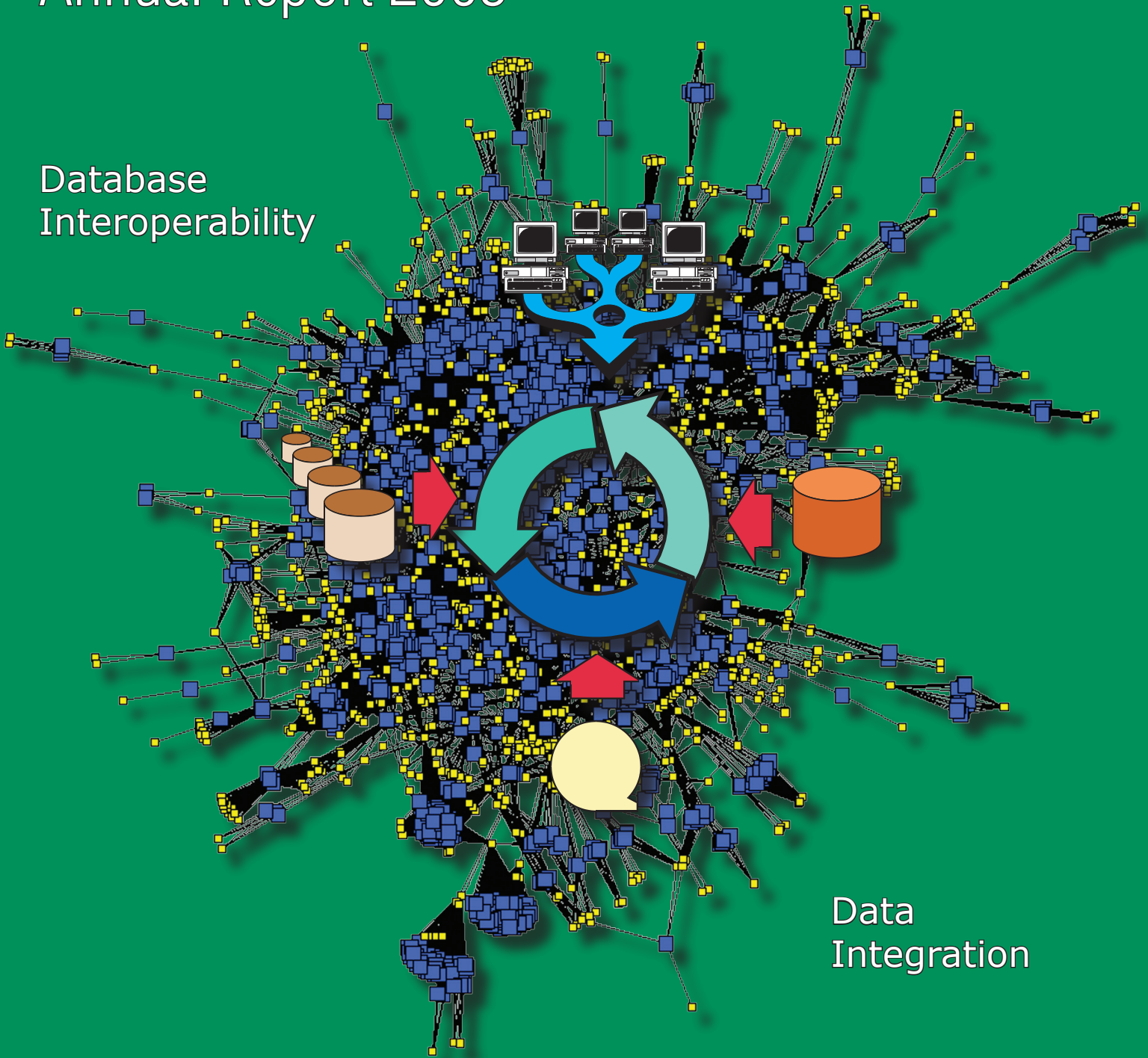# The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project

Annual Report 2005

Database
Interoperability

Data
Integration

The Multinational Coordinated
Arabidopsis thaliana
Functional Genomics Project
Annual Report 2005

Philip Benfey philip.benfey@duke.edu
Chair

Ian Small small@evry.inra.fr
Co-chair

Thomas Altmann altmann@mpimp-golm.mpg.de
Jorge Casal casal@ifeva.edu.ar
Bill Crosby bcrosby@cs.usask.ca
Martin Fellner emfee@prfholnt.upol.cz
Ian Furner ijf@mole.bio.cam.ac.uk
Mary Lou Guerinot Mary.Lou.Guerinot@Dartmouth.edu
Pierre Hilson pierre.hilson@gengenp.rug.ac.be
Heribert Hirt HEHl@gem.univie.ac.at
Gerd Jürgens gerd.juergens@uni-tuebingen.de
Jaakko Kangasjärvi jaakko.kangasjarvi@helsinki.fi
Keith Lindsey keith.lindsey@durham.ac.uk
Sean May arabidopsis@nottingham.ac.uk
Andrew Millar Andrew.Millar@ed.ac.uk
Peter McCourt mccourt@botany.utoronto.ca
Javier Paz-Ares jpazares@samba.cnb.uam.es
Ben Scheres b.scheres@bio.uu.nl
Kazuo Shinozaki sinozaki@rtc.riken.go.jp
Willem Stiekema w.j.stiekema@plant.wag-ur.nl
Paola Vittorioso paola.vittorioso@uniroma1.it
Geoffrey Wasteneys geoffw@rsbs.anu.edu.au
Yang Weicai wcyang@genetics.ac.cn

Isabell Witt isawi@duke.edu
Executive Secretary and Coordinator

The Multinational *Arabidopsis* Steering Committee · June 2005

Cover: The network diagram represents a qualitative model of the *Arabidopsis* metabolic and regulatory network by Rodrigo A. Gutierrez from the laboratory of Gloria Coruzzi at the Department of Biology, New York University (Gutierrez et al. (2005) Submitted).

The data flow diagram overlaying the *Arabidopsis* qualitative network model represents a software architecture designed for interoperability characteristic of the BioMoby protocol Wilkinson M, Schoof H, Ernst R, Haase D (2005). BioMOBY successfully integrates distributed heterogenous bioinformatics web services. The PlaNet exemplar case. Plant Physiol 138, p1-13.

Table of Contents

Foreword to the Report

This is the 2004/2005 annual report of the **Multinational *Arabidopsis* Steering Committee (MASC) on the status of the Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project.** This 10-year program initiated in 2001 was described in a long-range plan for this new phase of the *Arabidopsis* Genome Research Program in the 2002 MASC report. The goals of this project are to determine the function of every *Arabidopsis* gene and obtain a detailed and comprehensive understanding of the molecular processes underlying the development, metabolism and interaction with the environment of a flowering plant. The intent is that the knowledge gained on this experimental model organism will serve as the central reference and conceptual framework for all of plant biology. *Arabidopsis* is uniquely situated to play this role for a number of reasons: 1) Its genome is the best characterized among plants, 2) It has the most comprehensive reverse and forward genetic tools and resources, 3) The international research community that uses these tools and resources is among the most active and co-operative, and 4) The solution of most problems in plant biology, whether applied or basic, can be achieved rapidly and in a cost-effective manner through the use of *Arabidopsis*.

The results generated by the *Arabidopsis* Functional Genomics Project, which are currently made publicly available through central databases, not only provide unprecedented insight into plant function by uncovering basic biological concepts, but also greatly advance our knowledge of the genetic determinants of important traits in crop plants. Studies comparing the physiology, biochemistry, and development of *Arabidopsis* with that of other plant taxa and with economically important agricultural species will be of increasing importance.

The availability of the complete genome sequence of *Arabidopsis thaliana*, the ultimate accomplishment of the previous phase of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project, provided a "quantum leap" in the information base for plant molecular biological research. On the one hand, this information allowed the *Arabidopsis* research community to develop new approaches and research tools. On the other hand, it highlighted the enormous complexity of the plant biological system and the difficulty of deciphering the function of every gene. Nevertheless, a new long-term goal emerged: "To uncover the mechanisms and processes underlying the spatial, temporal-, and conditional control of the activity of the genes, the identity, function-, and localization of gene products and their interactions in the cellular context, which are the basis of the multitude of cellular, physiological and developmental processes of plant life." To accomplish this goal requires the use of newly developed high throughput technologies, novel experimental tools and comprehensive collections of plant resources as well as powerful procedures for data analysis, storage and display.

Arabidopsis research has provided the cutting edge in generating resources and analytical tools, providing an example for the investigation of other plant species. One of the most important determinants of the success of the *Arabidopsis* Functional Genomics Project is the integration of worldwide efforts. The nature and volume of the proposed work necessitates the marshalling of all resources to attain a maximum level of synergy as well as the avoidance of duplication of effort to enable the *Arabidopsis* community to achieve its ambitious goals. The Multinational *Arabidopsis* Steering Committee plays a key role in supporting this international coordination by collecting and disseminating information from the various initiatives and projects on technology development and functional analyses and by giving specific recommendations for further directions.

As is outlined in this report, the high level of cooperation and widespread willingness to share data throughout the *Arabidopsis* community as well as the support by the funding agencies has yielded important and exciting results. **These favorable developments indicate that the ambitious goal of determining the function of all *Arabidopsis* genes as a first step toward an in-depth understanding of the biology of higher plants to the benefit of our society can be achieved, as biological materials and services are continued to be made available around the world, and human resources are further developed.**

The Multinational *Arabidopsis* Steering Committee
June 2005

Executive Summary

As the first plant to have its genome sequenced in its entirety, *Arabidopsis* continues to play a leading role in providing insights into the functioning of flowering plants. The community of *Arabidopsis* researchers now numbers more than 13,000 world-wide and has set as its primary goal the understanding of the function of all *Arabidopsis* genes in their cellular, organismal and evolutionary contexts. This systematic and thorough analysis has already begun to shed light on fundamental biological processes as well as pathways of significant agronomic importance.

The specific aims of the multinational *Arabidopsis thaliana* functional genomics project comprise short-term, mid-term and long-term goals. Many of the short term goals have been achieved already:
- Genome annotation has improved dramatically each year through the use of different sources of expressed sequences and expert knowledge
- Comprehensive sets of sequence-indexed mutants have been created and are widely used for gene function determination
- Gene-specific probe sets covering the entire genome have been created and are widely used for gene expression analysis
- Full-length cDNA sequences have been defined for 70% of all *Arabidopsis* genes
- Databases have been greatly expanded providing access to increasingly large data sets of all formats
- Metabolic profiling facilities for global metabolite analysis have been established

This past year's progress included the following achievements:
- Availability of full length cDNA sequences covering 70% of the protein coding genes, sequences are available at the major data bases, clones are available from the stock centers.
- Knowledge of expression of 92% of all protein coding genes
- Increasing knowledge of gene function
- Identification of insertional mutations for 24,654 genes, of which a high percentage are knock-out mutations
- Availability of 20,000 knock down RNAi constructs
- A growing and freely accessible transcriptome reference data set with tools for mining are available at many different databases world wide

- Release of a full genome *Arabidopsis* Tiling chip for multiple applications
- A Workshop for Data Integration and Database Interoperability in April 2005
- Establishment of a Proteomics Working Group
- Establishment of a Working Group on Natural Variation and Comparative Genomics

The resources, data/information, and tools are freely available and accessible to the scientific community. Ultimate success of the Project will be judged by how well they contribute to advances in our understanding of the biology of plants. In just a few years since the initiation of the project, the field has seen many examples of new discoveries made possible because of the availability of the research resources. These include the recent identification of microRNAs controlling developmental processes (e.g. Guo HS, Xie Q, Fei JF, Chua NH., MicroRNA Directs mRNA Cleavage of the Transcription Factor NAC1 to Downregulate Auxin Signals for *Arabidopsis* Lateral Root Development., Plant Cell. 2005 17:1376), the characterization of an unusual form of non-mendelian inheritance (Lolle SJ, Victor JL, Young JM, Pruitt, Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. Nature. 2005, 434:443. the application of Tiling arrays (Mockler, T.C. and J.R. Ecker. 2005. Applications of DNA tiling arrays for whole genome analysis. Genomics 85:1-15.) and the global analysis of developmental AtGenExpress data (Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. Nat Genet. 37(5):501-6. Epub 2005 Apr 3.)

The MASC's (Multinational *Arabidopsis* Steering Committee) short term plan for the next year include:
- A complete collection of homozygous knock out lines and standardized description of their phenotypes
- Resources for studying protein localization, protein interaction networks including complete full length cDNA and ORFeome clones and tagged ORFeome clones for many different purposes
- Achieve better data integration and interoperability across data bases for optimal exploitation of genomic data resources

- Improve the usability of the Project's homepage at TAIR (the *Arabidopsis* Information Resources)

Strong international collaboration at all levels continues to be a hallmark of the multinational coordinated *Arabidopsis thaliana* functional genomics project. Many national projects are underway as outlined in this report. The MASC will continue to monitor the Project's progress overall and coordinate activities of the *Arabidopsis* research community to ensure progress toward the ultimate goal of understanding the function of all genes in *Arabidopsis*, the reference flowering plant.

Analysis and Recommendations

The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project has now completed its fourth year. In 2004/2005 there was an enormous increase in publicly accessible data and resources including SNP, MPSS, microarray, full length-cDNA information as well as full-length cDNA clones, ORF clones and T-DNA mutants. This report contains numerous examples of shared materials, resources, and data, which are widely used throughout the community. In analogy to the establishment of a reference transcriptome by international cooperation, driven by researchers with common goals and supported by national funding agencies it is now time to establish the reference proteome and metabolome. Again only international cooperation by ambitious researchers, a high level of coordination and sufficient funding will make this possible. Comparative genomics efforts between *Arabidopsis* accessions as well as with closely related species were started by an international group of researchers in 2004.

Current status of the program

This spirit of cooperation has contributed to fulfilling most of the short-term goals as well as the initiation of some mid-term goals (e.g. the reference transcriptome and the production of recombinant inbred lines) of the Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project. In the first four years the major accomplishments included:

1) The release of improved whole genome annotations, the most recent versions supported by full-length cDNA sequences and expert input. Improved genome annotation (i.e., identification of genes) is one of the most notable recent achievements. Of critical importance was the incorporation of information from full-length cDNA sequences. TAIR has taken over the task of maintaining and improving the *Arabidopsis* genome annotation in 2004 and a first release is planned for June 2005.

2) Generation of comprehensive sets of sequence-indexed mutants, listed in an integrated database and made available as seed stocks. Among the most highly utilized tools for the determination of gene function are vastly expanded resources for reverse genetics. Most notably, the sequence-indexed T-DNA and transposon mutants now cover insertions in 28,451 of 30,700 annotated genes (approximately 92% of all genes) including non protein-coding genes, of which null alleles constitute about 70% of all

Arabidopsis genes. The Salk Institute just started an AT2010 project to experimentally identify two homozygous insertion mutants for 25,000 *Arabidopsis* genes, thereby creating a "phenome-ready" genomics resource. AGRIKOLA, an EU funded project, offers a useful complement of 20,000 RNAi Gateway clones for the creation of knock-down lines.

3) Development of new tools for comparative genomics. In a project initiated by the Department of Molecular Biology at the Max Planck Institute for Developmental Biology (Detlef Weigel, PI), tiling arrays that cover the entire sequenced portion of the *Arabidopsis* genome with four 25mers per nucleotide on both strands (aggregate of one billion distinct oligonucleotides) are being used to interrogate 20 strains (ecotypes) of *Arabidopsis thaliana*. These 20 strains are a subset of the 96 strains analyzed by Bergelson, Kreitman and Nordborg as part of an AT 2010 project [http://walnut.usc.edu/2010.html]. The goal is to discover a large fraction of common SNPs as well as major deletions among these 20 strains. This project will create a data set unprecedented outside of humans. It will not only allow inferences about the microevolution of the *Arabidopsis thaliana* genome, but also enable large-scale association studies. The identification of ancestral SNPs exploiting the data mentioned above will be made possible by the announced sequencing by the DOE Joint Genome Initiative (JGI) of the closely related species, *Arabidopsis lyrata*. JGI will also sequence the more distantly related species, *Capsella rubella*, which should facilitate comparative genomics approaches to the identification of genes (particularly small ligands and microRNAs) as well as regulatory sequences.

4) The production of genome-wide sets of gene-specific probes for expression analysis and the establishment of reference transcriptomes. Functional analysis of *Arabidopsis* genes by free access to the rapidly growing genome-wide transcriptome data has been started by NASCArrays, the array facility and expression profile repository of the GARNet program. A major milestone in the establishment of community resources in *Arabidopsis* is the AtGenExpress reference transcriptome data set initially released in 2004 and still growing through contributions by various groups. This data set has been produced and compiled by a multinational initiative and covers genome-wide transcript profiles of all major

organs at various developmental stages, in response to diverse environmental stimuli and challenges (abiotic and biotic), and in response to different phytohormones. The data are publicly accessible and can be mined by different tools, databases and programs (Genevestigator, NASCArrays, TAIR, GEO, ArrayExpress, CSB.DB, Expression Angler etc.). This summer an *Arabidopsis* Affymetrix tiling chip covering the entire genomic sequence on two arrays (one chip covers the entire genome on 1 strand/ the second chip contains the entire reverse complementary strand of the whole genome) will be released.

5) Isolation of full-length cDNAs for about 70% of the genes, facilitating genome annotation and protein analyses. Resources for the characterization of the (biochemical) functions of the gene products have been vastly expanded: Full length cDNA clones of 13,386 genes are being distributed and almost 13,000 ORF clones useful for recombination cloning have been created and deposited at the *Arabidopsis* Biological Resource Center (ABRC). Beginning in May 2005, full length cDNAs in pUNI vectors (SALK/SS collection) will be transferred to pGATEWAY vectors by Invitrogen and made publicly available in the near future.

6) Initiation of working groups to establish multinational consortia for target areas such as proteomics, data base interoperability, and comparative genomics.

7) Appointment of a full-time MASC coordinator to foster information flow, international collaboration and coordination and to monitor progress of the program.

Some areas lag behind initial plans or are currently underrepresented:

1. Despite the long history of *Arabidopsis* as a model plant for genetics, we lack standardized ways of describing visible phenotypes (or for that matter the growth conditions that give rise to them). The recent development of standard ontologies for *Arabidopsis* anatomy and development holds promise that this gap can be filled in the near future. High-throughput phenotyping platforms are essential to fully leverage the potential of the fabulous reverse genetics and biodiversity resources now at our disposal.

2. The development of facile technology for heterologous protein expression has not yet been achieved for all proteins. Ongoing efforts to generate ORF clone collections suitable for recombination cloning are addressing this urgent need. Major projects on the elucidation of *Arabidopsis* protein structures and functions are ongoing.

3. Proteomics and comparative genomics need to be addressed with more emphasis.

4. In comparison with other model organisms, the physical interactions between macromolecules (protein-protein, protein-DNA, protein-RNA) are only patchily described in *Arabidopsis*, due mainly to technical difficulties with developing high-throughput systematic approaches to study such interactions in plants and to the problems with heterologous expression alluded to above.

Comprehensive interaction networks will be an essential prerequisite to understanding functions at a 'systems' level.

5. Although *Arabidopsis* databases were improved during the last year and offer access to rapidly growing sets of information, resources and tools, data integration and database interoperability needs to be established in new ways, The MASC Bioinformatics subcommittee held a workshop at TIGR in April 2005 to develop recommendations for data integration and database interoperability (for further information see the Bioinformatics subcommittee report).

6. Enabling resources for proteomics including production of antibodies against, or epitope tags on, all deduced proteins are needed. As a complement to the reference transcriptome, a catalogue of protein profiles at organ, cellular and subcellular levels under a wide range of environmental conditions should produce a reference proteome. To address the latter issue MASC is building a Proteome Working Group of experts in the field. The goal is to have an internationally coordinated initiative that develops strategies to unravel the *Arabidopsis* proteome.

7. For transcriptomics, proteomics and metabolomics, improved means of sampling at single cell resolution are needed.

Recommendations

Based on the analysis outlined above as well as direct community feedback to MASC and comments by the MASC subcommittees, MASC makes the following specific recommendations: A number of resources are urgently needed by the *Arabidopsis* community and their establishment as freely accessible materials and information should be given a high priority. These include

• A complete collection of verified homozygous knock-out lines and standardized descriptions of their phenotypes. Although this goal is addressed in a 2010 project (Ecker, SALK), broader community input is needed to complete the collection.

• Complementation of the comprehensive microarray repository by more experiments covering single cell analyses and several other experimental conditions that have not been investigated on the gene expresson level yet. Data need to be suitable for data mining.

• Resources including complete full-length cDNA/and ORFeome clone collections for studying protein localization and protein interaction networks, specifically comprehensive sets of tagged versions of ORFeome clones. Researchers expressed high interest in publicly available *Arabidopsis* libraries representing the full ORFeome for different two-hybrid systems.

• Development of high-throughput systematic approaches to study protein-protein, protein-DNA and protein-RNA interactions in plants. Arrays optimized for chromatin immunoprecipitation followed by microarray analysis (ChIP/chip) are a priority as well as generation of a comprehensive set of promoter-reporter lines

• Development of new means of data integration and database interoperability. Synchronisation of data syntax and semantics within different *Arabidopsis* derived data types but also with data

formats from other model organisms is an urgent need. The MASC Bioinformatics subcommittee is working on ways to find solutions for this complex goal and needs massive support by experts in the community.

- Improved metabolomics and proteomics resources
- Establishment of a reference proteome
- Establishment of a reference metabolome
- Co-financing of larger projects by co-operating funding agencies
- High-throughput genotyping methods to take advantage of the forthcoming comparative genomics information.

To achieve the long term goals of the program novel strategies need to be initiated soon to enable achievement of the desired level of in-depth knowledge on every *Arabidopsis* gene. Thus, systematic, high-throughput analyses of cellular networks including transcriptional, protein and metabolic networks need to be pursued. A key part of this "systems biology" approach will include protein-DNA interaction data, subcellular protein localization, protein-protein interactions, protein turnover rates, and protein modifications at cell-specific resolution.

A prerequisite for successful data mining is seamless access to all available information. This can only be achieved via well-connected high-performance databases that contain up-to-date information.

Progress and Activities of the Multinational *Arabidopsis* Steering Committee (MASC)

MASC activities in 2004/2005

A Workshop on Data Integration within the International *Arabidopsis* Community was organized by Chris Town and Heiko Schoof for the Bioinformatics subcommittee. The Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project Annual Report, 2004 made several recommendations intended to further the goals of the current international *Arabidopsis* functional genomics projects. One of these is: "Improved capabilities and integration of *Arabidopsis* database(s), with better ways to locate data and strongly enhanced mechanisms for import of data from individual researchers. In order to do simultaneous queries and analyses on large datasets, these finally have to be merged into one central database or a well-integrated network of databases." A workshop on data integration within the international *Arabidopsis* community, organized under the auspices of the Multinational *Arabidopsis* Steering Committee (MASC) and supported by the US National Science Foundation (NSF), took place at The Institute for Genomic Research (TIGR) on April 18-19, 2005. The purpose of the workshop was to respond to previous concerns, voiced by MASC, that better and more comprehensive data integration from both large and small data providers was required if the research community was to derive full benefit from the diverse resources and datasets being developed world-wide. The workshop brought together data providers and users to develop a strategy and propose solutions to address these issues. While some possibilities and experiences were presented, the main focus of the workshop was to articulate specific objectives and recommend strategies that could form the basis for discussion with the community, e.g. at the *Arabidopsis* conference June 2005 in Madison, and for efforts to implement these objectives. In addition to a roadmap of next steps, a number of self-identified working groups were set up to address more specific topics. For more specific information see the report of the Bioinformatics Subcommittee.

The full-time coordinator position is funded by the NSF for 2 years until the end of 2006. Isabell Witt moved from MPI Golm to Duke University to continue in this position. Tasks include communication within the international *Arabidopsis* community, within MASC, it's sub committees and working groups, with the funding agencies, other plant communities, major databases like TAIR, stock centers and several other groups. Monitoring progress of the functional genomics efforts, organizing the International Conference on *Arabidopsis* Research and overseeing the production of the annual MASC report belong to the core activities. The

major task for the time period after the *Arabidopsis* Conference is a major overhaul of the functional genomics web pages located at TAIR.

During the MASC business meeting in 2004 it became clear that two new MASC working groups are needed to address underdeveloped fields:
1) Proteomics (contact Wolfram Weckwerth)
2) Natural Diversity and Comparative Genomics (Tom Mitchell-Olds)

Both groups are in the final stages of establishment. The Natural Diversity and Comparative Genomics workshop (organized by Tom Mitchell-Olds) during the conference is a start up meeting for this working group.

Highlights of the Past Year

AtGenExpress follow up

Major insights from the evaluation of the AtGenExpress data sets are that the vast majority of annotated genes present on the ATH1 array are expressed. Hypothetical proteins are overrepresented in the non-expressed set, suggesting that many are not really genes. A very positive outcome was that all data generated at NASC, RZPD and MPI Tuebingen are overall very similar. Also, identical tissue types grown and collected at different sites but processed in Tuebingen all had very similar profiles, indicating that expression profiling with a common platform is more robust than anticipated. However, the RIKEN set is clearly distinct, reflecting a different scanner and different scanner settings. Because common samples with CAGE were included, the two projects can be cross-referenced in the future.

An initial analysis of global gene expression during *Arabidopsis* development in samples covering many stages from embryogenesis to senescence and diverse organs was published by the Weigel group* in Nature Genetics on data that are part of the AtGenExpress expression atlas. One general outcome is that transcription factor and signal transduction genes are expressed at a similar level as those of metabolic genes. Expresson patterns of large gene families are often more similar as expected by chance, suggesting that many gene families have been

*Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of Arabidopsis thaliana development. Nat Genet. 37(5):501-6. Epub 2005 Apr 3.

11

co-opted for specific developmental processes. Hypothetical genes were often found not to be expressed suggesting that many of them might not be genes or are expressed transiently or limited to few cells so that they could not be detected by the resolution of the experimental design.

A publication on the analysis of the whole AtGenExpress data set is in preparation.

Agrikola

To complement the efforts of saturating the *A. thaliana* genome with addressable insertion mutations, the EU-funded AGRIKOLA project (http://www.agrikola.org/) produced RNAi lines that cover the genome. So far, 20,000 hairpin constructs in pGATEWAY have been created. 3,000 of these have been introduced into *Agrobacterium* and 2,000 have been introduced in *Arabidopsis thaliana*. Preliminary analysis of the transformants indicates that (i) phenocopies of previously described knockout mutants can be obtained, (ii) viable 'knockdown' mutants of genes known to be essential can be obtained, (iii) the project reveals many informative phenotypes by inhibition of genes of currently unknown function and (iv) loss of phenotype in T2 or T3 is depending on the gene targeted and is not observed generally but rather exceptionally. The program has also started to produce inducible RNAi constructs that might help for certain difficult genes. Therefore the method can be considered as a very useful and robust complimentary tool to conventional and insertional mutagenesis. Further investigations will show at what level, mRNA or protein, the knockdown is most efficient. The AGRIKOLA source is so large that it will enable the community not only to learn a lot about gene functions but also about RNAi mechanisms in plants in general. All pGateway clones will be available from NASC in summer 2005.

Tiling Chips

This summer (2005) Affymetrix will release *Arabidopsis* tiling chips covering the full *Arabidopsis* genome sequence. Such oligonucleotide-based whole-genome microarrays are emerging as a preferred platform for genomic analysis beyond simple gene expression profiling. Uses include annotation of the transcriptome, chromatin-immunoprecipitation-chip studies, analysis of alternative splicing, characterization of the methylation state of the genome, polymorphism discovery and genotyping, comparative genome hybridization, and genome resequencing*. In a project initiated by the Department of Molecular Biology at the Max Planck Institute for Developmental Biology (Detlef Weigel), tiling arrays that cover the entire sequenced portion of the *Arabidopsis* genome with four 25mers per nucleotide on both strands (aggregate of one billion distinct oligonucleotides) are being used to interrogate 20 strains of *Arabidopsis thaliana*. These 20 strains are a subset of the 96 strains analyzed by Bergelson, Kreitman and Nordborg as part of an *Arabidopsis* 2010 project [http://walnut.usc.edu/2010.html]. The goal is to discover a large fraction of common SNPs as well as major deletions among these 20 strains. This project will create a data set unprecedented outside of humans. It will not only allow inferences about the microevolution of the *Arabidopsis thaliana* genome, but also enable large-scale association studies. The initial analysis will be performed as collaboration be-

tween the MPIs for Developmental Biology and Biological Cybernetics in Tübingen (Weigel and Schölkopf), the Center for Bioinformatics at the University Tübingen (Huson), the Salk Institute for Biological Studies (Ecker), and the USC Program in Molecular and Computational Biology (Nordborg). Data acquisition is ongoing and should be completed by fall 2005. The data will be released upon publication; at the same time, seed material of the exact 20 strains that have been used in the project will be deposited with the stock centers.

Measuring the Gene Function Knowledge

During the MASC annual meeting held during the 2003 conference, it was agreed that a better update on gene functions and quantification for how many genes the function is known for would be established. In the MASC annual report 2004 this was illustrated by thermometers. This year the thermometers are updated with data available at the end of April 2005.

Functional categories have been defined for easier quantification.

1. For genes that encode a protein
   - Protein activity/ Molecular function (catalytic or otherwise/ e.g., kinase, chaperone, phosphatase, proteinase) We should be using the GO ontology, trait ontologies for functional categorizations
   - Tertiary structure
   - Post-translational modification data
   - Expression pattern of protein at cell and tissue level
   - Subcellular localization
   - Protein interaction data
   - Phenotype of genetic knockout/ other loss-of-function alleles
   - Biological processes (e.g., photosynthesis, amino acid metabolism, cell wall biosynthesis)
2. For genes that do not code for a protein
   - Activity of RNA/ gene product
   - Expression pattern at cell and tissue level
   - Structure of RNA/ gene product
   - Subcellular location for RNA/ gene product
   - Phenotype of genetic knockout/ other loss-of-function alleles
   - Interaction data

The Gold Standard of a genes functional characterization is reached when we have full information for each of these categories. The opposite of the Gold Standard is "we don't know anything."
- Sequence has no homology to any sequence that we know the function of
- ORF has no expression
- no cDNA has been isolated, just predicted

*Mockler, T.C. and J.R. Ecker. 2005. Applications of DNA tiling arrays for whole genome analysis. Genomics 85:1-15.

12

Although it is evident that the Gold Standard is a very high standard, the final goal is that it will be applied to all genes. In the next six years, it should be possible to collect at least one category for every gene in the genome. As shown in the thermometers below, expression patterns for more than 80% of the genes can be extracted by the various expression profiling experiments, although not many at the cellular level. The progress made in each of the categories will be measured and illustrated with thermometers in the subsequent, annual MASC reports as well.

For this year's report, Eva Huala and colleagues from TAIR and Hank Wu from TIGR provided the actual numbers of genes that fall into different evidence codes, genes for which there are fullength c-DNAs, genes that have been detected in various expression profilings, and genes for which there was experimental evidence in the literature for a function. Information for the thermometers was also obtained through the *Arabidopsis* community. A questionnaire was sent to 2010/AFGN researchers by the MASC coordinator about the categories listed under 1.: functional categories for genes that code for proteins. Forty nine 2010/AFGN projects supplied data which were forwarded to TAIR and filtered for redundancy with other data. New non-redundant information was integrated into the thermometer and called community input (CIP).
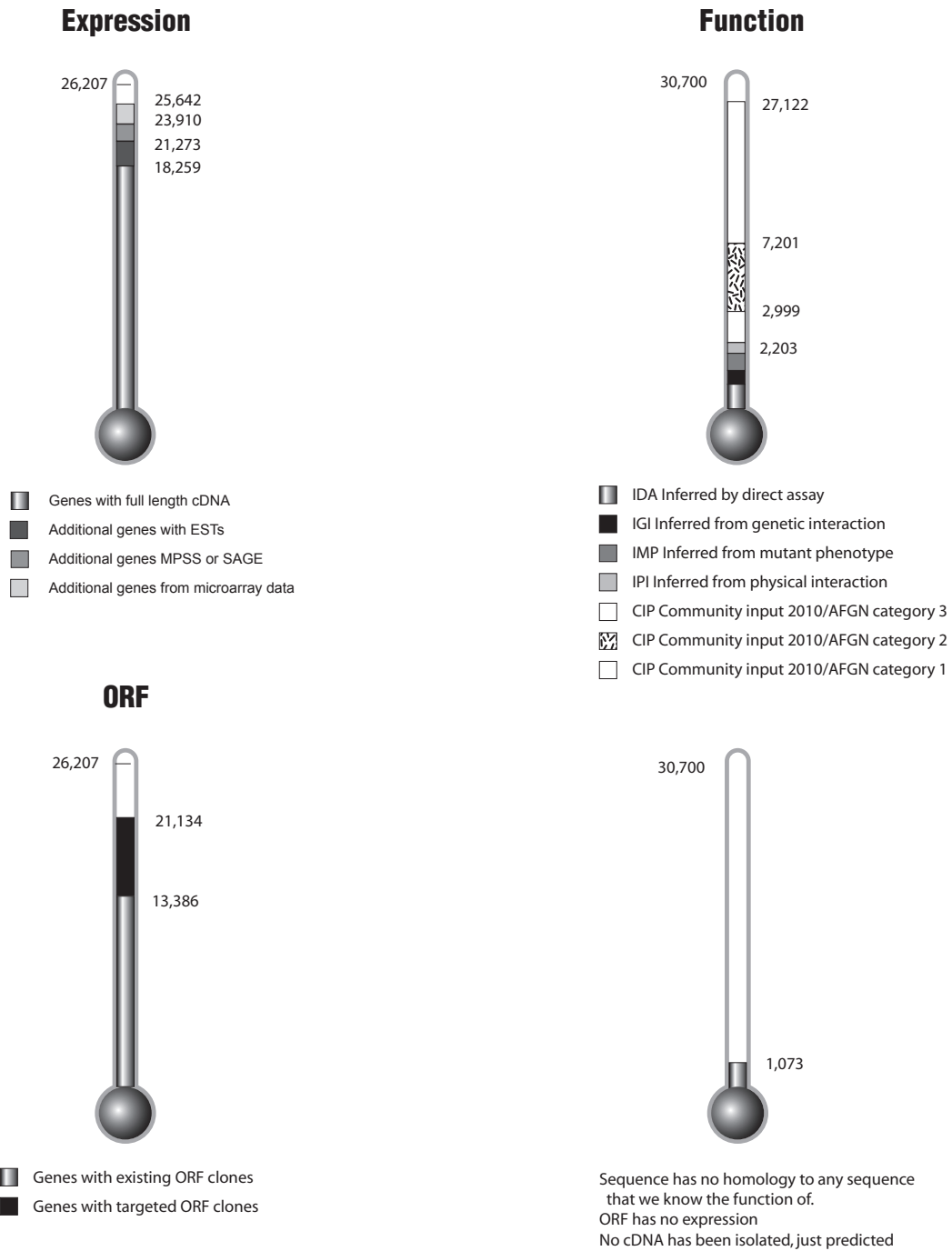
Expression

26,207 —
25,642
23,910
21,273
18,259

■ Genes with full length cDNA
■ Additional genes with ESTs
■ Additional genes MPSS or SAGE
■ Additional genes from microarray data

Function

30,700
27,122

7,201

2,999

2,203

■ IDA Inferred by direct assay
■ IGI Inferred from genetic interaction
■ IMP Inferred from mutant phenotype
■ IPI Inferred from physical interaction
□ CIP Community input 2010/AFGN category 3
▨ CIP Community input 2010/AFGN category 2
□ CIP Community input 2010/AFGN category 1

ORF

26,207 —
21,134

13,386

■ Genes with existing ORF clones
■ Genes with targeted ORF clones

30,700

1,073

Sequence has no homology to any sequence
  that we know the function of.
ORF has no expression
No cDNA has been isolated, just predicted

Figure 1: Measures of knowledge on *Arabidopsis* genes. Exact numbers for the different categories are as follows: Gene expression: genes with full length cDNA (18,259), additional genes with ESTs (3,014), additional genes from MPSS or SAGE (2,637), additional genes from microarray data (1,732). Genes with existing ORF clones (13,386), additional genes with targeted ORFs for cloning* (7,748). Gene function (includes loci annotated to GO function, process or component with the listed evidence codes): IDA (1589), additional IGI (173), additional IMP (377), additional IPI (64), CIP class 3 is when a characterization is almost finished (796), CIP class 2, additional genes have been partially characterized (4,202) and CIP class 1, additional genes were selected for characterization but have not been characterized yet (19,921). Please note that gene accessions were compared for redundancies. The numbers in each thermometer refer to non-redundant gene accessions. The total number of genes where the sequence has no homology to any sequence that we know the function of, ORF has no expression, no cDNA has been isolated, just predicted (1,073).

* genes targeted by Ecker's group (list provided to TIGR), CESG, Wisconsin (list provided to TIGR), ORPHEUS Group (data extracted from ORPHEUS DB, Ghent), Atome (taken from their web site), TIGR, and a few other minor sources.

14

Reports of the MASC Subcommittees

Bioinformatics

Prepared by Chris Town (Chair) and Heiko Schoof
MASC Bioinformatics Subcommittee

Last year's report included the committee recommendation that "improved capabilities and integration of *Arabidopsis* database(s), with better ways to locate data and strongly enhanced mechanisms for import of data from individual researchers are needed to take full advantage of the rapidly increasing amount of data from diverse sources" and that "in order to do simultaneous queries and analyses on large datasets, these finally have to be merged into one central database or a well-integrated network of databases."

To address these issues, a workshop organized by members of the bioinformatics subcommittee and supported by the US National Science Foundation was held at The Institute for Genomic Research (TIGR) on April 18-19, 2005. The workshop was attended by representatives of MASC[1], NAASC[2] and wet lab-oriented biologists as users, together with data providers, database hosts such as TAIR[3], application developers, data integration projects such as BioMoby[4] and PlaNet[5], and funding agencies including the NSF, the European Commission and the USDA. For a complete list of attendees please see the complete report at http://www.arabidopsis.org/info/2010_projects/MASC_Info.jsp. The goal was to discuss these issues and formulate specific objectives and recommended strategies that could form the basis for dialog with the community, e.g. at the *Arabidopsis* conference June 2005 in Madison. ,

The workshop began with a review of needs as perceived by data users and resources and possible future solutions for data management as perceived by database providers and developers including TAIR, PlaNet[6], DAS[7], BioMoby and PlaNet. In the course of this discussion, a list of data types currently being handled and generated was compiled. This showed that the complexity of the integration problem has been continually increasing in recent years, and while there are many encouraging advances in integrating sequence and related data, expression data is proving to be more difficult. Even more complications are expected for more qualitative data such as phenotypes and protein interactions. The following key questions were discussed:

- How do we enable high-throughput data integration?

While data can be integrated from multiple sources manually, this is not feasible given the large scale of current data sets.

- How do we ensure availability of high-quality data?

While enabling widespread availability of data what can or should be done to ensure that the data are reliable and of high quality?

Further discussion between computer science and database or application developers, users and data generators about the requirements for data integration identified the following goals:

1. User interface
- Simple search "one-stop-shop" interfaces.
- Multiple, domain- and problem-specific interfaces and views of the data.
- More user-friendly, powerful and versatile visualization tools.
- Intuitive user interfaces developed according to human-computer-interface standards.
- The ability to build and use data analysis pipelines that connect several tools without the need to invoke each step individually.
- The ability to create and manage groups or sets of data, e.g. to retrieve data from analysis and visualization tools, like a selected set of proteins, and to input these into further tools.
- The ability to save and share work sessions.
- A common user management and login for multiple sites.
- Integration of phylogenetic data and tools, such as orthologs/paralogs and phylogenetic trees, with the gene/genome data.
2. Data formats/Syntax
- Standards for data formats (common syntax).
- Standard operating procedures, e.g. for GenBank submissions, to ensure that metadata is entered consistently (e.g. "sequence class" for mRNA-derived sequences).
3. Data content/Semantics
- Standardized vocabularies, ontologies, and terms, as well as defined usage of terms and tokens so as to build a common semantics, e.g. mapping different fields corresponding to a "gene name" across varied data schemas to the shared concept of a "gene name."

- Coordinated naming, ensuring uniqueness and lack of ambiguity, and the use of generally recognized and accepted names e.g. in publications.
4. Data Quality
- Current, detailed and comprehensive annotation, including knowledge collected from literature.
- High-quality data resources and long-term archives
- Manual curation, e.g. of links between different sets of data and of synonym lists.
- Provenance information, experiment descriptors and evidence metadata linked to all data.
- Education about and dissemination and enforcement of analysis standards.
5. Data and Database Accessibility
- Availability of downloadable, integrated core datasets.
- Common interoperability standards and programmer interfaces to enable consolidation of development efforts.
- Public, machine-discernible interfaces to data and analysis resources.
- A one-stop-shop, so that users don't have to search for and visit a number of different sites on the web and compile the information themselves, but can access all information from a single entry point. This need not mean that all information is physically located in a single database, or accessible through a single web site: alternative user interfaces are of value to end users who can select an interface appropriate to their problem. However, the user should be able to navigate to all data from the entry point. Integration through web links was considered an initial solution for this, but better data integration strategies should be explored.
- A data discovery service.
6. Others
- Simplified update of information and capture of community knowledge.
- An easily extensible and open architecture, while at the same time preventing confusion that might be caused by illicit data sources hosting nonsensical or wrong data.
- Synchronization of core data like the genome sequence/assembly between data providers.
- Documentation and tutorials.
- While the focus of the workshop was on data integration the goals above also recognize issues of broad concern regarding data generation, data mining and the kinds of tools and data sets that the community might need.

To achieve these goals, service oriented architectures such as those used by Biomoby, PlaNet or Toolbus[8] were compared to integrated data warehouses such as TAIR and PlantGDB. It is clear that no single approach will answer all requirements. For example, while on the one hand, centralized curation of data can ensure quality, on the other hand, flexibility, extensibility, comprehensiveness and diversity can be more efficiently realized in a semantically aware, service-based, distributed system.

Problems specific to particular domains of data were addressed, and served to illustrate fundamental prerequisites for comparability or integration of data. For example, the necessity of controlled and carefully described experimental conditions for comparing microarray data from expression analysis, or the usefulness of a universal and unique identifier for microarray experiments were discussed. This illustrates that data integration is not an isolated topic, but must be coordinated with the experimenters, large-scale data generators and users, as some prerequisites already need to be met before the data are generated. As a result, communication with relevant groups, initiated through the MASC subcommittees, was proposed.

The group summarized the future objectives of this MASC-based data integration initiative as follows. To work towards more widespread and comprehensive data integration it is necessary to

- Ensure easy access to a comprehensive, integrated and high quality core data set
- Establish links between different data types
- Develop analysis pipelines and simplify access to them
- Facilitate generation of specialized (custom) datasets and combinatorial queries on data
- Develop and provide visualization and analysis tools
- In general, progress the field technologically through the levels of interconnectivity, interoperability, and towards eventual integration.

To achieve the above objectives, the group makes the following recommendations:

1. Establish standardized semantics (vocabularies, tokens (e.g. <tags>), ontologies, names/identifiers, etc.) for data representation in several realms:
- DNA sequence annotation,
- RNA properties
- protein properties
- metabolome
- phenotype
- evolutionary relationships
Examples: AGI Locus Code, Gene name/symbol, GO_term, expression experiment ID. This should be done in collaboration with data producers, data providers and analysts.
2. Propose standard data exchange formats for the above data types. If possible, use existing standards, and promote adoption of these standards by the community at large.
3. Support existing efforts to create controlled vocabularies/ontologies to describe e.g. plant anatomy (PO), phenotypes (PATO), biochemical function (in more depth than GO?) and localization (GO), sequence annotation (SO).
4. Create and promote the use of experiment and evidence descriptors to attach provenance information to the above data types for comparability and quality assessment purposes.

5. Form a working group comprised of both informatic and bio-logical expertise (database developers, providers and users) to explore, assess and recommend technology for database interoperability.
6. Provide training, tutorials and documentation to both bioinformaticians and biologists to facilitate *Arabidopsis* data production, mining, analysis and integration.

The following mechanisms were proposed to work towards these goals:

•   Tutorial workshop at the *Arabidopsis* conference
•   Tutorial documents on Web pages
•   White paper (to be widely disseminated – perhaps published)
•   Working groups on data types and formats to be developed in collaboration with the MASC subcommittees
•   A database interoperability and data integration workshop at the end of 2006 that will bring together people actively involved in the efforts recommended by this workshop.

(Footnotes)

1. Multinational *Arabidopsis* Steering Committee
2. North American *Arabidopsis* Steering Committee
3 The *Arabidopsis* Information Resource: www.arabidopsis.org
4. biomoby.org
5. www.eu-plant-genome.net
6. www.plantgdb.org
7. Distributed Annotation System, biodas.org
8. http://pathport.vbi.vt.edu/download/

cDNAs and Clone-Based

Functional Proteomics (ORFeomics)
Chair: Pierre Hilson
MASC cDNAs and Clone-Based Functional Proteomics
(ORFeomics) Subcommittee

Collectively, the *Arabidopsis* community has now gathered full-length (fl) cDNA sequence information for about 18,259 of the 26,207 protein-encoding genes, excluding transposable elements and pseudo-genes, identified by TAIR and TIGR May 2005. This experimental confirmation of gene models is crucial for a high quality annotation because, in many cases, the predicted models are corrected by the actual transcript sequences and, in other cases, some transcription units are simply not predicted at all. However, the isolation of novel fl cDNA clones becomes more laborious as it focuses increasingly on genes expressed at low level, in particular conditions or in specific cell types. Consequently, alternative approaches are welcome at this stage of the genome structural annotation. Such an example is the use of transcript profiling tiling arrays or the systematic RTPCR amplification and sequencing of cDNAs based on predicted gene models. New methods to intentionally capture cDNAs originating from uncharacterized transcription units will soon be necessary as the fraction of genes lacking experimental expressed sequence data narrows down.

Fl cDNA clones are not only important for genome annotation. They also constitute crucial reagents for the functional analysis of protein-encoding genes. In this respect, major projects have already resulted in the construction of open reading frame (ORF) collections that can be transferred at large-scale via recombinational cloning techniques from a reference clone to a wide-variety of expression vectors, each designed for a specific functional assay. These ORF collections are, or soon will be, publicly available (see Fl cDNA and ORF table). They will undoubtedly foster research projects that either focus on the analysis of selected gene subsets with various methods or on the systematic genome-scale characterization of certain protein properties. Because different applications dictate incompatible sequence constraints (ORF with or without stop codon including or not terminal tags) the ORF collection format cannot be unique and settled once and for all. However, the community will greatly benefit from the new source of pGATEWAY ORF clones that is produced by Invitrogen by transferring all pUNI ORF clones to this system and make it publicly available. Very useful would be a centralized database that would inform all potential users of the status of the cDNA/ORF cloning and sequencing progress for their genes of interest in any of the ORFeome projects, together with the restrictions that may or may not apply to their use. Obviously, less or no restriction is preferable to boost the *Arabidopsis* research community.

See "Table 1. Ongoing large-scale fl cDNA and ORF projects" on the next page.

Table1: Large-scale fl cDNA and ORF projects/resources

Institution	Project name	Project Goal	Sequences Completed	Sequences Deposited at Data Base	Clones deposited at Stock Center	Restrictions/Costs*
RIKEN, Genomic Sciences Center (GSC), http://pfgweb.gsc.riken.go.jp/projects/raflcdna.html, http://rarge.gsc.riken.go.jp/, PI: Motoaki Seki, Kazuo Shinozaki[1]	RIKEN GSC project,SSP Consortium, National BioResource Project(Japan)	determine the full-length sequences of 18,000 RAFL cDNA clones	15240 fully sequenced RAFL cDNA clones	15,240 full reading, 246,640 single-pass reading. Their data are available at Genbank.	151593 RAFL clones whose single-pass sequences have been determined are available at BRC	1. Clones available from RIKEN BRC: MTA with BRC (few restrictions). Similar costs as ABRC. 2. Clones not available from BRC: MTA with GSC. No costs.
Salk SIGnAL, http://signal.salk.edu/cdnastatus.html, PI: Joseph Ecker	SSP Consortium, 2000-2003 (completed)	10.500 fl CDNA clones	10556 fully sequenced ORF clones tailored for recombinational cloning	10556 at Genbank	1841Gateway clones/8659 pUNI clones at ABRC	No restrictions
Salk SIGnAL, http://signal.salk.edu/cdnastatus.html, PI: Joseph Ecker	Salk NSF2010 2003-2005 c-DNA sequencing project	experimental verification of the annotation of additional 3,330 genes by full-length cDNA sequencing and the construction/sequencing of 3,333 additional ORF clones for new transcriptional units. 3072 fl cDNAs to be constructed from EST and GSLT data	1035 Salk fl cDNAs	1035 Salk fl cDNAs at Genbank	1013 Salk fl cDNA clones at ABRC in pUNI	No restrictions
Salk SIGnAL, http://signal.salk.edu/cdnastatus.html, PI: Joseph Ecker	Salk NSF2010 2003-2005 c-DNA sequencing project	Gold Standard: "All SSP and SALK ORFs are fully sequenced to high quality and when conceptually translated show a perfect match to the corresponding protein". 1174 gold standard ORFs to be constructed from RAFL data.	1550 gold standard ORFs	1550 gold standard ORFs deposited at Genbank	1508 gold standard ORF clones deposited at ABRC in pUNI	No restrictions
TIGR, http://www.tigr.org/tdb/hypos/, PI: Christopher Town	TIGR 2010 cDNA	2,000 transcripts for which no experimental cDNA sequences are available	full or partial cDNA sequences from 1563 genes	478 deposited at Genbank	480 pDONR221 clones deposited at ABRC	No restrictions
Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, Peking University and Yale University, http://datf.cbi.pku.edu.cn/, PI: Xing Wang Deng[2,3], Yuxian Zhu[2,3]	Peking-Yale consortium	Full-length ORFeome clones for the transcription factors, in GATEWAY pENTRY vector, with end-sequence confirmed	ORFs of 1282 Arabidopsis transcription factors	Only clones with different sequences than reported deposited in genbank	ORFs of 1282 Arabidopsis transcription factors in GATEWAY entry vectors deposited at ABRC	No restrictions

1. Seki et al.(2002) Science 296: 141-145
2. Wei Gong, et al., Genome wide ORFeome cloning and analysis of Arabidopsis transcription factor genes. Plant Physiol. 135; 773 (2004).
3. Anyuan Guo, Kun He, Di Liu, Shunong Bai, Xiaocheng Gu, Liping Wei*, & Jingchu Luo* DATF: a Database of Arabidopsis Transcription Factors. Bioinformatics, 2005 in press
* If not indicated differently costs are only the stock center fees for maintenance and shipping.

Multiparallel Analytical Tools

Prepared by Mike Beale and Mary Lou Guerinot

The good news is there is more and more data available to be analyzed. The bad news is there is more and more data available to be analyzed! NASC arrays now has over 2,450, and AtGenExpress more than 1,340 new Affymetrix baseline data sets for a number of developmental stages (see Schmid et al., 2005), tissue types and stereotyped challenges. NASCArrays microarray data are freely available on the web http://affy.arabidopsis.info. Here one can find spot histories, two-gene scatterplots across all experiments, gene swinger, subset bulk gene downloader and other tools including ExpressionProfiler friendly files for clustering. In the past year, TAIR has also added gene expression profiling data from 1414 Affymetrix slides, including those of the AtGenExpress project: http://www.arabidopsis.org/servlets/Search?action=new_search&type=expression.

User interfaces for accessing microarray data have been revamped to improve usability. Additional tools, programs and databases that enable the community to access the NASCArray and AtGenExpress datasets are:

- Genevestigator (>1800 ATH1 arrays, 140 experiments) https://www.genevestigator.ethz.ch/
- Expression Angler (NASCArray data, AtGenExpress developmental Dataset) http://bbc.botany.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi
- CSB.DB (NASCArray data, AtGenExpress developmental, abiotic stress data) http://csbdb.mpimp-golm.mpg.de/
- GABI MapMan https://gabi.rzpd.de/projects/MapMan/data.shtml
- AMPL http://www.cbs.umn.edu/arabidopsis/
- ARACyc http://www.arabidopsis.org:1555/expression.html

This summer (2005) Affymetrix will release Arabidopsis tiling chips covering the full *Arabidopsis* genome sequence. Such oligonucleotide-based whole-genome microarrays are emerging as a preferred platform for genomic analysis beyond simple gene expression profiling. Uses include annotation of the transcriptome, chromatin-immunoprecipitation-chip studies, analysis of alternative splicing, characterization of the methylation state of the genome, polymorphism discovery and genotyping, comparative genome hybridization, and genome resequencing (See Mockler and Ecker, 2005 for an overview of whole genome arrays). The Ecker group at the Salk Institute provided all the "platform" information (sequences, xy coordinates etc.) to the NIH GEO database. GPL1979 and GPL1980 are the *Arabidopsis* tiling chips. http://www.ncbi.nlm.nih.gov/projects/geo/query/browse.cgi?view=platforms.

In a project initiated by the Department of Molecular Biology at the Max Planck Institute for Developmental Biology (Detlef Weigel), tiling arrays that cover the entire sequenced portion of the Arabidopsis genome with four 25mers per nucleotide on both strands (aggregate of one billion distinct oligonucleotides) are being used to interrogate 20 strains of *Arabidopsis thaliana.* These 20 strains are a subset of the 96 strains analyzed by Bergelson, Kreitman and Nordborg as part of an Arabidopsis 2010 project [http://walnut.usc.edu/2010.html]. The goal is to discover a large fraction of common SNPs as well as major deletions among these 20 strains. This project will create a data set unprecedented outside of humans. It will not only allow inferences about the microevolution of the Arabidopsis thaliana genome, but also enable large-scale association studies. The initial analysis will be performed as collaboration between the MPIs for Developmental Biology and Biological Cybernetics in Tübingen (Weigel and Schölkopf), the Center for Bioinformatics at the University Tübingen (Huson), the Salk Institute for Biological Studies (Ecker), and the USC Program in Molecular and Computational Biology (Nordborg). Data acquisition is ongoing and should be completed by fall 2005. The data will be released upon publication; at the same time, seed material of the exact 20 strains that have been used in the project will be deposited with the stock centers.

One challenge of course is integrating all the different types of information. The AREX (ARabidopsis gene Expression) is one example of a database that is set up to collect Arabidopsis gene expression data from various sources (microarrays, in situs, promoter::reporter constructs etc.) into a single searchable database (http:www.arexdb.org). Although currently the database is limited to root gene expression data, the goal is to extend it to other plant organs. All interested laboratories can enter data and decide whether they will restrict access to their own laboratory, to certain laboratories, or open their data to the general public. The database will therefore be useful not only for data sharing but also for different laboratories to easily store and search their own data.

Phenotypes

Since digital images of plants are being archived in an increasing number of projects, it is essential that a controlled vocabulary be used when describing such images. TAIR plans to begin using PATO (for attribute and value) along with the GO vocabularies and the Plant Ontology terms for plant structure and developmental stage to begin phenotype annotation at the end of 2005. For more information on PATO see the SourceForge OBO site (http://obo.sourceforge.net/cgi-bin/detail.cgi?poav). There is also a mailing list for interested parties: obo-phenotype@lists.sourceforge.net. NASC has implemented a Plant Ontology browser to search for phenotypes and mutants. NASC has a plant picture library of 600 photographs on this web page: http://seeds.nottingham.ac.uk/Nasc/action.lasso?-token.user=18513023411&-response=/Nasc/picbook/picture_book.lasso. With a program developed by INRIA http://www-rocq.inria.fr/cgi-bin/imedia/ikona/ a vocabulary free analysis tool is also provided.

Metabolomics

The goal proving to be the most difficult to achieve is to simultaneously quantify all of the metabolites at the cell, organ or plant level. Traditional analytical chemistry based on chromatographic separation of metabolites and subsequent identification by techniques such as GC-MS and LC-MS has played an important role in opening up this area. Most work published so far utilizes these techniques to profile crude plant extracts or to home in on particular classes of compounds in purified extracts. Recently, the application of 'fingerprinting' to unchromato-

graphed extracts by NMR and direct injection ESI-MS or FT-ICR-MS have proved to be promising techniques. They are perceived as a way forward for high-throughput mass screening of mutants and natural variants. Much Arabidopsis metabolomics is being pursued in the private sector. Service and large-scale activities in publicly funded Arabidopsis metabolomics are less prevalent than the other 'omics'. Nevertheless, the UK GARNet project contains an Arabidopsis metabolomics service and activities in Arabidopsis metabolomics are also beginning to emerge in the Netherlands (http://www.biosystemsgenomics.nl/) and in Sweden (http://wcn.ntech.se/platforms/Metabolomics.htm). In addition, the MeT-RO project (Metabolomics at Rothamsted) is a newly funded initiative in the UK which has built on the GARNet project to establish a National Centre for Plant and Microbial Metabolomics has opened metabolite profiling services to the plant community http://www.metabolomics.bbsrc.ac.uk. The plant metabolomic community holds an International Congress (Potsdam, 2003; Iowa State, 2004, Reading 2006) and has formed a platform to further international discussions http://www.metabolomics.nl. The integration of metabolomics data with other functional genomic data is a difficult goal to achieve and concerns the community. Pathway databases (http://www.arabidopsis.org/tools/aracyc) and (http://www.genome.ad.jp/kegg/pathway.html) are being developed but they do not yet contain metabolomic datasets. Problems of alignment of datasets, databases and effective query tools are still being researched. However, software solutions to some of the problems are emerging (see for example metAlign at http://www.plant.wageningen-ur.nl/default.asp?section=products). The Metablomics Society formed to further metabolomics research, including plants http://www.metabolomicssociety.org/. The 2nd phase of UK (BBSRC) initiative to fund plant and microbial metabolomics was launched. MeTNetDB, a resource for mapping metabolite, expression profiling and proteome data has been developed under an *Arabidopsis* 2010 pilot project http://www.public.iastate.edu/~mash/MetNet/.

References

Mockler, T.C. and J.R. Ecker. 2005. Applications of DNA tiling arrays for whole genome analysis. Genomics 85:1-15.

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU. 2005. A gene expression map of *Arabidopsis thaliana* development. Nat Genet. 37(5):501-6. Epub 2005 Apr 3.

Reverse and Forward Genetic Stocks

Chair: Bernd Weisshaar
MASC Reverse and Forward Genetic Stocks Subcommittee

Fast and reliable access to mutants in selected genes is crucial for systematic reverse genetic approaches. The MASC Forward and Reverse Genetic Stocks subcommittee addresses issues of coordination and communication among the existing projects in this field. The next meeting will take place during the International Conference on *Arabidopsis* Research in 2005.

The integration and data exchange between the various projects has progressed well. Most providers of flanking sequence tag (FST)-based mutant collections do allow access to the primary FST sequence information, either from their web sites or via GenBank. In addition, the discussion on, for example, what constitutes a "potential FST gene hit" has resulted in more detailed evaluation and annotation of FSTs in terms of the location of the insertion within a given gene. Analysis of the current resources has shown that coverage of the *Arabidopsis thaliana* gene inventory with knockout mutations is already impressive, but it also demonstrates that the number of really useful insertion lines, namely those that are likely to be a null mutation, is still not saturating. If one considers insertions in coding exons as good candidates for NULL alleles, then there are about 22.647 different genes covered with insertions. However, it is clear that alleles with insertions at the end of the ORF may still result in a (partially) functional protein. Insertions in introns are also often good candidates for NULL alleles (24,654 insertions in exons and introns). In conclusion, we can argue that for more than 70% of all *Arabidopsis thaliana* genes useful NULL alleles are available. On the other hand out of 30,700 total genes 2,249 genes were never hit by an insertion taking all resources into account (SALK, SAIL, Wisc, FLAG, RIKEN, SM, IMA, GABI-KAT and CSHL lines), this information is of April 26, 2005 by Huaming Chen and Paul Shinn (Ecker lab at SALK). The Salk Institute Genomic Analysis Laboratory (SIGnAL, http://signal.salk.edu/cgi-bin/tdnaexpress) has integrated FST data from GABI-KAT, SAIL and FLAGdb as well as data from RIKEN, Wisconsin and several other FST resources into their TDNA express database. As a result, a quite comprehensive collection of sequence-indexed T-DNA insertion mutants can be searched at a single location on the basis of FST sequence information. This allows users of reverse genetic resources a "one-stop" access to almost all available information on T-DNA insertions in a given gene. The "*Arabidopsis* Knockout Facility" at the University of Wisconsin-Madison has announced the availability of a new collection of T-DNA lines containing Ds-Lox launching pads and Cre/Lox recombination sites (see http://www.hort.wisc.edu/krysan/DS-Lox/). The specific features of these lines can be used to delete tandemly duplicated gene family members, or to generate insertion mutants at flanking loci that are not covered by currently available T-DNA collections. The regions flanking the insertions of more than 10,000 of these lines have been sequenced and can be located on the map of **SIGnAL.** The sequences are being incorporated into TAIR. To complement the efforts of saturating the *A. thaliana* genome with addressable insertion mutations, the EU-funded AGRIKOLA project (http://www.agrikola.org/) produced RNAi lines that cover the genome. So far, 20,000 hairpin constructs in pGATEWAY have been created. 3,000 of these have been introduced into *Agrobacterium* and 2,000 have been introduced in *Arabidopsis thaliana*. Preliminary analysis of the transformants indicates that (i) phenocopies of previously described knockout mutants can be obtained, (ii) viable 'knockdown' mutants of genes known to be essential can be obtained, (iii) the project reveals many informative phenotypes by inhibition of genes of currently unknown function and (iv)

loss of phenotype in T2 or T3 is depending on the gene targeted and is not observed generally but rather exceptionally. The program has also started to produce inducible RNAi constructs that might help for certain difficult genes. Therefore the method can be considered as a very useful and robust complimentary tool to conventional and insertional mutagenesis. Further investigations will show at what level, mRNA or protein, the knockdown is most efficient. The AGRIKOLA source is so large that it will enable the community not only to learn a lot about gene functions but also about RNAi mechanisms in plants in general. All pGateway clones will be available from NASC by summer 2005.

Also, projects based on TILLING (http://www.*arabidopsis*.org/abrc/ henikoff.jsp) allow access to additional mutations, including change-of-function alleles of a given gene. A comprehensive summary of forward genetic stocks, recombinant inbred line (RIL) populations and other such resources is available at http://www.inra.fr/qtlat/NaturalVar/RILSummary.htm. At the moment, nine RIL populations are available as seed stocks from the public stock centers, but more than 56 different RIL populations and two populations of genetic substitution lines (nearly isogenic lines, NILs) are presently being established.

Single Nucleotide Polymorphisms (SNPs) detection is of increasing importance in the forward genetics tool kit. Several large SNP collections are available through TAIR, including those of Cereon/Monsanto (approximately 37,500 SNPs), the Stanford Genome Technology Center

(at least 11,000 SNPs) and GABI-MASC ("MASC" stands here for the Max-Planck *Arabidopsis* SNP Consortium; over 8,000 SNPs; these are also available via http://www.mpizkoeln.mpg.de/masc/). With the exception of the GABI-MASC SNPs, which were obtained by re-sequencing between 6 and 12 accessions, the SNPs above were identified as a difference between a single accession (typically Ler) and the reference genome. Thus, little is known about their frequency in other accessions. In contrast, over 17,000 polymorphisms obtained through re-sequencing of 1.500 short fragments of 96 accessions are available through http://walnut.usc.edu/2010 , and most of them are available through TAIR as well. The 96 accessions, which include many of those being used to generate RILs are available as a set from the stock centers. A further development of this source was established by Norman Warthmann (Weigel group MPI Tuebingen) together with Genaissance Pharmaceuticals (identification of 300 SNPs in comparison to columbia). SNP assays for about 180-250 markers will work for any cross to columbia, while 50 to 70 markers will still work for any other cross. Community value: The assays are developed and paid for and the *Arabidopsis* scientist is invited to submit DNA to Genaissance and pay only the genotyping cost, which will vary depending on the size of the project (markers x lines). Contact: Don Benson (d.benson@genaissance.com), Min Seob Lee (m.lee@genaissance.com). The SNP set is very soon available from TAIR.

The International *Arabidopsis* Functional Genomics Community

Argentina

Several university-associated groups are actively engaged in *Arabidopsis* research in Argentina. Funding for *Arabidopsis* research is available from the organizations listed below:

- Analysis of transcriptome in plant-pathogen interactions: plant genes required for susceptibility to fungal infection.

Malena Alvarez,
malena@dqb.fcq.unc.edu.ar
CIQUIBIC-CONICET
Facultad Ciencias Quimicas, Universidad Nacional de Cordoba. Province of Córdoba
http://www.fcq.unc.edu.ar/ciquibic

- The genetic network involved in plant responses to the light environment, analysis of transcriptome in phytochrome and cryptochrome mutants. This group has produced a set of RILs between Landsberg *erecta* and Nossen in collaboration with the groups of Allan Lloyd (University of Texas) and Javier Botto (University of Buenos Aires), which will become available at ABRC.

Jorge J. Casal,
casal@ifeva.edu.ar
IFEVA, Facultad de Agronomía, Universidad de Buenos Aires. Buenos Aires
http://www.ifeva.edu.ar/staff/perpages/casal.htm

- Functional analysis of genes involved in the biogenesis of the cytochrome c-dependent respiratory chain

Daniel H. Gonzalez,
dhgonza@fbcb.unl.edu.ar
Facultad de Bioquímica y Ciencias Biológicas
Universidad Nacional del Litoral. Province of Santa Fe

- Role of senescence associated genes in the formation of lytic vacuoles during senescence.

Juan José Guiamet,
jguiamet@museo.fcnym.unlp.edu.ar
Instituto de Fisiología Vegetal
Universidad de La Plata.
Province of Buenos Aires

- Genes involved in Potassium and Sodium transport.

Guillermo E. Santa-Maria
gsantama@pop.unsam.edu.ar
Instituto de Investigaciones Bioteconológicas, Universidad Nacional de San Martin
Province of Buenos Aires

- Regulatory genes involved in the control of transcription of genes of the photosynthetic antenna

Roberto J. Staneloni
RStaneloni@Leloir.org.ar
Instituto Leloir
Buenos Aires

- Functional analysis of oxidative stress-regulated genes

Estela M. Valle
evalle@fbioyf.unr.edu.ar
Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), Facultad Ciencias Bioquimicas y Farmaceuticas, Universidad Nacional de Rosario. Province of Santa Fe

- Identification of key components for retrograde signaling between mitochondria and nucleus in higher plants by transcriptomic, proteomic and functional analyses of respiratory complex mutants in *Arabidopsis*

Eduardo Zabaleta (ezabalet@mdp.edu.ar)
Universidad de Mar del Plata. Province of Buenos Aires

- Regulatory genes involved in the biogénesis of mitochondrial Fe-S proteins. Metabolic analysis of *Arabidopsis* mutants defficient in enzymes involved in carbon metabolism.

Diego Gómez-Casati (diego.gomezcasati@intech.gov.ar)
Instituto de Investigaciones Bioteconológicas, Universidad Nacional de San Martin, Province of Buenos Aires.

- The main sources of financial support are:

ANPCYT (Agencia Nacional de Promoción Científica y Técnológica).
CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas)
FUNDACION ANTORCHAS (Argentina)
FIRCA (NIH)
TWAS (Third World Academy of Sciences)

First Affymetrix workstation in Latin America goes to *Arabidopsis* research groups ANPCYT has granted an Affimetrix workstation to a consortium integrated mainly by research groups listed above. The equipment will be installed at IFEVA.

Australia & New Zealand

Australia has a strong tradition in plant scientific research. Many institutions are engaged in *Arabidopsis* Functional Genomics work including the Plant Industry Division of the Commonwealth Scientific and Industrial Research Organization (CSIRO), the major Universities and private enterprise. Their work ranges from individual projects to international collaborations and major resource development. Funding is mainly available through the Australian Research Council's (ARC's) Discovery and Linkage Grant Schemes and the Grains Research and Development Corporation of Australia (GRDC).

Researchers in all Australian States and the Capital Territory now use *Arabidopsis* functional genomics approaches. Projects are generally highly focused but increasingly involve international collaborators. Canberra, Australia's capital city, remains a major node for *Arabidopsis* research activity. Together, CSIRO's Division of Plant Industry, the Australian National University (ANU) and the Center for the Application of Molecular Biology to International Agriculture (CAMBIA) form a remarkable unit of fundamental, industrial and application-driven research.

The Australian Center for Plant Functional Genomics is a major initiative announced in 2001, and it is now underway at the University of Adelaide. Established jointly by the ARC and the GRDC, the center's objective is to contribute to ensuring that Australia remains internationally competitive in plant science research. However, its current focus on major crop plants with little emphasis on *Arabidopsis*. New Zealand has a small population but is nevertheless home to several *Arabidopsis* research programs. Increasing numbers of New Zealand plant scientists are incorporating *Arabidopsis thaliana* into their research, and at least six groups are using functional genomics approaches. Funding is principally available through the Royal Society of New Zealand's Marsden Fund and the New Zealand Foundation for Research, Science and Technology. Geographically, *Arabidopsis* research seems to be concentrated in three regions: in the North Island cities of Auckland and Palmerston-North and at the University of Otago in Dunedin, on the South Island. In addition to the projects being conducted at the universities, research programs are carried out at the Governmentowned Crown Research Institutes, including Horticulture and Food Research Institute of New Zealand (HortResearch) and the New Zealand Institute for Crop & Food Research Limited (Crop & Food Research).

The horticultural industry is a big part of the New Zealand economy and, reflecting this, much of the *Arabidopsis* research impinges on reproductive development and fruiting. Other functional genomics programs include work on a magnesium transporter gene family and a recently initiated study on the role and function of carboxylesterases.

Austria

http://www.*Arabidopsis*.org/info/2010_projects/Austria.html
Contact: Heribert Hirt
Vienna Biocenter, Vienna
HEHI@gem.univie.ac.at

In recent years, major changes have taken place in the development of molecular biology research facilities in Austria. One of the hot spots of constant change is the Vienna BioCenter, a newly established science campus close to the city center. In addition to several smaller biomedical companies, the Vienna BioCenter has become home of various research institutes from the University of Vienna, the Academy of Sciences and the pharmaceutical company Boehringer-Ingelheim. These developments prompted the government, local authorities and the University of Vienna to concentrate plant molecular research groups from Botany, Microbiology and Genetics, Biochemistry and Molecular Cell Biology, and Medical Biochemistry institutes at the Pflanzen Molekularbiologie Zentrum (PMZ). The PMZ facilities are already constructed and the center is expected to open in early 2005.

Adjacent to the PMZ, the Austrian Academy of Sciences is establishing two new institutes: the Gregor-Mendel-Institute of Molecular Plant Sciences (GMI) and the Institute of Molecular Biotechnology (IMBA). Whereas the IMBA will concentrate on generating knowledge that aims ultimately at curing major human diseases, the goal of the GMI is a basic understanding of how plants work. Construction of both institutes has just begun and their opening is scheduled for 2005. Also, the Gregor-Mendel-Institute has accepted to buy an Affymetrix workstation so that Austrian researchers can process Affychip microarray data of the various *Arabidopsis* genomics consortia. These new developments add considerable value to Austria's research potential and provide the necessary critical mass for starting a coordinated thematic program on *Arabidopsis* biology.

It is the intention of the Austrian Platform of *Arabidopsis* Research (APAR) consortium to function as a research platform coordinating and promoting *Arabidopsis* research in Austria. The activities of APAR are tightly linked to several programs of the European Union and to the worldwide coordination efforts by MASC. Additional Austrian project partners will be incorporated into APAR in the future. APAR currently comprises several projects. For example, (i) molecular regulation of cytokinesis during plant development, (ii) molecular analysis of MAPK-mediated ethylene signaling in *Arabidopsis thaliana*, (iii) analysis of glycogen synthase kinase/shaggy-like kinases, (iv) novel aspects of salt stress signaling in plants, (v) specificity and functional analysis of a PP2C protein phosphatase gene subfamily, (vi) calcium-dependent protein kinases in *Arabidopsis* signal transduction, and (vii) the functional study of the Ku complex at *Arabidopsis* telomeres. One hundred and fifty participants joined the trilateral (Austrian, German and Swiss) APAR meeting held in Vienna, 15-17 April, 2004. Additional activities on *Arabidopsis* research in Austria include projects examining structure-function relationships of ribonucleoproteins, signal transduction and cell cycle regulation, auxin and cytokinin, transport and cell differentiation, epigenetics, chromosome biology, genes involved in the reprogramming of microspores, and MAP kinase signal transduction in plants.

Funding for *Arabidopsis* research in Austria is available from Fonds zur Förderung der wissenschaftlichen Forschung (FWF; basic research only) http://www.fwf.ac.at, Wiener Wissenschafts-, Forschungs- und Technologiefonds (Vienna region) http://www.wwtf. at, Bundesministerium für Bildung, Wissenschaft und Kultur (BMBWK) http://www.bmbwk. gv.at/, and the Austrian Industrial Research Promotion Fund (FFF; applied research), http://www.fff.co.

Canada

http://www.*Arabidopsis*.org/info/2010_projects/Canada.html
Contacts:
Bill Crosby
University of Saskatchewan, Saskatoon, Saskatchewan
bcrosby@cs.usask.ca
Peter McCourt
University of Toronto, Toronto, Ontario
mccourt@botany.utoronto.ca

Arabidopsis functional genomics efforts are ongoing at several major institutions in Canada. The *Arabidopsis* Research Group (ARG) at the University of Toronto, which includes eight research groups housed out of the Department of Botany, was originally established to provide resources and expertise for the *Arabidopsis* community in Canada. Programs sponsored by ARG are jointly funded through the Ontario Genomics Initiative (OGI), Genome Canada, the National Science and Engineering Research Council (NSERC) and by private industry. All resources and data will be made publicly available through various databases and international stock centers. Contacts for each program are listed at http://www.genomecanada. ca/GCprogrammesRecherche/projets/index. asp?l=e or the ARG program director, John Coleman, can be reached directly at coleman@botany.utoronto.ca.

The functional genomics program at the University of British Columbia includes participants from the Biotechnology Laboratory, Botany and Plant Science Departments, along others. The program has recently received diverse funding input to support its projects, including CFI, NSERC, OTIP, FRBC, HFSP, Genome BC, and Genome Canada. Select program elements include the exploitation of *Arabidopsis* as a model system for studying development and the development of TILLing resources.

The recently implemented University of Saskatchewan program derives from activities initiated in late 1999, under the auspices of the National Research Council Genomics in Health and Agriculture Initiative (NRC - GHI). The program was additionally funded by Genome Canada, the Saskatchewan-Canada Agriculture- Food Innovation Fund and, more recently, it has been linked to an NSF 2010 project concerned with the functional genomics of the Ubiquitin-Protein Ligase (E3) families in *Arabidopsis*. In addition, the United States have supported a new Bioinformatics group that includes a research emphasis involving plant genomics and Systems Biology.

The ongoing program at the NRC Plant Biotechnology Institute continues to explore the interface between *Arabidopsis* functional genomics for its implication to Brassica crop improvement with a new emphasis on food quality and secondary metabolism. The Saskatoon Research Center of Agriculture Canada is conducting an active program designed to exploit *Arabidopsis* model system in support of genomics approaches to Brassica crop development. The program is funded by the Agriculture Canada Genomics Program and is supplemented by recent support from Genome Canada. Program elements include genetic, physical and bioinformatics approaches to defining the relationship between the *Arabidopsis* and Brassica genomes and the development of an *Arabidopsis* activation-tagged T-DNA insert population.

China

http://www.*Arabidopsis*.org/info/2010_projects/China.html
Contact: Yang Weicai
Institute of Genetics and Developmental Biology
Chinese Academy of Sciences, Beijing
wcyang@genetics.ac.cn

The *Arabidopsis* community has rapidly expanded in China these past few years. More than 250 participants attended the Annual Workshop on *Arabidopsis* Research, held in Shanghai on November 30, 2003. The workshop was organized by Zhihong Xu, President of Peking University, and featured eighteen oral presentations. In 2002, the National Science Foundation of China (NSFC) provided a grant of US$1.5 millions for a major international collaborative project aimed at the proteomic characterization and functional studies of approximate 1,600 *Arabidopsis* transcription factors. The project involves multiple leading academic institutions in China including Peking University, the Institute of Genetics and Developmental Biology of the Chinese Academy of Sciences (CAS), Fudan University, Wuhan University, Shanghai Jiao Tong University, and Shanghai Institute of Plant Physiology and Ecology of CAS. The coordinators of the project are Xing-Wang Deng (Peking University/ Yale University/CAS Center for Plant Molecular Genetics and Agrobiotechnology and Yale University) and Yuxian Zhu (Peking University). During the first phase of the project, an ORFeome collection for the *Arabidopsis* transcription factor genes has been generated in a Gateway high-through-put cloning vector (Gong *et al.*, Plant Physiology, in press). ORFs of 1282 *Arabidopsis* transcription factors in GATEWAY entry vectors are deposited at ABRC (http://www.*Arabidopsis*.org/news/news.jsp#orf). In a separate effort, funded by the Ministry of Science and Technology of China (MOST; US$ 350,000), an inducible enhancer/ promoter vector was used to generate activation tagging lines (Jianru Zuo, Institute of Genetics and Developmental Biology, CAS). More than 55,000 T1 transgenic lines had been collected by the end of 2003, 35,000 of which were generated in Zuo's lab and 20,000 lines generated in Yingtang Lu's lab at Wuhan University.

Funding for *Arabidopsis* functional genomic research is available from the Ministry of Science and Technology of China (www.most.gov.cn), National Science Foundation of China (NSFC - www.nsfc.gov.cn), CAS (www.cashq.gov.cn), and other sources on a competitive basis.

Eastern European *Arabidopsis* Activity

Contact: Martin Fellner
Palacky University
Department of Cell Biology and Genetics and
Institute of Experimental Botany ASCR
Olomouc, Czech Republic
emfee@prfholnt.upol.cz

Arabidopsis research in former communist countries is relatively new, small and often isolated. The goal of the Eastern European *Arabidopsis* Activity (EEAA) is to integrate the *Arabidopsis* community in Eastern Europe and incorporate its program into the international *Arabidopsis* effort. The purpose of the EEAA is to initiate a joint research project, potentially in collaboration with some of the already established *Arabidopsis* laboratories around the world. EEAA's long-term objective is to grow the *Arabidopsis* community and boost the prestige of plant science in Eastern Europe. During 2003, laboratories from six countries demonstrated repeated interest in the EEAA and are currently investigating various topics.

Czech Republic

Interaction between blue light signaling and abiotic stress (supported by Academy of Sciences of the Czech Republic). Identification of genes integrating hormone and light signaling. Martin Fellner, emfee@prfholnt. upol.cz, http://genetika.upol.cz/. Analysis of 1500 *Arabidopsis* insertional lines (containing insert of T-DNA with tetramer of enhancer from 35S promoter) with respect to flower and root mutations, mutations in responses to elevated boron concentrations and in response to Plasmodiophora brassiceae infection.

Tomá_ Kocábek, kocabek@umbr.cas.cz,
http://www.umbr.cas.cz/805_www/Kocabek/index.htm.

Hungary

Function of phosphoprotein phosphatases (supported by the Hungarian Scientific Research Fund).

Ilona Farkas, farkas@jaguar.dote.hu,
http://www.dote.hu/tudomany/whoiswho98/119.htm

Collecting cDNA clones and insertion mutants representing protein family designated as "Receptor-like cytoplasmic kinases VI" involved in Rop GTPase-dependent signaling cascades (supported by the National Grant Agency).

Attila Fehér, fehera@nucleus.szbk.u-szeged.hu.

Lithuania

Functional activity of the plasmatic and vacuolar membrane.
Plant responses to salt stress.

Gemir Maksimov, gemir@botanika.lt,
http://ml.lms.lt/200203/disertacijos.htm.

Poland

Identification and characterization of enzymatic activity of all *Arabidopsis* ORF's containing Nudix/MutT domain and *Arabidopsis* protein Ku70.
Marta Dobrza_ska, martad@ibb.ww.pl, http://www.ibb.waw.pl/
Investigation of plant genes transcriptional activation and repression mechanisms through remodeling of chromatin structure. Biological functions of linker (H1) histones.

Andrzej Jerzmanowski, andyj@ibb.waw.pl, http://www.ibb.waw.pl/.

Russia

Characterization of the state of phytochromes and (proto) chlorophylls in their native state in the cell (supported by state funding and by the Russian Foundation).

Vitaly Sineshchekov, V.Sineshchekov@mtu-net.ru.

Uzbekistan

Thionines-cystein rich peptides and Isolation and physico-chemical characterization of hormone-binding proteins.

O. Veshkurova, G. Mavlonov, ali@ibchem.ccc.uz.

European Union

http://www.*Arabidopsis*.org/info/2010_projects/EU.html
Contact: Bernard Mulligan
bernard.mulligan@cec.eu.int

Opportunities for functional genomics research on all organisms can be found in several areas of the current 6th Framework Program "FP6"(2002-2006), the European Union's research funding program. To be eligible for FP6 funding, research programs must involve laboratories from several European Member States. However, many opportunities also exist for researchers from countries outside Europe to be involved in programs funded through FP6. In fact, in certain cases, researchers from countries outside Europe can receive FP6 funding. FP6 funds large scale "networks of excellence" and "integrated projects" with grants of Euro 10 million or more as well as smaller targeted projects and individual research fellowships. Funding opportunities for coordination projects and for activities (e.g., conferences and workshops) to support the development of European Union science policy (e.g., in areas relating to functional genomics research) are also available. Details about all these opportunities can be found at http://fp6.cordis.lu/ fp6/home.cfm and http://europa.eu.int/comm/research/fp6/ index_en.html.

The large-scale projects are often very multidisciplinary in nature. A good example is the integrated project "Grain Legumes". This highly multidisciplinary project will develop new genetic, genomic, post-genomic, and bioinformatics tools to improve and sustain grain legume seed production and quality. Notably, the project will contribute to the complete sequencing, within an international project, of the gene-rich regions of the Medicago truncatula genome which is a relevant model system for European grain legumes. "Grain Legumes" fully recognizes the value of the model plant *Arabidopsis* and consequently will fully integrate *Arabidopsis* research or data derived from this model system in several of its activities. With 54 partners in 18 countries, this project is expected to build a European area for Grain Legumes research. Further information about the project can be found at http://www.eugrain-legumes. org/.

Another example is the network of excellence "Epigenetics" which includes a joint research program in the field of epigenetics. Epigenetics involves 25 research teams of top European scientists with a proven track record as leaders in their field. They will constitute the 'virtual core center' by combining their expertise and resources. The 25 core research teams are geographically clustered around eight established centers of epigenetic research and in some cases benefit from established collaborations and synergies that have emerged from previous European Union programs (e.g., 5th Framework Program). The research program of the core teams addresses the functional analysis of epigenetic control in many different organisms (e.g., S.cerevisiae, S.pombe, plants, Drosophila, Xenopus, mouse, human) and applies varied and wide ranging genetic, biochemical and cytological approaches. The strength of such a core program lies in its focus on the molecular mechanisms underlying epigenetic control rather than on purely descriptive and phenomenological analyses. For further details please see http://www.epigenome.imp. ac.at/.

The ERA-NET grant scheme is a novel feature of the 6th Framework Program. It provides support for transnational networking and coordination of national research programs. Therefore, the scheme's participants are program managers working in national ministries and funding agencies. The "European Research Area – Plant Genomics", with a grant of 2.2 million Euros, focuses on networking of national programs to help maximize the return on the Euro 80 million invested in plant genomics across Europe each year. The network will formulate long-term research goals and objectives for plant genomics in Europe and identifying areas in which Europe should contribute to international programs (see http://www. cordis.lu/coordination/publications.htm and http:// www.genomics.nl/homepage/research/ funding_opportunities/eranet_(fp6)_projects/).

In addition, a project database is being set up for projects funded under FP6 (http://www.cordis.lu/ fp6/projects.htm). A database of previously funded European Union projects is available at http://www. cordis.lu/en/home.html. There will be a call for grant proposals to the EU shortly (see http://www.cordis.lu/lifescihealth/workprogramme.htm) with a deadline around November 2005. One of the topic areas open specifically features *Arabidopsis*.

France

http://www.*Arabidopsis*.org/info/2010_projects/France.html
Contact:Ian Small
Evry Plant Genomics Research Laboratoy (URGV) Institut National de la Recherche Agronomique (INRA), Evry
small@evry.inra.fr

The major source of funding in France for the *Arabidopsis* functional genomics projects is Génoplante (http://www.genoplante.com/), a joint venture created by public funding agencies (INRA, CNRS, CIRAD, IRD) and several French ag-biotech companies such as Biogemma, Aventis CropScience, and Bioplante). Since 1999, Génoplante has directed over 62 million of research on *Arabidopsis* (around 50 million of this coming from the recurrent budgets of the government agencies, universities and companies involved and 12 million from Génoplante). The major part of this effort has been on targeted functional analyses of specific biological questions or protein families, too numerous to list here.

Some large generic programs are still ongoing

FLAGdb++ (http://genoplante-info.infobiogen.fr/FLAGdb/),an *Arabidopsis* genomics database including amongst many other things an inventory of flanking sequence tags from the Versailles *Arabidopsis*T-DNA collection. FLAGdb++ now also includes the rice genome and its annotation.

CATMA (http://www. catma.org/), a complete *Arabidopsis thaliana* microarray containing more than 24000 gene-specific tags.This is a program involving several EU countries that is being continued in the EU-funded CAGE project (http://www.psb.ugent.be/CAGE/index.htm).

ATOME (http://genoplante-info.infobiogen.fr/Databases/CT_Nouveaux_Outils/NO2001054/index.html) An *Arabidopsis thaliana* ORFeome. ATOME is aiming at cloning up to 5000 *Arabidopsis* ORFs into Gateway entry vectors.

The Génoplante-Info database (http://genoplante-info.infobiogen.fr/) contains data from many *Arabidopsis* projects including those listed above. Génoplante funded projects in 2004 include efforts on biotic stress (D Roby, I Jupin, P Saindrenan, L Jouanin), epigenetics (V Colot, O Vionnet), seed development (M Caboche, L Lepiniec) and new methods for purifying tagged protein complexes (H Mireau). Several of these projects are jointly funded projects with similar German and Spanish initiatives.

The prospects for 2005 are on paper rather bright, with estimates of at least 5 million in funding for generic plant research via the newly created National Research Agency (ANR), to be distributed primarily via the GenAgro program administered by Génoplante. It remains to be seen whether the promised funding will actually be forthcoming.

A major addition to the French research infrastructure in the last year has been the creation of the National Resources Centre for Plant Genomics (CNRGV) in Toulouse (http://cnrgv.toulouse.inra.fr/ENG/index.html). The Centre is distributing *Arabidopsis* cDNA and BAC clones.

Germany

http://www.*Arabidopsis*.org/info/2010_projects/Germany.html
Contacts:
Thomas Altmann
Max Planck Institute for Molecular Plant Physiology, Golm
Altmann@mpimp-golm.mpg.de
Gerd Jürgens
University of Tübingen, Tübingen
gerd.juergens@uni-tuebingen.de

Arabidopsis functional genomics research has received strong support in Germany through the implementation of two major research programs supported by the Federal Ministry for Education and Research (BMBF) and the Deutsche Forschungsgemeinschaft – German Research Foundation (DFG).

The first of these programs is **G**enom**a**nalyse im **bi**ologischen System Pflanze (GABI), genome analysis in the plant biological system (http://www.gabi.de/). GABI was initiated in 1999 aiming at strengthening plant genome research in Germany, establishing a network of competence including public, private research groups and corporations, and enhancing international collaboration and transfer of knowledge into application. The second phase of the program has recently been started and will last until the end of 2007 with a budget of Euro 10 million per year. The support of private partners involved in the program has increased from 10% in the first program phase to 20% in the second phase.

About 50% of GABI's funding in the first program phase was devoted to work on the model system *Arabidopsis thaliana*. In the second program phase, support for the model organism is reduced. However, the interlocking of research on a model organism and the transfer of these results to crop plants is a fundamental principle of GABI. Therefore, "bridging projects" embed research on the model *Arabidopsis* with crops within single research consortia. Established rules regulate disclosure and use of research results obtained though GABI activities. Several GABI projects provided major recent contributions to the international efforts on *Arabidopsis* functional genomics: a large collection of sequence-indexed T-DNA insertion lines (GABI-KAT; http://www.mpiz-koeln.mpg.de/GABI-Kat/), a database of membrane proteins (Aramemnon; http://crombec.botanik.uni-koeln.de/index.html), and extensive SNP information for 13 different *Arabidopsis* accessions (MASC-DB; http://www.mpiz-koeln.mpg.de/masc/). Maintenance and further development of MAtDB at MIPS http://mips.gsf.de/proj/thal/ are also being supported by GABI.

One of GABI's major targets is the establishment and support of international collaborations. A first step towards setting up direct collaborative efforts in Europe has been the establishment of joint research projects between the French plant genome program, Génoplante, and the German GABI initiative. This bilateral interaction was expanded to a trilateral co-operation including the Spanish genome program. In 2004 nine trilateral research projects plus five bilateral projects between France and Germany started. A recently funded European Research Area Network Plant Genomics (ERA Net PG; http://www.erapg.org) is another example of plant genomics as a front runner. Both German programs, GABI and AFGN (see below), played an important role during the establishment of this network and, consequently, performed a significant function in the creation of the European Research Area. The second major funding initiative for *Arabidopsis* functional genomics research is the *Arabidopsis* Functional Genomics Network (AFGN), funded since 2001 by the DFG. AFGN was founded in close coordination with the 2010 Project of the United States National Science Foundation. Both programs were established with the goal of elucidating the function of all *Arabidopsis* genes by the year 2010. Eleven AFGN projects started in 2001 and 21 more projects in 2002, ending in 2004. The second phase of the program began in 2004 and will end in 2007 funding 24 projects. As a result of the increasing interaction between these two funding agencies, in 2004, AFGN proposals submitted to the DFG and 2010 Project proposals submitted to NSF were co-reviewed by a joint AFGN-NSF panel. Transnational co-operative projects were especially encouraged. Information about the AFGN project can be found at http://www.uni-frankfurt.de/fb15/botanik/mcb/AFGN/AFGNHome.html and information about individual AFGN funded projects at http://www.uni-frankfurt.de/fb15/botanik/mcb/AFGN/Members.html or at the functional genomics web page http://www.*Arabidopsis*.org/info/2010_projects/AFGN_Abstracts.jsp.

AFGN has taken the lead in setting-up an international joint effort to establish a comprehensive genome-wide *Arabidopsis* transcriptome reference database. AtGenExpress is a multinational coordinated effort to uncover the transcriptome of the multicellular model organism *Arabidopsis thaliana* coordinated by Detlef Weigel, Thomas Altmann and Lutz Nover. The overall database derived from about 1300 microarrays (i.e., more than 30 million data points) is accessible via TAIR and is released to the public Gene Expression Omnibus (GEO) and ArrayExpress databases. The data are accessible at different data bases and tools as mentioned earlier in the report.

Italy

http://www.*Arabidopsis*.org/info/2010_projects/Italy.html
Contact: Paola Vittorioso
University of Rome "La Sapienza", Rome
paola.vittorioso@uniroma1.it

Several Italian groups have been actively engaged in *Arabidopsis* research in recent years. Most of these groups are involved in national and international plant functional genomics network projects. In 2003, a common technological platform was developed creating a network among groups of the highest qualification active in Italian universities, public research institutes and the most relevant plant biotechnology companies. This national network, funded by the Italian Ministry of Research (MIUR; www.miur.it), could represent a first step towards the establishment of a National Plant Biotechnology Center (From *Arabidopsis* to tomato: A scientific network and a technological platform for the functional genomics of plant development).

This network intends to exploit a functional genomics approach to analyze selected regulatory aspects of *Arabidopsis* development through the analysis of the function and interactions of members of different families of regulatory and structural genes. On these genes, laboratories involved in the network have achieved results and know-how of the highest international standards. The scope of this project is to gain knowledge on the function of individual genes involved in the different developmental processes analyzed and to identify regulatory networks and interactions between different genes and different processes. It has become increasingly evident that in higher organisms, individual genes influence several processes and, therefore, a satisfactory comprehension of developmental processes can only be achieved through a functional genomics approach.

Analyzed in this research are genes from the:
- Dof family (Vittorioso-Costantino) involved in auxin-dependent meristem formation, in the control of seed germination and in the response to light and gibberellins;
- HD-Zip family involved in the regulation of primary and secondary meristem activity (Morelli) and in developmental processes as a response to the environment, such as shade-avoidance response (Ruberti);
- MYB family (Tonelli) involved in morphogenesis, stress response and in the biosynthesis of nutritionally relevant polymers;
- NF-Y family which interact with several families of transcription factors crucial in differentiation and development in eukaryotes (Tonelli);
- MADS family (Colombo/Kater) involved flower development
- HMG and TAF/TBP families known as important factors in modulating transcription (Colombo/Kater);
  and from the
- E2F family (Cella, Albani) involved in cell-cycle regulation and development.

This research analyzed also genes involved in response to red and far-red light (PHY; Bowler), response to blue light (CRY; Bowler, Benvenuto), in signalosome assembly (DET; Bowler), in the biosynthesis of carotenoids (UR Benvenuto, Cellini), and genes important for photosynthetic activities and nutritional quality. Included were also genes for proteins of the oligogalacturonides (OG)-mediated signal transduction pathway to identify key regulators of the defense response (Cervone), members of the 14.3.3 protein class (Aducci) involved in cell cycle control and in several signal-transduction pathways, and members interacting with 14.3.3 (Soave) and genes involved in iron homeostasis (ferritin) and in detoxification of ROS (Soave).

The different lines of research on these genes are coordinated. New post-genomic technologies will be set up and the use of existing technologies will be made available to all partners of the network. The network will develop and utilize technologies for the functional analysis of the genes (i.e., RNA interference, negative and positive dominant, chemical gene-machine/Tilling), technologies for the analysis of interactions between genes (i.e., *Arabidopsis* macro- and micro-arrays, real-time PCR) as well as technologies for the identification of protein partners and targets (i.e., Surface Plasmon Resonance, two hybrid in yeast and plant, stable antibodies phage display libraries). Moreover, B. Mattei is one of the eight european partners involved in the project "Functional Genomics for Biogenesis of the Plant Cell Wall" which has been recently funded by the UE Marie Curie Training Network. This project will be developed on *Arabidopsis* as a model system. The project is supposed to start in June 2005 and to last 4 years. Futhermore L. Colombo is the coordinator of the EU FP6 Marie Curie Training Project "TRANSISTOR" (Trans-cis element regulating key switches in plant development). In addition, Mariotti, Marmiroli, Migliaccio, and Perata groups are involved in *Arabidopsis* projects funded by the Italian Space Agency, the European Space Agency, and the Institut Pasteur.

Japan

http://www.*Arabidopsis*.org/info/2010_projects/Japan.html
Contact: Kazuo Shinozaki
RIKEN Plant Science Center
sinozaki@rtc.riken.go.jp

Japan has been a worldwide leader in *Arabidopsis* research and is continuing that tradition by moving forward into the world of functional genomics. In Japan, ongoing programs for *Arabidopsis* functional genomics are found at RIKEN Genomic Sciences Center, Plant Functional Genomics Research Group (http://pfgweb.gsc.riken. go.jp/), RIKEN Plant Science Center (http://www.psc.riken.go.jp/indexE.html), Kazusa DNA Research Institute (http://www.kazusa. or.jp/en/plant/), the CREST program of the Japan Science &Technology Corporation, and NEDO project.

Both the RIKEN Genomic Sciences Center Plant Functional Genomics Research Group and the Kazusa DNA Research Institute have ongoing bioinformatics programs as well.

Arabidopsis functional genomics research at RIKEN Genomic Sciences Center (GSC, PIs are Kazuo Shinozaki and Minami Matsui) includes (i) collection and phenotype analysis of Ds-tagged lines (Takashi Kuromori), (ii) collection of full-length cDNAs (Motoaki Seki), (iii) collection and phenotype analysis of activation tagging lines (Miki Nakazawa), (iv) full-length-cDNA-overexpressing (FOX) transgenic lines (Takanari Ichikawa), (v) structural proteomics of plant regulatory proteins with novel structures in collaboration with Protein Research Group of RIKEN GSC (PI: Dr. Shigeyuki Yokoyama) (http://protein.gsc.riken. go.jp/Research/index_at.html). and (vi) transcriptome analysis of genes expression in response to both abiotic and biotic stress using RAFL full-length cDNA microarray analysis (Motoaki Seki). Database is open from WEB site (http://rarge.gsc.riken.go.jp/). Further work on reverse proteomics for functional analysis of in vitro expressed proteins using the wheat germ cell-free protein synthesis system is taking place at RIKEN GSC, in collaboration with a group at Ehime University (Yaeta Endo, Principal Investigator). The RIKEN GSC (Takashi Kuromori) is active in phenotype analysis of Ds-tagged lines in collaboration with RIKEN PSC (Takuji Wada and Kiyotaka Okada). PSC has contributed to AtGenExpress since 2004. Recently, RIKEN PSC (director: Kazuo Shinozaki) has started a project "Understanding metabolic systems for plant productivity" to integrate metabolomics with transcriptomics.

At the Kazusa DNA Research Institute (Satoshi Tabata), ongoing projects include the collection of T-DNA tagged lines and *Arabidopsis* and Lotus japonicas ESTs. A major project is the genomic sequencing of Lotus japonicas.. In addition, *Arabidopsis* T87 cultured cells have been transformed with RAFL cDNAs and other cDNAs for metabolic profiling of primary and secondary metabolites (Daisuke Shibata and Kazuki Saito). Kazusa group has opened a new web site for data search, KaPPA-View: Integration of transcriptome and metabolome data in plant metabolic pathways (Dr. Tashiaki Tokimatsu) http://kpv.kazusa.or.jp/kappa-view/ and KATANA (Kazusa Annotation Abstract): Integration of major database sites of *Arabidopsis* genome annotation (Dr. Kentaro Yano) http://www.kazusa.or.jp/katana/. Several groups at other centers and universities are also involved in *Arabidopsis* functional genomics. The projects involve metabolic profiling in *Arabidopsis* (Chiba University - Kazuki Saito), genome-wide analyses of the two-component system (Takeshi Mizuno) and transcription factor function using repressor domain and overexpressors (Agency of Industrial Science & Technology in Tsukuba - Ohme-Takagi and Kaoru Suzuki).

RIKEN BRC (http://www.brc.riken.jp/lab/epd/Eng/) is funded by the National Bioresource Project of Japan and collects various plant resources from Japanese research institutes and universities. The RAFL clones, Ds-tagged lines and Activation tagging lines mentioned above are distributed from tRIKEN BRC. This year, RIKEN BRC takes over the distribution service of the ecotypes and mutants of *Arabidopsis* from the Sendai *Arabidopsis* Seed Stock Center (SASSC; Nobuharu Goto). Since established in 2001, RIKEN BRC has already distributed approximately 7,000 *Arabidopsis* resources to the world. Masatomo Kobayashi (kobayasi@rtc.riken.jp) is in charge of *Arabidopsis* resources distribution at RIKEN BRC (http://www.brc.riken.go.jp/lab/epd/Eng/).

Other funding opportunities for *Arabidopsis* functional genomics in Japan include CREST of Japan Science and Technology Corporation (http://www.jst.go.jp/EN/), the Program of Promotion of Basic Research Activities for Innovative Biosciences (http:// www.brain. go.jp/welcome-e.html), NEDO, and Grants-in Aid for Science from the Ministry of Education, Science, Culture and Sports (MEXT).

The Netherlands

http://www.*Arabidopsis*.org/info/2010_projects/Netherlands.jsp
Willem Stiekema
Plant Research International, Wageningen
w.j.stiekema@plant.wag-r.nl
Ben Scheres
Department od Molecular Cell Biology, University of Utrecht
b.scheres@bio.uu.nl
Centre for BioSystems Genomics, the Netherlands Plant Genomics Network

Within this national plant genomics research programme the *Arabidopsis* research focuses on the analysis of the regulatory network of genetic, biochemical, physiological and environmental interactions that control plant performance and the complex traits involved in plant-oomycete interactions and adaptation to stresses. Fully integrated large-scale activation tag screening, gene expression, proteome and metabolite profiling is aimed for and based on the full exploration of the available genetic variation with emphasis on control of metabolic composition. Understanding of the adaptive traits relevant for research in potato and tomato is another important subject as is the development of concepts and technologies based on the availability of the whole *Arabidopsis* genome sequence and efficiency of *Arabidopsis* genetics. Four different projects respectively focus on:

1. *Arabidopsis* quality: the genetic and genomics analysis of metabolic composition. Koornneef/Vreugdenhil, Wageningen University; Pereira, Plant Research International, Wageningen; Smeekens, Utrecht University.
2. The analysis of ligand-receptor interaction networks in *Arabidopsis*.
   Liu, Plant Research International, Wageningen; Heidstra, Utrecht University; De Vries, Wageningen University.
3. The role of chromatin structure in gene expression of *Arabidopsis* and tomato.
   Bisseling/de Jong, Wageningen University.
4. Priming of defence gene expression in plant-oomycete interactions.
   Pieterse/van Ackervecken, Utrecht University.

Arabidopsis projects funded by other sources such as first flow university funds, second flow Netherlands Organisation for Scientific Research, EU etc. and third flow contract research.

Wageningen University

1. QTL express: identification of plant performance traits in *Arabidopsis* by combining high-throughput mapping and expression profiling (M. Koornneef, D. Vreugdenhil).
2. Heavy metal tolerance and accumulation in Thlaspi caerulescens, a heavy metal hyperaccumulating plant species (M. Aarts).
3. Do plants love heavy metals? *(*A. Assunção)
4. The role of tomato serine and cysteine proteases in defence signalling (R. van der Hoorn)
5. A molecular genetic approach to chemical ecology and community ecology (M. Dicke)
6. Cross-talk between signal-transduction pathways in induced defence of *Arabidopsis* against microbial pathogens and herbivorous insects. M. Dicke (joint projects with C. Pieterse, Utrecht University)
7. Development of a method for breeding of cucumber for improved attraction of biological control agents (M. Dicke, H. Bouwmeester)
8. From genetic code to ecological interactions: molecular, phytochemical and ecological aspects of a glucosinolate polymorphism in Barbarea vulgaris. (N. van Dam)
9. *Arabidopsis*: the system to study structure and function of heterochromatin (T. Bisseling)
10. Chromatin genomics: functional analysis of *Arabidopsis* chromatin remodeling genes in development (T. Bisseling)
11. Wageningen Phytoinformatics: the added value from plants (W. Stiekema)

Plant Research International, Wageningen

1. Identification and characterization of genes for drought tolerance (A. Pereira).
2. Identification of plant genes for abiotic stress resistance (A.Pereira).

3. Isolation and characterisation of key-genes in the formation of germination stimulants of the parasitic weeds Striga and Orobanche (H. Bouwmeester).
4. LRR receptor-like proteins and their functions in plant signaling (C-M Liu).
5. MADS box transcription factor functioning (G. Angenent)
6. Signalling Pathways Controlling Embryogenic Cell Development in *Arabidopsis* (K. Boutilier)
7. Signalling in the shoot apical meristem: A question of determinate or indeterminate growth (R. Immink)

Utrecht University

1. Interaction between sugar-and hormone signalling pathways in plants (J. Smeekens)
2. Trehalose-6-phosphate as a regulatory molecule in plants (H. Schlüpmann)
3. Control of plant architecture (M. Proveniers)
4. Dormancy as survival mechanism in plants *(*L. Bentsink)
5. Induced disease resitance signaling in *Arabidopsis* (C. Pieterse)
6. Cross-talk between signal-transduction pathways in induced defence of *Arabidopsis* against microbial pathogens and herbivorous insects (C. Pieterse, joint projects with M. Dicke, Wageningen University)
7. Plant innate immunity: cross-talk between signaling pathways to fine tune defense (C. Pieterse)
8. A functional proteomics approach to identify phospho-proteins involved in plant innate immunity (F. Menke)
9. Priming in plant-pathogen interactions: the molecular mechanism of the alarmed state (J. Ton)
10. Signalling at the host-microbe interface: pathogen-induced modulation of the plant plasma membrane (A. van den Ackerveken)
11. Genetic networks in root development: Interplay between cell polarity information, pattern formation cues, and control of cell division; Chromatin dynamics; Ubiquitination and cell cycle control are investigated with an emphasis on the APC complex and its regulation (B. Scheres)
12. Genomics for multicellular development: Function of the quiescent center in regulation of pattern formation and differentiation within the *Arabidopsis thaliana* root meristem (R. Heidstra)
13. Analysis of the hyponastic and differential growth response of *Arabidopsis thaliana* petioles induced by submergence and low light conditions (T. Peeters, R. Voesenek)

Leiden University

1. Characterization of a novel regulator of plant secondary metabolism (J. Memelink)
2. Effect of NHR mutations on genome stability and development in *Arabidopsis* (P. Hooykaas)
3. ORA EST: Functional analysis of jasmonate-responsive AP2/ERF-domain transcription factors in *Arabidopsis thaliana* (J. Memelink)
4. Regulation of polar auxin transport by PINOID protein kinase signaling to the actin cytoskeleton. (R. Offringa)

University of Amsterdam

1. Role of PA kinase in plant stress signalling (T. Munnik)
2. Targets for the novel lipid second messenger, phosphatidic acid (C. Testerink)
3. SUMO-signaling in plants (H. van den Burg)

Vrije Universiteit, Amsterdam

1. Function of meristem identity in flower and inflorescence development (R. Koes)
2. Genetic control and evolution of inflorescence architecture (R. Koes)

University of Groningen

1. Molecular biology of programmed cell death in higher plants (Dijkwel, J. Hille)

All Scandinavian countries have their own national research funding system. The Nordic Research Board, NordForsk, is funding a 5-year (2001-2005) Nordic Network for research groups in Finland, Sweden, Norway, Iceland and Denmark that are involved in research with *Arabidopsis*. The Nordic *Arabidopsis* Network (www.*Arabidopsis*.no) aims at keeping the groups in regular contact with each other, and it also offers small mobility grants for graduate students and post docs for short-time exchange between groups.

Nordic *Arabidopsis* Network

http://www.*Arabidopsis*.org/info/2010_projects/Nordic.html
Contact: Jaakko Kangasjärvi
University of Helsinki
Helsinki, Finland
jaakko.kangasjarvi@helsinki.fi

The Norwegian Plant Functional Genomics Program (NARC) started in 2003 and is fully operative. NARC is 1 out of 11 genomics technology platforms forming the national functional genomics program (FUGE). The plant platform includes service activities within transcriptional profiling (full genome arrays and custom designed arrays) and bioinformatics (Bones lab, NTNU) genotyping and clone collection (Rognli lab, UMB), in situ hybridization and yeast two-hybrid screening (Aalen lab, UIO). Most of the activity is directed against *Arabidopsis thaliana*. The service facilities are open for all scientists at equal conditions. The program is coordinated by Atle M. Bones (NTNU) and information about the services can be found at www.narc.no or by request to narc@bio.ntnu.no . Norway is a partner of the EU Plant Genomics network ERA-PG and hosted the Nordic *Arabidopsis* meeting 2004.

In Sweden, the Umeå Plant Science Center (UPSC) has been created by moving plant groups from the Umeå University and Swedish University of Agricultural Sciences (Umeå) to the same building. UPSC groups have also received National Center of Excellence status and funding for functional genomics. Their activities are mainly concentrated in trees (hybrid poplar). However, *Arabidopsis* functional genomics is heavily utilized for the determination of the function of poplar genes that have a well-conserved counterpart in *Arabidopsis*. The UPSC is also a partner in the European CATMA-project. Groups from the Uppsala University are involved in two EU-projects that aim at the elucidation of several transcription factors groups in *Arabidopsis*.

The Finnish groups involved in *Arabidopsis* research are concentrating on stress-physiology and functional genomics of plant stress responses, developmental and hormone biology, and in photosynthesis. They are using genomics, proteomics, and metabolomics to determine plant defense and adaptation to biotic and abiotic stresses and the function for the proteins in chloroplast thylakoid membranes. *Arabidopsis* genomic information is also used in functional and comparative genomics of the lower plants as a template for the eurosids. Information is stored and made available at http://sputnik.btk.fi. The outcrossing relative *Arabidopsis lyrata* is being used in studies of population genetics of adaptation to abiotic conditions. The eight chromosomes of the species differ from the A. *thaliana* genome mainly by a small number fusions and reciprocal translocations. The Finnish Plant Functional Genomics Project Program was created spring of 2003 in order to increase collaboration in functional genomics between the participating groups. It is also member in the European plant functional genomics network ERA-PG.

In Denmark, a number of groups at The Veterinary and Agricultural University, Copenhagen University, Risø National Laboratory, Danish Institute of Agricultural Sciences and Aalborg University work on *Arabidopsis*. The research, which in most cases is funded by the national research councils, involves studies of several aspects of plant life. The activities are coordinated through the Plant Biotech Denmark-network (www.plant-biotech.dk). The 2005-workshop in the Nordic *Arabidopsis* Network will be organised in Denmark on October 13-15.

The Icelandic investigators involved in *Arabidopsis* research have promoted *Arabidopsis thaliana* as a model research plant within the Icelandic research community.

United Kingdom

http://www.*Arabidopsis*.org/info/2010_projects/United_Kingdom.jsp
Contact: Andrew Millar
University of Edinburgh, Rutherford Building, Mayfield Road, Edinburgh EH9 3JH
Andrew.Millar@ed.ac.uk

UK Funding

The major funding agency for plant science in the UK is the Biotechnology and Biological Science Research Council, BBSRC. BBSRC encourages applications that use genomic technologies and has launched several initiatives to stimulate research in this area, including exploiting genomics, proteomics, metabolomics and systems biology (see http://www.bbsrc.ac.uk/science/initiatives/). A report to BBSRC in 2004   (http://www.bbsrc.ac.uk/about/pub/reports/crop_sci_review12_05_04.html) recommended greater emphasis on the exploitation of *Arabidopsis* research for crop science. Other UK funding bodies supporting plant science include NERC (Natural Environmental Research Council), DEFRA (Department for Enviornment Food and Rural Affairs and SEERAD (Scottish Executive Environment and Rural Affairs).

GARNet

GARNet, the **G**enomic ***A**rabidopsis* **R**esource **Net**work was established in 2000 to provide reliable, efficient, user-driven and publicly available functional genomics resources for *Arabidopsis* research. The original three year funding period for GARNet was extended in 2003 to allow the services provided by GARNet to move to full cost recovery. Coordination activities were funded for a further 5 years from 2005, to keep services up to date, provide an information resource (website http://garnet.*Arabidopsis*.info/, newsletter, annual meeting) and point of contact for other UK plant communities and international plant genomics programmes. Resources initiated by GARNet include transcriptomics, proteomics and metabolomic . Insert clone libraries and a screening service are available from GeTCID and additional insertional mutagenesis populations generated in the first funding period are now available at the Nottingham *Arabidopsis* Stock Center, NASC. Data from GARNet-funded Affymetrix and proteomics are available at NASC (http://*Arabidopsis*.info), and from metabolomics results at http://www.metabolomics.bbsrc.ac.uk/. Many leading universities and institutes in the UK have established their own functional genomics resource centers, further information on these can be found at http://garnet.*Arabidopsis*.info/*Arabidopsis*_resources.htm.

European Stock Centre

NASC (http://*Arabidopsis*.info/) makes a wide range of seeds, DNA and database information available to the research community as an open international resource. NASC and the ABRC hold and curate duplicated lines as safety copies such that the onus of acquiring, curating, bulking, and distributing is shared by both centers. Distribution from NASC alone is about 35,000 tubes of seed per year worldwide. In 2005 NASC has actively hand-curated all of its archival phenotype data into the international plant ontology and Phenotype/Trait ontology standards (http://*Arabidopsis*.info/bioinformatics/Ontology_details.html) not only for improved analysis and access today but also to enhance future automated / machine access to the data. Browser resources from NASC include http://atensembl.*Arabidopsis*.info, a comprehensive genome oriented "one-stop-shop" bringing together a variety of resources including MIPS and TIGR annotation linked to germplasm information; an extensive database of Affymetrix GeneChip data; comprehensive insertion line coverage; and for the Madison meeting we will have achieved full integration of recently generated AGRIKOLA EU consortium RNA-I germplasm and gateway clone information and physical resources. NASC also provides an open international (not-for-profit) genechip hybridization service. To date NASC has proactively released data from over 2,000 genechips into the community originating from researchers operating on 5 continents. Most of NASCs resources are also now available as BioMOBY web services (70+ services) that may be openly accessed by other databases, automated data-mining engines and desktop workflow applications such as Taverna. These next-generation services were made possible through grant QLRI-CT-2001-00006 (PLANET) awarded by the EU commission. The UK BBSRC government funding body funds all other NASC services.

Other Programmes

In addition to GARNet funded projects, UK *Arabidopsis* researchers are also involved in numerous functional genomic programmes including CATMA, Agrikola, METRO, Phosophoprotemics and Protein Localisation. This is in addition to projects aimed at functional analysis of specific gene families/functions such as cell cycle, chloroplasts development, meiosis and microtubules.

UK *Arabidopsis* Meetings

GARNet hosts an annual meeting for plant scientists across the UK and Europe to disseminate information about new technologies and resources. GARNet 2005 will be held at the John Innes Centre Norwich 5-6th September (see http://garnet.*Arabidopsis*.info/garnet_meeting.htm). GARNet is involved in the organization of PlantGEMS 2005 (http://www.plantgems.org/). The Genetics Society hosts a one-day meeting on *Arabidopsis* annually, in 2005 this is in Edinburgh on May 8 (http://www.genetics.org.uk).

Future

Arabidopsis researchers need to understand the contemporary questions in crop science. Multinational funding mechanisms, such as the ERAnet in plant genomics, should facilitate the development of investigator-led projects in this area. Long-term international funding and coordination will also be required to develop the informatics infrastructure, mathematical models and data that are required for a comprehensive framework of *Arabidopsis* biology in silico.

United States

http://www.*Arabidopsis*.org/info/2010_projects/United_States.html
Contacts:
Philip Benfey
Duke University, Durham, North Carolina
philip.benfey@duke.edu
Mary Lou Guerinot
Dartmouth College, Hanover, New Hampshire
Mary.Lou.Guerinot@Dartmouth.edu

The *Arabidopsis* research community in the United States is coordinated by the North American *Arabidopsis* Steering Committee (NAASC http://www.*Arabidopsis*.org/info/2010_projects/NAASC_Info.jsp) which consists of 6-8 elected members who serve four-year terms. Two members rotate off every year. Two members of the Committee represent the U.S. on the Multinational *Arabidopsis* Steering Committee. The NAASC is organizing the International Conference on *Arabidopsis* Research when held in the US. The National Science Foundation (NSF) initiated the *Arabidopsis* 2010 Project in fiscal year 2001. The program's goal is to determine the function of 25,000 genes in *Arabidopsis* by the year 2010. The current foci of the Project are to determine the function of a network of genes and to develop research tools and resources that enable the entire research community to participate in the 2010 activities. NSF requires that the 2010 awards be coordinated with similar activities worldwide, that the investigators post publicly the identity of genes under investigation, and that the outcome of the awards (data, information and materials) be made available to the public according to the timetable approved by NSF.

Twenty-seven projects were funded under this program in 2001, a further twenty projects in 2002 and 20 more projects were funded in 2003. In May 2004, for the 2010 year 4, grant proposals were co-reviewed with the AFGN grant proposals at the NSF in order to support further collaboration between the two projects. Abstracts can be found at http://www.nsf.gov/bio/pubs/awards/2010awards.htm. NSF awarded 17 additional grants in 2004. In 2004 there was a new call for 2010 projects reviewed in 2005. The NSF expects to continue the *Arabidopsis* 2010 Project for 10 years, although the focus of the project may change. In addition to the *Arabidopsis* 2010 Project, other activities related to *Arabidopsis* research are supported by various programs at NSF, including projects in the plant genome program where *Arabidopsis* is often used as reference plant, individual research projects, workshops/meetings, information resources including TAIR and informatics tools development, and the biological resource center, ABRC. NSF award information can be found at https://www.fastlane.nsf.gov/a6/A6AwardSearch.htm. The Center for Eukaryotic Structural Genomics (http://www.uwstructuralgenomics.org/) has been funded by the National Institutes of Health (NIH) to solve three dimensional structures for many of the proteins of the *Arabidopsis* proteome. The U.S. Department of Agriculture, the U.S. Department of Energy and the NIH, especially the National Institutes of General Medical Sciences, support many research projects involving *Arabidopsis*, although they do not have a funding program specifically targeted to *Arabidopsis* research. NIH awards can be searched at http://commons.cit.nih.gov/ crisp3/Crisp_Query.Generate_Screen.

Members of the Multinational *Arabidopsis* Steering Committee

Thomas Altmann
Representing Germany
taltmann@rz.uni-potsdam.de
Universitaet Potsdam
Golm, Germany

Philip Benfey, Chair
Representing the United States of
    America
philip.benfey@duke.edu
Duke University, USA

Jorge Casal
Representing Argentina and Brazil
casal@ifeva.edu.ar
University of Buenos Aires
Buenos Aires, Argentina

Bill Crosby
Representing Canada
bcrosby@cs.usask.ca
University of Saskatchewan
Saskatoon, Saskatchewan,
Canada

Ian Furner
Representing the United Kingdom
ijf@mole.bio.cam.ac.uk
University of Cambridge
Cambridge, United Kingdom

Mary Lou Guerinot
Representing the United States
Mary.Lou.Guerinot@Dartmouth.
    edu
Dartmouth College
Hanover, New Hampshire, USA

Pierre Hilson
Representing Belgium
pierre.hilson@gengenp.rug.ac.be
Ghent University, Belgium

Heribert Hirt
Representing Austria
HEHI@gem.univie.ac.at
Vienna Biocenter,
Vienna, Austria

Gerd Jürgens
Representing Germany
gerd.juergens@uni-tuebingen.de
University of Tübingen, Germany

Jaakko Kangasjärvi
Representing the Nordic
    Arabidopsis Network
jaakko.kangasjarvi@helsinki.fi
University of Helsinki
Helsinki, Finland

Keith Lindsey
Representing the
United Kingdom
keith.lindsey@durham.ac.uk
University of Durham
Durham, United Kingdom

Sean May
Representing the
United Kingdom
arabidopsis@nottingham.ac.uk
Nottingham *Arabidopsis* Stock
    Center
Loughborough, United Kingdom

Peter McCourt
Representing Canada
mccourt@botany.utoronto.ca
University of Toronto
Toronto, Ontario, Canada

Andrew Millar
Representing United Kingdom
Andrew.millar@ed.ac.uk
University of Edinburgh
Institute of Molecular Plant
    Sciences

Lutz Nover
Representing Germany
nover@cellbiology.uni-frankfurt.de
Biozentrum
Universitaet Frankfurt
Frankfurt, Germany

Javier Paz-Ares
Representing Spain
jpazares@samba.cnb.uam.es
Centro National de
Biotechnoligia
Madrid, Spain

Ben Scheres
Representing the Netherlands
b.scheres@bio.uu.nl
University of Utrecht
The Netherlands

Kazuo Shinozaki
Representing Japan
sinozaki@rtc.riken.go.jp
RIKEN Genomic Sciences Center
RIKEN Tsukuba Institute
Tsukuba, Ibaraki, Japan

Ian Small Co-chair
Representing France
small@evry.inra.fr
INRA-Evry
Evry, France

Willem Stiekema
Representing The Netherlands
w.j.stiekema@plant.wag-ur.nl
Plant Research International
Wageningen, The Netherlands

Paola Vittorioso
Representing Italy
paola.vittorioso@uniroma1.it
University of Rome
"La Sapienza"
Rome, Italy

Geoffrey Wasteneys
Representing Australia and New
    Zealand
geoffw@rsbs.anu.edu.au
Australian National University
Canberra, Australia

Yang Weicai
Representing China
wcyang@genetics.ac.cn
Chinese Academy of Sciences
Bejing, China

Isabell Witt
Executive Secretary and
    Coordinator
isawi@duke.edu
Duke University
Durham, NC, USA

Members of the MASC Subcommittees

Bioinformatics

Fabrice Legeai
flegeai@infobiogen.fr

Sean May
arabidopsis@nottingham.ac.uk

Klaus Mayer
Kmayer@gsf.de

Yasukazu Nakamura
yn@kazusa.or.jp

Yves van de Peer
yvdp@gengenp.rug.ac.be

Heiko Schoof
schoof@mpiz-koeln.mpg.de

Chris Town (chair)
cdtown@tigr.org

cDNAs and Clone-Based Functional Proteomics (ORFeomics)

Pierre Hilson (chair)
pihil@gengenp.rug.ac.be

Kazuo Shinozaki
sinozaki@rtc.riken.go.jp

Ian Small
small@evry.inra.fr

Chris Town
cdtown@tigr.org

Multiparallel Analytical Tools & Phenotype Analyses

Gerco Angenent
gerco.anagenet@wur.nl

Mike Beale (chair)
mike.beale@bbsrc.ac.uk

Jerry Cohen
cohen047@umn.edu

Paul DuPree
p.dupree@bioc.cam.ac.uk

Oliver Fiehn
ofiehn@ucdavis.edu

Mary Lou Guerinot
mary.lou.guerinot@dartmouth.edu

Heribert Hirt
hehi@gem.univie.ac.at

Martin Kuiper
martin.kuiper@psb.ugent.be

Keith Lindsey
keith.lindsey@durham.ac.uk

Sean May
arabidopsis@nottingham.ac.uk

Lutz Nover
nover@cellbiology.uni-frankfurt.de

Kiyotaka Okada
Kiyo@ok-lab.bot.Kyoto-u.ac.jp

Göran Sandberg
goran.sandberg@genfys.slu.se

Motoaki Seki
mseki@rtc.riken.go.jp

Mike Sussman
msussman@facstaff.wisc.edu

Reverse and Forward Genetic Stocks

Thomas Altmann
taltmann@rz.uni-potsdam.de

Steve Henikoff
steveh@howard.fhcrc.org

Maarten Koornneef
maarten.koornneef@wur.nl

Minami Matsui
minami@postman.riken.go.jp

Magnus Nordborg
magnus@usc.edu

Randy Scholl
scholl.1@osu.edu

Ian Small
small@evry.inra.fr

Mike Sussman
msussman@facstaff.wisc.edu

Satoshi Tabata
tabata@kazusa.or.jp

Paola Vittorioso
paola.vittorioso@uniroma1.it

Bernd Weisshaar (chair)
bernd.weisshaar@uni-bielefeld.de