

2008 Phenomics Subcommittee Report

Prepared by Eva Huala (Co-chair, huala@acoma.stanford.edu) and Sean May (Co-chair, sean@arabidopsis.info)

Phenomics seed and DNA resources

- Homozygous mutant T-DNA collection

A set of genetically purified, confirmed T-DNA insertion lines consisting of two alleles for each Arabidopsis gene is under development by The Salk Institute Genome Analysis Laboratory (SIGnAL) and is being made available from the Arabidopsis Biological Resource Center (ABRC). This phenome-ready population will consist of 50,000 lines, the goal being to confirm and purify two insertion alleles each for approximately 25,000 loci. To date, 19,516 of these confirmed lines have been received by ABRC. A complete set of the lines as well as a one-allele per locus (“unigene”) set are being made available from ABRC at an economical price. First installments of 8,889 lines of a complete set and 6,868 lines of a one-allele set are presently being distributed. The confirmed population is also being organized into pools of different sizes to allow efficient forward phenotypic screening for traits that can be identified within larger populations. More details can be found in the MASC thermometers (Figure 2).

- RNAi clone resources:

The AGRİKOLA consortium has constructed a collection of approximately 29,000 Gateway entry plasmids, with a subset of about 27,000 transferred into hairpin RNA expression vectors, each capable of triggering RNAi against a defined target sequence in an Arabidopsis transcript. In addition, a small number of RNAi clones were made by the Chromatin Functional Genomics Consortium (CFGC) against chromatin remodeling transcripts, and more recently, a collection of artificial microRNAs (amiRNAs) was developed by researchers at Cold Spring Harbor Laboratory. There are nearly 36,000 RNAi plasmids targeting 22,969 unique loci, with transformed plant lines available for 3,592 loci. AGRİKOLA and CFGC clones are available at low cost through the main stock centers ABRC and NASC (See Table 4 below and Figure 2 in the MASC thermometers section of the report). A subset of higher-priced clones (due to cost-recovery for clone purification and sequence validation) are distributed through BCCM/LMBP. The amiRNAs are currently only being distributed at substantially higher cost through the Open Biosystems company, although they are expected to be deposited into ABRC eventually (www.openbiosystems.com/RNAi/ArabidopsisthalianaamiRNA).

Table 4. RNAi resources for Arabidopsis.

Creator	Format	Focus	Validation	Scale	URL	Stock center
<i>RNAi clones</i>						
AGRIKOLA	Gateway entry		PCR sized insert	28,000	www.agrikola.org	NASC
AGRIKOLA	Constitutive hp RNA expression		PCR sized insert	26,000	www.agrikola.org	NASC
AGRIKOLA	Gateway entry		Pure, sequence validated	1,000	bccm.belspo.be/db/lmbp_gst_clones/	BCCM/LMBP
AGRIKOLA	Constitutive hp RNA expression		Pure, sequence validated	800	bccm.belspo.be/db/lmbp_gst_clones/	BCCM/LMBP
CFGC	ds RNA expression	Chromatin remodeling	Single pass seq	200	www.chromdb.org	ABRC
amiRNA Central	Artificial miRNA		Full sequence	10,000	2010.cshl.edu	Open Bio-systems Inc.

Table Reference: Bleys, A., Karimi, M., and Hilson, P. (2008) Cloned-based functional genomics. *Methods Mol. Biol.* In press.

Stock centers distributing Arabidopsis clone repertoires:

- Arabidopsis Biological Resource Center (ABRC, USA), <http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm>
- RIKEN BioResource Center (BRC, Japan), <http://www.brc.riken.jp/lab/epd/Eng/catalog/pDNA.shtml>
- GABI Primary Database (GABI/RZPD, Germany), <http://gabi.rzpd.de/>
- National Resources Centre for Plant Genomics (CNRGV, France), <http://cnrgv.toulouse.inra.fr/ENG/index.html>
- European Arabidopsis Stock Centre (NASC, United Kingdom), <http://arabidopsis.info/>
- BCCM/LMBP Plasmid and DNA library collection (BCCM/LMBP, Belgium), http://bccm.belspo.be/db/lmbp_gst_clones/
- Open Biosystems Inc., www.openbiosystems.com/

Phenotype annotation tools and ontologies

Phenote, a curation tool to facilitate annotating phenotypes using ontologies, has been developed by National Center for Biomedical Ontology (NCBO) in collaboration with Berkeley Bioinformatics and Ontologies Project (BBOP) (www.phenote.org). It's currently in use by several model organism databases including Flybase, ZFIN, WormBase and others. Data annotated with Phenote is based on the EQ model for representing phenotypes, combining entities from any ontology (for example GO or PO) with qualities (such as those found in PATO). The main funding for Phenote development is due to end in July 2008. Additional funding is being sought but there will be a funding gap of at least a few months with the exception of an effort to adapt Phenote for annotation of images, which will continue as part of another project.

Several ontologies useful for controlled vocabulary phenotype annotation can be accessed at the Open Biomedical Ontologies website (www.obofoundry.org/). These include the Plant Ontology (PO) Structure (for plant anatomical parts) and Growth/Developmental Stage ontologies, the GO biological process and cellular component ontologies and the PATO ontology for phenotypic qualities. In addition, the OBO website contains an ontology for units of measurement which could be used in combination with other ontologies to capture quantitative phenotypes, and ontologies of experimental conditions (OBI) and chemical entities of biological interest (CHEBI) useful for capturing conditional phenotypes that are only apparent after experimental manipulation.

High throughput phenotyping projects and data

- Barry Pogson, Plant Phenomics project:

Funding for the \$50M Australian Plant Phenomics Facility (APPF, www.plantphenomics.org.au) has now been secured. The APPF will be based across two nodes located at CSIRO/ANU in Canberra and UA in Adelaide. Construction of both APPF facilities will begin in May 2008 with full commissioning of stage one of the Arabidopsis screening module at the High Resolution Plant Phenomics Centre in Canberra (medium throughput growth and chlorophyll fluorescence screening with mathematical morphological analysis and phenomic database capability) occurring at the end of 2008. Throughput will increase with time until the full HTP Arabidopsis module is completed in Canberra at the end of 2009, by which time the Plant Accelerator automated glasshouse facility in Adelaide will also be commissioned. The NCRIS funded National Facility will be available to researchers at the marginal cost of running the facility and several international collaborations are being established and encouraged. The resulting phenotype data will be freely released to the international community following a quarantine period (6-12 months) to allow data to be prepared for publication.

- Minami Matsui, RIKEN:

Phenotype data for 140 Ac/Ds transposon insertion lines (RAPID) generated by Dr. Takashi Kuromori and having visible phenotypes (out of 4000 lines examined) are freely available (<http://rarge.gsc.riken.jp/phenome/>) and their associated phenotype data will be integrated into TAIR. An additional set of 500 activation-tagged lines with phenotypes (generated by Dr. Youichi Kondou) can be

found at (<http://amber.gsc.riken.jp/act/top.php>). Note: the current requirement to sign a Materials Transfer Agreement before viewing the data is expected to be lifted soon. A new project to generate Arabidopsis Full-length cDNA overexpressing (FOX) lines for 13,000 Arabidopsis full-length cDNAs generated by Dr. Takanari Ichikawa is now underway and these will be made accessible from RIKEN BRC. A similar project for overexpressing rice cDNAs in Arabidopsis can be found at (<http://ricefox.psc.riken.jp/login/>) but is currently only accessible to Japanese scientists (this restriction will be lifted by the end of 2008). A hub database project has been organized by Dr. Tetsuro Toyoda at RIKEN to connect Arabidopsis genome and phenome information, including Ac/Ds and Activation-tagged line projects described above. See (<http://omicspace.riken.jp/>) for a full description of the new database and (www.psc.riken.go.jp/english/database/index.html) for access to current RIKEN databases.

- Detlef Weigel, Max Planck Institute:

Population-wide association mapping has the advantage that it can potentially resolve very finely the location of causal variants affecting a trait but this approach has the inherent disadvantage that population structure has a profound confounding effect. Essentially the converse is true for Quantitative Trait Locus (QTL) mapping. One way to combine the advantages of both is perform parallel QTL mapping across several populations, and to exploit both QTL interval and haplotype information across populations. This approach, dubbed nested association mapping, has recently been introduced by Ed Buckler and colleagues for maize. To evaluate the appropriateness of this approach for *A. thaliana*, the Weigel laboratory has generated 15 F2 populations of 480 individuals each. Because the parents were drawn from the 20 accessions investigated with Perlegen arrays (Clark et al., Science 2007), detailed SNP information – on average, 1 SNP per kb in each accession – is available for the parents, and moderately dense genotyping with 400 intermediate-frequency markers will allow almost complete reconstruction of haplotypes in the 7200 F2 individuals. All individuals have been phenotyped for a number of life history traits and images have been acquired at regular intervals for each individual during vegetative growth. A database that accepts both genotype and phenotype, including image data, is being developed and the data will be made freely available following publication.

- Pierre Hilson, Christine Granier, Cyril Pommier, AGRON-OMICS project

The Agron-omics project, which stands for Arabidopsis Growth Network integrating OMICS technologies (<http://www.agron-omics.eu/>), will conduct an in-depth study of leaf growth in the model plant species *Arabidopsis thaliana*. Over a five year period starting November 2006, this network of European plant biology researchers will perform experiments to identify the molecular components controlling growth and build mathematical models to explain how these components interact. A major component is PHENOPSIS, an automated platform for Arabidopsis leaf growth phenotyping developed at INRA (Granier et al., 2006 New Phytologist). Phenotyping is focused on three main areas: 1) high-throughput phenotyping of leaf growth response to environmental stresses in different collections of accessions (ERANET, ARABRAS project, 2007-2010); 2) identification of leaf growth QTLs in different populations of recombinant inbred lines grown in different environmental conditions (at this time, Ler x An-1, Ler x Cvi-0 and Bay-0 x Sha in different day-lengths, different incident light and different soil water contents) [GENOPLANTE DNV project (2007-2010)]; 3) high-throughput phenotyping of leaf growth in genotypes affected in cell cycle, endoreduplication, cell wall properties, metabolism, hormonal status, circadian rhythm and flowering time (European Integrated Project, FP6, AGRON-Omics, 2006-2011). Data from PHENOPSIS will be made freely available following publication. A second phenotyping platform with higher throughput but with a less precise description of the phenotype and environment is under development by INRA at the Versailles campus (www-ijpb.versailles.inra.fr/en/sgap/equipes/crg/phenotypage/index.html). A prototype has already been created in collaboration with ISEP (Institut Supérieur d'Electronique de Paris) and CEGS-DESTEC. Currently the IJPB, with financial support from the Région Ile de France, INRA and Génoplante, is building a new facility for 300 m² of culture chambers, which will eventually accommodate the IJPB high throughput phenotyping platform. A dedicated chamber of 60m², fully automated, will allow experiments with more than 10,000 individuals in parallel