer WA, Mahmoudian M, Freeman JL, Lahner B, Richards EL, Reeves RD, Murphy AS, Salt DE: Assessment of plants from the Brassicaceae family as genetic models for the study of nickel and zinc hyperaccumulation. *New Phytologist* 2006, 172:248-260.
- 7. Assuncao AGL, Pieper B, Vromans J, Lindhout P, Aarts MGM, Schat H: Construction of a genetic linkage map of *Thlaspi caerulescens* and quantitative trait loci analysis of zinc accumulation. *New Phytologist* 2006, 170:21-32.
- 8. Deniau AX, Pieper B, Ten Bookum WM, Lindhout P, Aarts MGM, Schat H: QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator Thlaspi caerulescens. *Theoretical and Applied Genetics* 2006, 113:907-920.
- 9. Hammond JP, Bowen HC, White PJ, Mills V, Pyke KA, Baker AJM, Whiting SN, May ST, Broadley MR: A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist* 2006, 170:239-260.
- 10. van de Mortel JE, Almar Villanueva L, Schat H, Kwekkeboom J, Coughlan S, Moerland PD, Ver Loren van Themaat E, Koornneef M, Aarts MGM: Large Expression Differences in Genes for Iron and Zinc Homeostasis, Stress Response, and Lignin Biosynthesis Distinguish Roots of Arabidopsis thaliana and the Related Metal Hyperaccumulator Thlaspi caerulescens. *Plant Physiol.* 2006, 142:1127-1147.
- 11. Rigola D, Fiers M, Vurro E, Aarts MGM: The heavy metal hyperaccumulator *Thlaspi* caerulescens expresses many species-specific genes, as identified by comparative expressed sequence tag analysis. *New Phytologist* 2006, 170:753-765.
- 12. Peer WA, Mamoudian M, Lahner B, Reeves RD, Murphy AS, Salt DE: Identifying model metal hyperaccumulating plants: germplasm analysis of 20 Brassicaceae accessions from a wide geographical area. *New Phytologist* 2003, 159:421-430.
- 13. Talke IN, Hanikenne M, Kramer U: Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator *Arabidopsis halleri*. *Plant Physiol*. 2006, 142:148-167.
- 14. Mirouze M, Sels J, Richard O, Czernic P, Loubet S, Jacquier A, Francois IEJA, Cammue BPA, Lebrun M, Berthomieu P, Marques L: A putative novel role for plant defensins: a

- defensin from the zinc hyper-accumulating plant, Arabidopsis halleri, confers zinc tolerance. *The Plant Journal* 2006, 47:329-342.
- 15. Pauwels M, Frerot H, Bonnin I, Saumitou-Laprade P: A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: Arabidopsis halleri (Brassicaceae). *Journal of Evolutionary Biology* 2006, 19:1838-1850.
- 16. Knight C, Vogel H, Kroymann K, Shumate A, Witsenboer H, Mitchell-Olds T: Expression profiling and local adaptation of *Boechera holboellii* populations for water use efficiency across a naturally occurring water stress gradient. *Molec Ecol* 2006, 15:1229-1237.
- 17. Schranz M, Windsor A, Song B, Lawton-Rauh A, Mitchell-Olds T: Comparative genetic mapping in *Boechera stricta*, a close relative of *Arabidopsis*. *Plant Physiol*. 2007:in press.
- 18. Windsor AJ, Schranz ME, Formanová N, Gebauer-Jung S, Bishop J, Schnabelrauch D, Kroymann J, Mitchell-Olds T: Partial shotgun sequencing of the *Boechera stricta* genome reveals promoter conservation and microsynteny with *Arabidopsis*. *Plant Physiology* 2006, 140:1169-1182.
- 19. Schranz ME, Mitchell-Olds T: Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 2006, 18:1152-1165.
- 20. Ayele M, Haas BJ, Kumar N, Wu H, Xiao Y, Van Aken S, Utterback TR, Wortman JR, White OR, Town CD: Whole genome shotgun sequencing of *Brassica oleracea* and its application to gene discovery and annotation in *Arabidopsis*. *Genome Res.* 2005, 15:487-495.
- 21. Haberer G, Mader M, Kosarev P, Spannagl M, Yang L, Mayer K: Large-scale cis-element detection by analysis of correlated expression and sequence conservation between *Arabidopsis* and *Brassica oleracea*. *Plant Physiol*. 2006:in press.
- 22. Hay A, Tsiantis M: The genetic basis for differences in leaf form between *Arabidopsis* thaliana and its wild relative *Cardamine hirsuta*. *Nature Genetics* 2006, 38:942-947.
- 23. Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiate DJ: Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* 2005, 171:765-781.
- 24. Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I: Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl. Acad. Sci. USA* 2006, 103:5224-5229.
- 25. Li X, Jewell E, Erwin T, Lim G, Love C, Batley J, Spangenberg G, David Edwards D: Brassica Bacterial Artificial Chromosome (BAC) and Gene Annotator. *Cruciferae Newsletter* 2007, 27:in press.
- 26. Lim G, Jewell E, Li X, Erwin T, Love C, Spangenberg G, David Edwards D: CMap: A Comparative Mapping Tool for Brassica. *Cruciferae Newsletter* 2003, 27:in press.