

Microarray expression profiling resources for plant genomics

Willem Albert Rensink and C. Robin Buell

The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

Large volumes of genomic data have been generated for several plant species over the past decade, including structural sequence data and functional annotation at the genome level. Various technologies such as expressed sequence tags (ESTs), massively parallel signature sequencing (MPSS) and microarrays have been used to study gene expression and to provide functional data for many genes simultaneously. This review focuses on recent advances in the application of microarrays in plant genomic research and in gene expression databases available for plants. Large sets of *Arabidopsis* microarray data are publicly available. Recently developed array platforms are currently being used to generate genome-wide expression profiles for several crop species. Coupled to these platforms are public databases that provide access to these large-scale expression data, which can be used to aid the functional discovery of gene function.

Plant genomics

To date, near-complete genome sequences are available for only two model species, *Arabidopsis* and rice [1,2]; draft sequences are available or in progress for poplar [3], *Medicago truncatula* [4], *Lotus*, tomato and maize. Annotation of the genome, both at the structural level in which the genes are identified and at the functional level in which the function of the genes and other genomic features are curated, is an iterative process essential to the interpretation of the genome sequence. Indeed, the *Arabidopsis* genome has been re-annotated since its completion [5] and the annotation of the finished rice genome sequence is an on-going process [6]. Structural annotation of genes and gene models relies on abundant experimental evidence such as ESTs and full-length cDNA sequences (FL-cDNA) to improve gene finder output on genomic sequence. By contrast, functional annotation relies primarily on the use of bioinformatics to determine gene function because experimental evidence that can be used to assign gene function is available for only a small portion of the genes within a genome. On a whole genome level, microarrays provide a high-throughput platform to measure gene expression and thereby generate functional data for many genes simultaneously. Over the past decade, advances in genomic technologies have resulted in a variety of microarray platforms that can be used for

global gene expression profiling (Box 1). Microarrays only measure expression levels for those genes for which a probe, either a clone or sequence, is available and, therefore, should be considered a targeted approach to expression profiling. There are several different approaches that can be employed to determine gene function using expression data. The primary method is correlative using the temporal and spatial expression patterns that are associated with a specific phenotype or response. A second method involves association of unknown genes with 'known' genes based on co-expression and co-regulation.

Functional discovery through microarray expression profiling

Access to a large collection of reference expression data from mutants, tissues or treatments can provide a tool for identifying the function of unknown genes (Figure 1). This concept of functional discovery via a compendium of expression profiles was first demonstrated in yeast [7]: the expression profile of uncharacterized yeast mutants was compared to the larger expression dataset of reference profiles and matched to similar expression profiles of other conditions that displayed similar global expression changes. An alternative approach is the analysis of co-expression. An analysis of the co-expression relationships among a large quantity of expression data from humans, flies, worms and yeast identified relationships that were conserved across evolution [8]. Although the yeast genome is well annotated and is a simpler eukaryote compared with that of higher plants, with the growing availability of microarray expression data, similar approaches could also be used to analyze plants. However, additional validation might be needed and full interpretation could be more complex because of the lower quality of annotation and the poorer understanding of plant genomes compared with those of yeast. For example, a recent study used publicly available *Arabidopsis* arrays to identify genes required for cellulose synthesis through the analysis of co-regulation with cellulose synthase genes [9], thereby illustrating the potential of mining many expression profiles for functional discovery via co-regulation. A co-response database (CSB.DB) based on publicly available expression data for *Arabidopsis* is available for the identification and mining of co-regulated genes [10]. Its usefulness was illustrated by the identification of brassinosteroid-related genes by co-regulation analysis of known signaling components [11]. Additional datasets such as protein-protein interaction

Corresponding author: Rensink, W.A. (wrensink@tigr.org).

Available online 7 November 2005

Box 1. Microarray technology

Definitions

Probe: gene-specific DNA spotted on the array that will hybridize with the target.

Target: for an RNA sample, typically either total RNA or mRNA is isolated and labeled using a first strand cDNA synthesis step, either by direct incorporation of a fluorescent dye or by coupling the dyes to a modified nucleotide. For GeneChips, RNA is labeled by incorporating biotin into amplified cRNA, which is detected by fluorescently labeled streptavidin. For non-expression-based experiments, DNA rather than RNA can be labeled and hybridized to the array.

Array platforms

On-slide synthesized arrays: probes are synthesized on the array surface using DNA synthesis chemistry. The activation for oligonucleotide elongation is achieved using a mask (Affymetrix) or maskless (Nimblegen) method. Alternatively, the reagents are delivered to each spot using ink-jet technology (Agilent).

Spotted cDNA arrays: clones from a cDNA library are amplified by the polymerase chain reaction (PCR) using generic primers for the vector. PCR products are purified and spotted on glass slides using a robotic arrayer.

Spotted gene-specific sequence tag arrays: unique segment(s) of the gene are amplified from genomic DNA or bacterial artificial chromosome clones using primers specific for each gene. The PCR products are purified and spotted on glass slides using a robotic arrayer.

Spotted long oligonucleotide arrays: oligonucleotides ranging from 50–70 bases are synthesized for a unique region of the genes of interest. Oligonucleotides are spotted on glass slides using a robotic arrayer.

data, genetic interactions and metabolic data can provide additional support for networks of co-regulated genes. Functional networks have been successfully constructed in yeast using two-hybrid data and co-expression [12,13].

One consideration for integrating data into networks in any study is quality assessment of the high-throughput data. Microarray expression data can be of low technical quality, particularly for spotted arrays, which are typically processed manually, which can lead to a low level of reproducibility or false associations. Staffan Persson *et al.* [9] performed a quality analysis to remove outlier arrays before proceeding to the co-regulation analysis. Currently, most expression databases do not provide insight in to the actual quality of the expression data (see below).

***Arabidopsis* and rice: reference species and reference microarray expression data**

The *Arabidopsis* community adopted microarrays as a tool when they were first developed. The initial AG Affymetrix array represented only ~8000 unique genes [14], whereas the more recent ATH1 array represents ~23 750 unique genes [15]. It has been shown that expression data obtained with both array versions are largely similar and comparable [16]. In addition to the Affymetrix platform, long oligonucleotide arrays for *Arabidopsis* are available as well as spotted arrays with gene-specific sequence tags [17]. However, based on the predominance of literature reports and available datasets, the ATH1 array now appears to be the platform of choice for the *Arabidopsis* community. Together with ArrayExpress [18], The Nottingham *Arabidopsis* Stock Center (NASC) now provides the largest

resource for *Arabidopsis* expression data because The *Arabidopsis* Information Resource [19] recently stopped collecting expression data. Large quantities of expression data have been generated using the ATH1 array in individual studies and as part of the expression profiling service at NASC [20] (Table 1). All hybridizations and data processing are performed at NASC using user-supplied RNA. Another large *Arabidopsis* expression dataset is provided through the AtgenExpress project, which is also available through NASC. This dataset includes a gene expression map of *Arabidopsis* covering samples of many stages of development [21]. A detailed gene expression map of the *Arabidopsis* root is also available [22].

Currently, only limited amounts of expression data are available for rice. Several published studies have used relatively small-scale spotted cDNA microarrays addressing rice pollination and fertilization [23], responses to elicitors [24], and salt stress [25]. Two studies have been published that used a proprietary rice Affymetrix array to study grain filling of rice [26] and drought tolerance [27]. No whole genome expression data providing a comprehensive expression map for rice are available to date but work is in progress at the Virtual Center for Cellular Expression Profiling of Rice in which laser capture microdissection is being used for whole genome expression profiling of rice cell types (Table 1). Several public array platforms are now available for rice (Table 1). Coupled with the public availability of genome-scale rice arrays, expression databases dedicated to housing rice expression data have also been developed (Table 1).

Microarray expression profiling of crop species

Microarray platforms for crop species have been developed as well. Because the research communities for some of these species are smaller, several projects have been organized as consortia to provide a microarray expression platform for these species, either from Affymetrix or as synthesized long oligonucleotide sets (Table 1). Because most of these microarrays have become available only recently, little data are publicly available. However, studies using arrays with crop species has been reported for barley [28,29], grape [30], maize [31–34], pine [35], poplar [36], potato [37–39], tomato [40,41], soybean [42–44] and wheat [45].

Crop species-specific gene expression databases are publicly accessible. Several reference experiments using the barley 22K GeneChip [28] are available through BarleyBase [46]. In addition to the barley expression data, BarleyBase is expanding to PLEXdb, which is pursuing collection of expression data for several plant species. In addition, a tool is available for linking probes between the different species and platforms within PLEXdb (Table 1). Expression data for soybean generated using a spotted cDNA array are available through the Soybean Genomics and Microarray Database (SGMD) [47]. Large quantities of expression data for the Solanaceae, including potato and tomato, are available through the Solanaceae Gene Expression Database, and tomato gene expression data are available through a tomato expression database (Table 1). Maize gene expression data are available through the Zeamage database [48]. In the

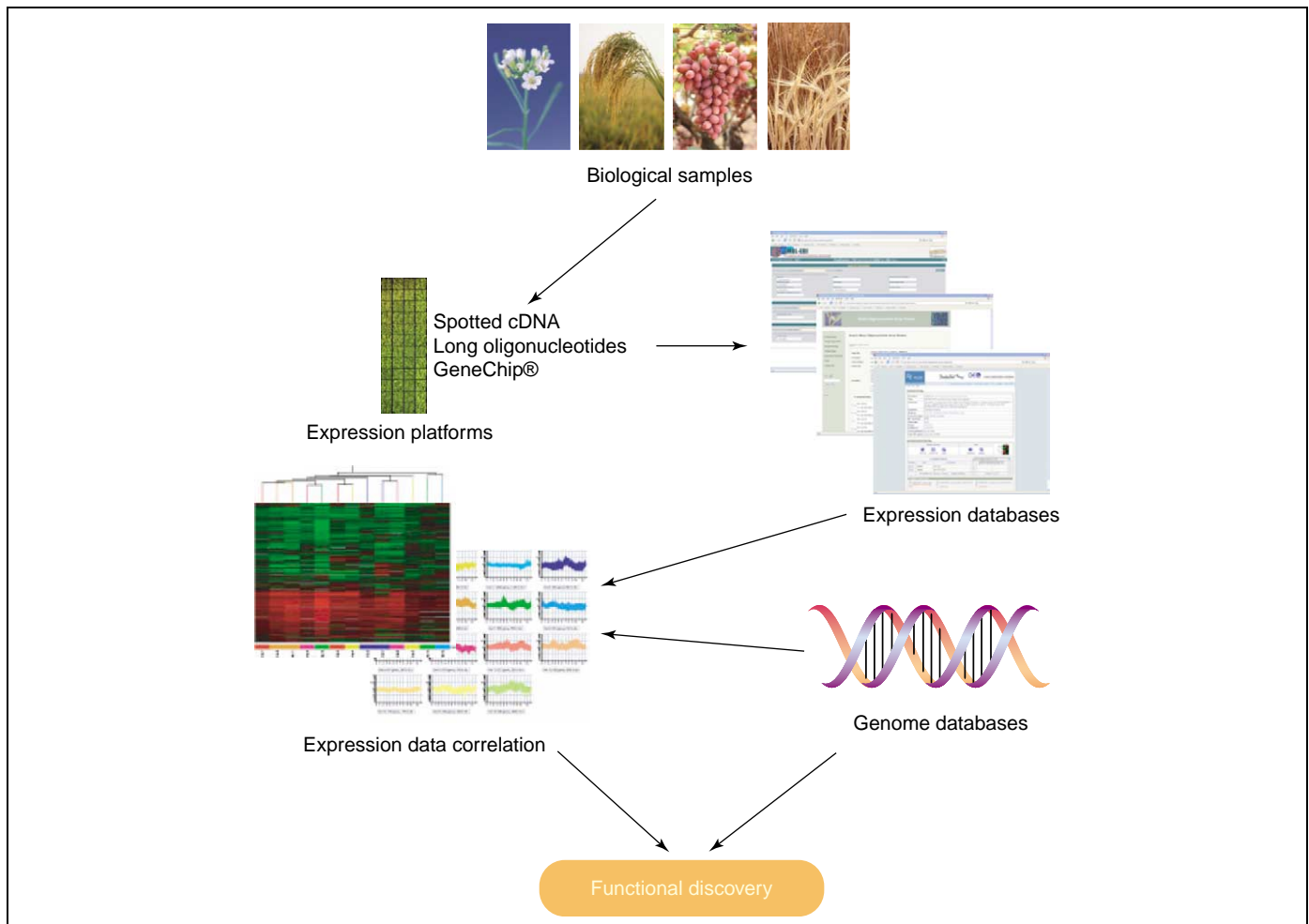


Figure 1. Functional discovery through expression profiling. Expression data from various sources and plant species provide expression data using a range of platforms. Expression databases provide access to the data. Through data analysis and integration with available genome sequences and annotation data, putative gene functions can be revealed. GeneChip is a registered trademark of Affymetrix.

near future, it can be assumed that large amounts of expression data will be generated; for data integration and cross-species comparisons, special care should be given to the identification of homologous genes and probes representing these genes.

On-line analysis tools for microarray expression data

To navigate the large quantity of data now present in expression databases, tools are being developed to identify experiments of interest as well as expression profiles for genes of interest. On-line expression analysis tools provide some capability for expression data analysis. Expression Profiler is a platform for web-based data analysis [49] in which analysis tools for gene expression data preprocessing, filtering, clustering and visualization methods are integrated with the ArrayExpress repository of expression data. Analysis tools for the *Arabidopsis* expression data are available at NASC with a more comprehensive set of analysis tools available through Genevestigator [50], which is a web-based platform for the selection and visualization of *Arabidopsis* expression data. Mapman was developed as a tool to display genomics data onto diagrams of biological processes and metabolic pathways [51]. Although the tool can be downloaded and installed locally to analyze *Arabidopsis* expression data, some

sample data are available for online analysis and display. Barleybase [46] has implemented several on-line visualization and analysis tools for statistical analysis of the data present in the database. In addition to on-line analysis tools, several open-source software packages have been developed for microarray data analysis and interpretation, such as the Bioconductor project, TM4 software and the web-based and BASE system (reviewed in [52]).

Microarray expression data integration and exchange

Even though early expression profiling studies relied mainly on publishing lists of up- or down-regulated genes, it is now well recognized that all experimental information and data associated with expression studies should be made available upon publication. Thus, any published expression study should provide enough information for other researchers to reproduce the experiment and to validate the findings. However, the complex nature of microarray expression data, such as experimental conditions, target and probe generation and relative expression ratios, prevent simple data exchange mechanisms. The MGED (Microarray Gene Expression Data) society has taken the lead in the development of standards accompanying microarray expression data [53]. Most

Table 1. Resources for publicly available plant microarray expression platforms and databases

Species	Platform	Contact
Platforms^a		
<i>Arabidopsis</i>	Spotted Oligo	http://www.ag.arizona.edu/microarray
<i>Arabidopsis</i>	Affymetrix	http://www.affymetrix.com/products/arrays/specific/arab.affx
<i>Arabidopsis</i>	Agilent	http://www.chem.agilent.com/scripts/pds.asp?IPage=33767
Barley	Affymetrix	http://www.affymetrix.com/products/arrays/specific/barley.affx
<i>Brassica</i>	Affymetrix	http://www.affymetrix.com/community/research/consortia.affx
<i>Citrus</i>	Affymetrix	http://www.affymetrix.com/community/research/consortia.affx
Grape	Affymetrix	http://www.affymetrix.com/products/arrays/specific/vitis.affx
Maize	Spotted Oligo	http://www.maizearray.org
Maize	Affymetrix	http://www.affymetrix.com/products/arrays/specific/maize.affx
<i>Medicago</i>	Affymetrix	http://www.affymetrix.com/community/research/consortia.affx
<i>Populus</i>	Affymetrix	http://www.affymetrix.com/community/research/consortia.affx
Potato	Spotted cDNA	http://www.tigr.org/tdb/potato/microarray_comp.shtml
Rice	Spotted Oligo	http://www.ricearray.org
Rice	Affymetrix	http://www.affymetrix.com/products/arrays/specific/rice.affx
Rice	Agilent	http://www.chem.agilent.com/Scripts/PDS.asp?IPage=12133
Soybean	Spotted cDNA	http://soybeangenomics.cropsci.uiuc.edu
Soybean	Affymetrix	http://www.affymetrix.com/products/arrays/specific/soybean.affx
Sugarcane	Affymetrix	http://www.affymetrix.com/products/arrays/specific/sugarcane.affx
Tomato	Spotted cDNA	http://bti.cornell.edu/CGEP/CGEP.html
Tomato	Affymetrix	http://www.affymetrix.com/community/research/consortia.affx
Wheat	Affymetrix	http://www.affymetrix.com/products/arrays/specific/wheat.affx
Expression databases		
<i>Arabidopsis</i>	NASC Arrays	http://affymetrix.Arabidopsis.info/narrays/experimentbrowse.pl
<i>Arabidopsis</i>	Genevestigator	http://www.genevestigator.ethz.ch/
Barley	Barleybase	http://www.barleybase.org
Generic	Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/projects/geo
Generic	ArrayExpress	http://www.ebi.ac.uk/arrayexpress
Generic	PLEXdb	http://www.barleybase.org/plexdb/html/index.php
Maize	Zeamage	http://www.maizearray.org
Rice	Rice Expression Database	http://red.dna.affrc.go.jp/RED
Rice	NSF Rice Oligonucleotide Array Project and Database	http://www.ricearray.org
Rice	Virtual Center for Cellular Expression Profiling in Rice	http://130.132.8.83/rc
Solanaceae	Solanaceae Gene Expression Database	http://www.tigr.org/tdb/potato
Soybean	Soybean Genomics and Microarray Database	http://psi081.ba.ars.usda.gov/SGMD/default.htm
Tomato	Tomato Expression Database	http://ted.bti.cornell.edu

^aIn addition to these publicly available arrays, synthesized oligonucleotide sets for several plant species are commercially available (<http://www.operon.com/arrays/omad.php>).

importantly, the Minimal Information About Microarray Experiments (MIAME) guidelines describe requirements for an accurate description of the biological samples used in the study as well as description of the microarray platform and the experimental procedures used to generate the expression data. Currently, most of the MIAME guidelines have been adopted by the various plant expression databases.

The majority of institutions engaged in large-scale expression studies employ expression databases either for internal data storage and/or use or for exporting the data to the public. Unfortunately, although most expression databases are MIAME-compliant, most institutions have adopted their own database schema, which complicates data exchange. As a consequence, the MGED group has developed standards for microarray data and microarray data exchange called MAGE-ML [54].

The requirement to disseminate microarray data publicly, preferably outside the authors' website, resulted in the establishment of public repositories for expression data (Table 1). Currently, there are two main repositories: the Gene Expression Omnibus (GEO) at the National Center of Biotechnology Information (NCBI) [55] and ArrayExpress at the European Bioinformatics Institute (EBI) [18]. Both repositories are organized in a similar way: first, a microarray platform is defined with spot layout and annotation for the probes on the platform;

second, expression data from experiments performed with these platforms are linked to these platform definitions. These expression data repositories contain several expression platforms for plants but are not plant-specific and contain expression data for a wide range of organisms. At GEO, 280 datasets are available for the Viridiplantae (green plants), with the bulk of the datasets derived from *Arabidopsis* (240). At neither GEO nor NCBI are the expression platforms linked by identifiers to enable cross-species expression analysis, although both repositories provide basic search capabilities and GEO allows for the retrieval of gene profiles for several experiments. Thus, these databases provide a repository and archive for expression data rather than a primary resource for expression data analysis.

Technical issues with microarrays

To make use of all available expression data for gene function discovery, data from different studies need to be integrated, requiring that the data be reliable and reproducible not only between platforms but also between laboratories. Most microarray platforms provide a reasonably robust and reproducible platform for gene expression measurements. However, not all platforms share identical probes either in sequence or type because oligonucleotides (of differing length) or PCR products can be spotted. As a consequence, low reproducibility among platforms as a

result of probe-specific effects can be observed [56]. Platforms that provide a high level of technical reproducibility result in higher quality data, an increased capability of detecting differentially expressed genes and more reproducible results among different microarray platforms [57]. An important consideration for cross-platform comparisons is the method used to 'link' the probes between the platforms because this will influence the correlation between the platforms. Sequence-based linking of gene identifiers rather than text-based matching identifiers increases cross-platform comparisons [58]. Another level of variability can be derived from between laboratory comparisons. Expression data generated with the same RNA on the same platform resulted in large differences between laboratories, but the best-performing laboratories coupled with the most robust platforms resulted in expression data with good correlation [59]. The application of standard protocols and methods among laboratories using the same platform will increase the reproducibility of expression data among laboratories [60]. Cross-platform reproducibility is best achieved when the sample labeling, hybridization and data processing is technically well performed. This was illustrated by the comparison of various microarray platforms for *Arabidopsis* that showed good correspondence among platforms of gene-specific PCR products, as well as long and short oligonucleotides [17]. If available, high-quality data generated using robust platforms with alternative probes and technology can be integrated. One complication will be the selection of datasets that meet quality standards. Thus, it can be preferable to use either datasets generated with the same platform or by the same laboratory for functional discovery of gene function. But in general, conclusions derived from expression data using alternative platforms should be reproducible with an alternative microarray expression platform. Overlying all these considerations is the need for replicate data, technical and biological, coupled with appropriate statistical analyses, something that until recently has typically been lacking in plant gene expression studies.

Tiling arrays to identify all transcripts

Another approach to study genome-wide expression is the detection of all transcriptional activity on the chromosomes using tiling arrays. Rather than using gene-specific probes to detect the expression of genes, the complete genome, including the intergenic regions is represented by probes on the array. In addition to detecting transcripts, other applications for tiling arrays include comparative genome hybridizations to detect deletions and polymorphisms, methylation profiling and the analysis of chromatin immuno-precipitation samples [61] because these types of experiments cannot be performed as well with the common expression arrays. The first plant species for which tiling expression data are available are rice and *Arabidopsis*. The *Arabidopsis* genome sequence was represented on high-density Affymetrix arrays. RNA samples were hybridized to these arrays to identify genes and gene structures experimentally [62]. For *Arabidopsis*, this resulted in the adjustment of gene models as well as the identification of novel transcripts.

A commercial version of the *Arabidopsis* tiling array will become available sometime next year. Because the genome size of rice is much larger than the genome of *Arabidopsis*, more arrays or higher density arrays are required for tiling the complete rice genome. The transcriptome of rice chromosome 10 has been analyzed using a short oligonucleotide tiling array [63] and chromosome 4 has been analyzed using a PCR-fragment-based tiling array [64]. Both studies provide evidence for expression of genes and insights into regulation of transcription along the chromosome. Owing to the lower density of these rice tiling arrays compared with those of *Arabidopsis*, the ability to refine gene structure using these arrays is limited.

An advantage of tiling arrays over expression arrays is that they are less biased for the expressed sequences because no assumption is made about the gene structure and which part of the genome sequence is coding. Another non-biased high-throughput method to determine expression levels for many genes simultaneously is provided by massively parallel signature sequencing (MPSS), which produces short sequence signatures and whose frequency is correlated to the expression level. To be maximally successful, both approaches require the availability of the genome sequence. These approaches seem to be well suited to generate a reference for gene expression in various tissues. MPSS was successfully applied to *Arabidopsis* to analyze the transcriptome in 14 libraries [65] and is currently being used to examine the rice transcriptome via 60 different libraries (Table 1).

Future perspectives

The plant community has embraced gene expression profiling technology, which is evident by the number of species for which one or more array platforms are available to date. As illustrated above, large expression databases have the potential to accelerate functional discoveries. Although these expression databases provide access to expression data and the experimental conditions used to generate the expression data, to complement genome annotation efforts effectively and to increase gene functional discovery, additional efforts are desirable. A first step would be the generation of reference expression datasets of cell, tissue or developmental-specific gene expression for each plant species. These gene expression compendia, such as AtgenExpress and the rice expression project, would provide a starting point to catalogue gene expression throughout the life cycle for each species. A second step would be the addition of gene expression data of mutants and treatments that will aid in the identification of expression and regulatory networks. A third step would be the integration of multiple data types because with the increased availability of expression platforms for multiple plant species, it is likely that the function of several unknown genes will be derived from coupling expression data with other functional and/or phenotypic data. Thus, to exploit these data fully and to embrace a complete systems approach, alternative data types such as knockout phenotypic data, metabolomics and proteomics data will need to be integrated with expression data and/or made readily available. A fourth step would be increased availability of pre-analyzed statistical and

clustering analyses because, owing to computational limitations, on-line analysis tools have limited capabilities to perform data integrations on a genome-scale. Improving large-scale databases could accelerate these efforts. Currently, plant gene expression data are scattered and stored in multiple databases, often separated between species, which inhibits cross-species comparisons. Special attention should be given to the identification of true orthologs among these plant species. These advances will be made available in the coming years not only for *Arabidopsis* and rice but also for other species. Coupled with these advances will be the improved structural annotation of the genomes and an increase in the number of genes with a known function. With the non-expression applications of arrays, even more functional information about the genome should be available, accelerating the discovery of the function of all the genes within a plant genome.

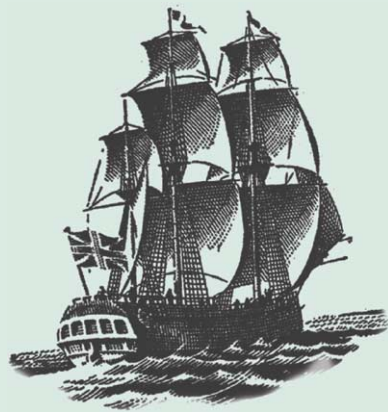
Acknowledgements

Work on plant gene expression databases at The Institute for Genomic Research is provided by funding to C.R.B. by the National Science Foundation Plant Genome Research Program (DBI 0218166, DBI 0313887 and DBI 0321663).

References

- Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
- International Rice Genome Sequencing Project. (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800
- Brunner, A.M. *et al.* (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. *Trends Plant Sci.* 9, 49–56
- Cannon, S.B. *et al.* (2005) Databases and information integration for the *Medicago truncatula* genome and transcriptome. *Plant Physiol.* 138, 38–46
- Wortman, J.R. *et al.* (2003) Annotation of the *Arabidopsis* genome. *Plant Physiol.* 132, 461–468
- Yuan, Q. *et al.* (2005) The institute for genomic research *osa1* rice genome annotation database. *Plant Physiol.* 138, 18–26
- Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
- Stuart, J.M. *et al.* (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302, 249–255
- Persson, S. *et al.* (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8633–8638
- Steinhauser, D. *et al.* (2004) CSB.DB: a comprehensive systems-biology database. *Bioinformatics* 20, 3647–3651
- Lisso, J. *et al.* (2005) Identification of brassinosteroid-related genes by means of transcript co-response analyses. *Nucleic Acids Res.* 33, 2685–2696
- Kemmeren, P. *et al.* (2005) Predicting gene function through systematic analysis and quality assessment of high-throughput data. *Bioinformatics* 21, 1644–1652
- Lee, I. *et al.* (2004) A probabilistic functional network of yeast genes. *Science* 306, 1555–1558
- Zhu, T. and Wang, X. (2000) Large-scale profiling of the *Arabidopsis* transcriptome. *Plant Physiol.* 124, 1472–1476
- Redman, J.C. *et al.* (2004) Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array. *Plant J.* 38, 545–561
- Hennig, L. *et al.* (2003) *Arabidopsis* transcript profiling on Affymetrix GeneChip arrays. *Plant Mol. Biol.* 53, 457–465
- Allemeersch, J. *et al.* (2005) Benchmarking the CATMA microarray. A novel tool for *Arabidopsis* transcriptome analysis. *Plant Physiol.* 137, 588–601
- Parkinson, H. *et al.* (2005) ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 33, D553–D555
- Rhee, S.Y. *et al.* (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* 31, 224–228
- Craigon, D.J. *et al.* (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res.* 32, D575–D577
- Schmid, M. *et al.* (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* 37, 501–506
- Birnbaum, K. *et al.* (2003) A gene expression map of the *Arabidopsis* root. *Science* 302, 1956–1960
- Lan, L. *et al.* (2004) Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10K cDNA microarray. *Plant Mol. Biol.* 54, 471–487
- Akimoto-Tomiyama, C. *et al.* (2003) Rice gene expression in response to *N*-acetylchitooligosaccharide elicitor: comprehensive analysis by DNA microarray with randomly selected ESTs. *Plant Mol. Biol.* 52, 537–551
- Kawasaki, S. *et al.* (2001) Gene expression profiles during the initial phase of salt stress in rice. *Plant Cell* 13, 889–905
- Zhu, T. *et al.* (2003) Transcriptional control of nutrient partitioning during rice grain filling. *Plant Biotechnol. J.* 1, 59–70
- Hazen, S.P. *et al.* (2005) Expression profiling of rice segregating for drought tolerance QTLs using a rice genome array. *Funct. Integr. Genomics* 5, 104–116
- Close, T.J. *et al.* (2004) A new resource for cereal genomics: 22K barley GeneChip comes of age. *Plant Physiol.* 134, 960–968
- Oztur, Z.N. *et al.* (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. *Plant Mol. Biol.* 48, 551–573
- Waters, D.L.E. *et al.* (2005) cDNA microarray analysis of developing grape (*Vitis vinifera* cv. Shiraz) berry skin. *Funct. Integr. Genomics* 5, 40–58
- Fernandes, J. *et al.* (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and microarray hybridization. *Plant Physiol.* 128, 896–910
- Cho, Y. *et al.* (2002) Gene-expression profile comparisons distinguish seven organs of maize. *Genome Biol.* 3, research0045
- Wang, H. *et al.* (2003) Temporal progression of gene expression responses to salt shock in maize roots. *Plant Mol. Biol.* 52, 873–891
- Yu, L.X. and Setter, T.L. (2003) Comparative transcriptional profiling of placenta and endosperm in developing maize kernels in response to water deficit. *Plant Physiol.* 131, 568–582
- Watkinson, J.I. *et al.* (2003) Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. *Plant Physiol.* 133, 1702–1716
- Moreau, C. *et al.* (2005) A genomic approach to investigate developmental cell death in woody tissues of *Populus* trees. *Genome Biol.* 6, R34
- Ducreux, L.J.M. *et al.* (2005) Metabolic engineering of high carotenoid potato tubers containing enhanced levels of β -carotene and lutein. *J. Exp. Bot.* 56, 81–89
- Restrepo, S. *et al.* (2005) Gene profiling of a compatible interaction between *Phytophthora infestans* and *Solanum tuberosum* suggests a role for carbonic anhydrase. *Mol. Plant-Microbe Interact.* 18, 913–922
- Rensink, W.A. *et al.* (2005) Gene expression profiling of potato responses to cold, heat, and salt stress. *Funct. Integr. Genomics* 5, 201–207
- Baxter, C.J. *et al.* (2005) Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits. *J. Exp. Bot.* 56, 1591–1604
- Eriksson, E.M. *et al.* (2004) Effect of the colorless non-ripening mutation on cell wall biochemistry and gene expression during tomato fruit development and ripening. *Plant Physiol.* 136, 4184–4197
- Moy, P. *et al.* (2004) Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Mol. Plant-Microbe Interact.* 17, 1051–1062

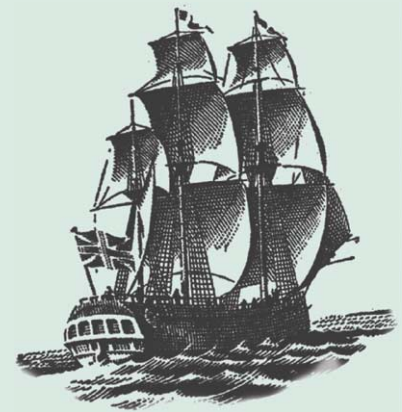
- 43 Vodkin, L. *et al.* (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. *BMC Genomics* 5, 73
- 44 Thibaud-Nissen, F. *et al.* (2003) Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. *Plant Physiol.* 132, 118–136
- 45 Clarke, B. and Rahman, S. (2005) A microarray analysis of wheat grain hardness. *Theor. Appl. Genet.* 110, 1259–1267
- 46 Shen, L. *et al.* (2005) BarleyBase – an expression profiling database for plant genomics. *Nucleic Acids Res.* 33, D614–D618
- 47 Alkharouf, N.W. and Matthews, B.F. (2004) SGMD: the Soybean Genomics and Microarray Database. *Nucleic Acids Res.* 32, D398–D400
- 48 Gardiner, J. *et al.* Design, production and utilization of long oligonucleotide microarrays for expression analysis in maize. *Maydica* (in press)
- 49 Kapushesky, M. *et al.* (2004) Expression Profiler: next generation – an online platform for analysis of microarray data. *Nucleic Acids Res.* 32, W465–W470
- 50 Zimmermann, P. *et al.* (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol.* 136, 2621–2632
- 51 Thimm, O. *et al.* (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.* 37, 914–939
- 52 Dudoit, S. *et al.* (2003) Open source software for the analysis of microarray data. *Biotechniques (Suppl)*, 45–51
- 53 Brazma, A. *et al.* (2001) Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat. Genet.* 29, 365–371
- 54 Spellman, P.T. *et al.* (2002) Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.* 3, RESEARCH0046
- 55 Barrett, T. *et al.* (2005) NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res.* 33, D562–D566
- 56 Kuo, W.P. *et al.* (2002) Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 18, 405–412
- 57 Yauk, C.L. *et al.* (2004) Comprehensive comparison of six microarray technologies. *Nucleic Acids Res.* 32, e124
- 58 Mecham, B.H. *et al.* (2004) Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res.* 32, e74
- 59 Irizarry, R.A. *et al.* (2005) Multiple-laboratory comparison of microarray platforms. *Nat. Methods* 2, 345–350
- 60 Bammler, T. *et al.* (2005) Standardizing global gene expression analysis between laboratories and across platforms. *Nat. Methods* 2, 351–356
- 61 Martienssen, R.A. *et al.* (2005) Epigenomic mapping in *Arabidopsis* using tiling microarrays. *Chromosome Res.* 13, 299–308
- 62 Yamada, K. *et al.* (2003) Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* 302, 842–846
- 63 Li, L. *et al.* (2005) Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. *Genome Biol.* 6, R52
- 64 Jiao, Y. *et al.* (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. *Plant Cell* 17, 1641–1657
- 65 Meyers, B.C. *et al.* (2004) The use of MPSS for whole-genome transcriptional analysis in *Arabidopsis*. *Genome Res.* 14, 1641–1653



Endeavour

the quarterly magazine for the history
and philosophy of science

You can access *Endeavour* online via
ScienceDirect, where you'll find a
collection of beautifully illustrated
articles on the history of science, book
reviews and editorial comment.



featuring

Selling the silver: country house libraries and the history of science by Roger Gaskell and Patricia Fara

Carl Schmidt – a chemical tourist in Victorian Britain by R. Stefan Ross

The rise, fall and resurrection of group selection by M.E. Borello

Mary Anning: the fossilist as exegete by T.W. Goodhue

Caroline Herschel: 'the unquiet heart' by M. Hoskin

Science in the 19th-century zoo by Oliver Hochadel

The melancholy of anatomy by P. Fara

and coming soon

Etienne Geoffroy St-Hillaire, Napoleon's Egyptian campaign and a theory of everything by P. Humphries

Losing it in New Guinea: The voyage of HMS *Rattlesnake* by J. Goodman

The accidental conservationist by M.A. Andrei

Powering the porter brewery by J. Sumner

Female scientists in films by B.A. Jones

and much, much more . . .

Locate *Endeavour* on *ScienceDirect* (<http://www.sciencedirect.com>)