Naturwissenschaften (2002) 89:235–249
DOI 10.1007/s00114-002-0321-3

REVIEW ARTICLE

Hauke Holtorf · Marie-Christine Guitton · Ralf Reski

Plant functional genomics

Published online: 9 April 2002

Abstract Functional genome analysis of plants has entered the high-throughput stage. The complete genome information from key species such as *Arabidopsis thaliana* and rice is now available and will further boost the application of a range of new technologies to functional plant gene analysis. To broadly assign functions to unknown genes, different fast and multiparallel approaches are currently used and developed. These new technologies are based on known methods but are adapted and improved to accommodate for comprehensive, large-scale gene analysis, i.e. such techniques are novel in the sense that their design allows researchers to analyse many genes at the same time and at an unprecedented pace. Such methods allow analysis of the different constituents of the cell that help to deduce gene function, namely the transcripts, proteins and metabolites. Similarly the phenotypic variations of entire mutant collections can now be analysed in a much faster and more efficient way than before. The different methodologies have developed to form their own fields within the functional genomics technological platform and are termed transcriptomics, proteomics, metabolomics and phenomics. Gene function, however, cannot solely be inferred by using only one such approach. Rather, it is only by bringing together all the information collected by different functional genomic tools that one will be able to unequivocally assign functions to unknown plant genes. This review focuses on current technical developments and their impact on the field of plant functional genomics. The lower plant *Physcomitrella* is introduced as a new model system for gene function analysis, owing to its high rate of homologous recombination.

H. Holtorf · M.-C. Guitton · R. Reski (✉)
Plant Biotechnology, Albert-Ludwigs-Universität Freiburg,
Sonnenstrasse 5, 79106 Freiburg, Germany
e-mail: ralf.reski@biologie.uni-freiburg.de
Tel.: +49-761-2036969, Fax: +49-761-2036967

Introduction

Large-scale genome projects have greatly changed the face of biology. Genomics has often been referred to as a new field that has led to a paradigm shift in the way science is performed. Meanwhile the post-genomic era has emerged by taking full advantage of the vast amount of genome sequence data. It has become possible to look at biology in a different way in the sense that nowadays researchers do not need to approach biological questions in a hypothesis-driven way but instead can collect and analyse data in a more non-biased and broader fashion (Rounsley and Briggs 1999; Brent 2000). New types of scientific questions can be asked and new kinds of experiments can be performed at an unprecedented pace. Recent technological advances and the rapid development of novel tools now permit the interrogation of a complete genome all at once and in a single experiment. Currently the mass of genome data is being converted into gene-function data, meaning that value is added to the nucleotide sequence collections. Knowing the exact sequence and location of all the genes of a given organism is only the first step towards understanding how all the parts of a biological system work together. In this respect functional genomics is the key approach to transforming quantity into quality. Functional genomics is a general approach toward understanding how the genes of an organism work together by assigning new functions to unknown genes. Information about the hypothesized function of an unknown gene may be deduced from its sequence structure using already known functions of similar genes as the basis for comparison. In addition, the location of given gene in the context of the chromosome may allow speculation with respect to gene function, providing that the function and chromosomal location of genes with similar sequence is known. However, to pin down the exact function of unknown genes it is necessary to understand each gene's role in the complex orchestration of all gene activities in the plant cell. Gene function analysis therefore necessitates the analysis of temporal and spatial gene expression patterns. The most conclusive information about changes

in gene expression levels can be gained from analysis of the varying qualitative and quantitative changes of messenger RNAs, proteins and metabolites. New technologies have been developed to allow fast and highly parallel measurements of these constituents of the cell that make up gene activity. The basic principles underlying the different analytical steps have been known for quite some time. But it is only the recent adaptation of these techniques to the simultaneous analysis of large numbers of genes, proteins, chemical constituents and mutant plants that has allowed the creation of the current functional genomic technology platform. Some techniques have already become routine tools for analysing gene function whereas others are still in their infancy. These new developments have led to the creation of new fields of research within functional genomics, named comparative genomics, transcriptomics, proteomics, metabolomics and phenomics. Together these approaches will allow a comprehensive and systematic functional analysis of genomes with the potential to greatly accelerate the rate of gene function prediction.

Plant functional genomics has benefited largely from genomic research performed on model organisms such as yeast, nematodes, flies and mice, not forgetting the technological spin-offs that were inspired by the human genome project. As a consequence, modern plant biology has become more and more characterized by extensive instrumentation and an increasing rate of mutant plant analysis. By the end of last year the complete genome sequence of the small weed plant *Arabidopsis thaliana* had been made accessible to the public, representing a very important cornerstone for modern plant molecular biology (The Arabidopsis Genome Initiative 2000). Recently rice, as a major staple crop species, has also been sequenced to completion. Interestingly, a large proportion of genes in these two species share no homology to genes of known or hypothesized function (Eckardt 2000). Uncovering the exact function of these genes will certainly lead to interesting new discoveries. In addition, the tremendous amount of biological diversity in plant systems will allow the identification of novel gene functions. Important functions will be matched with agronomically important traits, leading to the improvement of crop plants. This review highlights the innovative technological developments behind the different sets of multiparallel approaches that make up modern functional genomics and describes how they add a new dimension to the functional characterization of plant genomes.

Genome sequencing projects

The public effort to sequence the genome of the model flowering plant, *Arabidopsis thaliana*, had been completed by the end of last year (The Arabidopsis Genome Initiative 2000). This tiny mustard plant, a common weed of the Brassicaceae family, had been chosen as the first reference plant to be sequenced because it has several advantages over other species. Its nuclear genome is very small (120 Mbp), the generation time is short, and it is genetically very well characterized (Salanoubat et al. 2000). The deciphered genome of *Arabidopsis* was found to code for approximately 25,500 genes with one quarter of genes believed to be plant-specific. More recently, two companies have managed to finish sequencing the rice genome using a shotgun approach (Eckardt 2000). The public Rice Genome Project (RGP), pursuing a more accurate approach, is still in the process of sequencing (Sasaki and Burr 2000). In contrast to *Arabidopsis*, rice is a model monocotyledonous plant and a major crop. It also has a relatively small genome size (approximately 400–430 Mbp) (Arumuganathan and Earle 1991) and it shows a high degree of synteny, i.e. conserved gene order, with other cereals. It is estimated that the rice genome codes for 50,000 genes (Eckardt 2000). The completely annotated reference genomes of *Arabidopsis* and rice will certainly serve as a starting point for the large-scale functional analysis of other plant genomes by comparative genomics. Many other plant species have entered the stage of large-scale expressed sequence tag (EST) sequencing (Table 1). This approach of randomly selecting and sequencing a large set of cDNA clones allows to put together a collection of sequence fragments of expressed genes. A new example is the initiation of the Medicago Genome Initiative. *Medicago truncatulata*, a nitrogen-fixing legume has a close phylogenetic relationship with many agronomically important legume crop plants such as pea, soybean and lucerne (alfalfa). *M. truncatula* has a small diploid genome (450 Mbp) and will be developed into a model for legume genetics and genomics (Bell et al. 2001).

Once a plant genome has been sequenced to completion, all genes can be identified and their putative function may be deduced from algorithms which allow the comparison of sequence and presumed protein structure of an unknown gene with genes with known function. Such approaches of *in silico* gene function prediction allow the rapid identification of candidate genes that can

Table 1 A selection of internet sites of current plant EST collections

Public plant EST collections	
Soybean EST Project	http://macgrant.agron.iastate.edu/soybeanest.html
Tomato EST Project	http://www.tigr.org/tdb/Lgi/index.html
Physcomitrella EST Programme	http://www.moss.leeds.ac.uk/
Medicago Genome Initiative	http://sequence.toulouse.inra.fr/Mtruncatula.html
Pine Gene Discovery Project	http://www.cbc.umn.edu/ResearchProjects/Pine/DOE.pine/index.html
EST collections from other plant species	http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html

be analysed further. Gene function prediction by sequence comparison, however, will not always lead to conclusive results. For example, knowing the class of protein that a gene belongs to does not immediately tell researchers the exact function of a gene. Gene functions may also vary even between close members of a gene family. To date approximately 45% of all predicted *Arabidopsis* genes cannot be assigned a function (Somerville and Dangl 2000). In such cases gene function must be inferred on the basis of gene product activity, and by studying the expression of the gene in question over the course of a developmental process or in response to certain stress conditions. This will allow us to assess a biological role for the gene in question.

Forward and reverse genetic approaches to gene function

Classical forward genetics are performed by phenotypic screening of mutant populations that have been obtained by exposing wild-type plants to chemical [e.g. ethyl methanesulfonate (EMS)] or physical (X-rays) mutagens. Once an interesting phenotype has been found, this approach requires time-consuming positional cloning strategies to identify and isolate the gene lesion. Such conventional forward approaches are inefficient when it comes to performing high-throughput functional genomic analyses, despite the existence of dense genetic and physical maps in *Arabidopsis* (Camilleri et al. 1998). Random insertional mutagenesis, in contrast, allows the isolation of mutated genes in a much more direct way. Foreign DNA, either a piece of T-DNA (Krysan et al. 1999) or a transposon (Martienssen 1998), is introduced into the plant and used to disrupt genes by way of random integration. This creates loss-of-function mutants harbouring mutant alleles which are tagged by the interrupting piece of DNA. Such lines can be used for forward genetic approaches but their potential can only be exploited fully when using such lines for reverse genetics (Parinov and Sundaresan 2000). The known inserted DNA sequence allows for relatively easy isolation of any mutated gene by using screens based on PCR methodology (Sussmann et al. 2000; Ramachandran and Sundaresan 2001; Young et al. 2001). Currently databases are set up that contain sequence information of DNA stretches flanking the insertion sites (Parinov et al. 1999; Tissier et al. 1999). Accordingly, by using reverse genetic screening methods, it becomes more easy to find an *Arabidopsis* mutant with an insertion in the gene of interest. Insertional mutagenesis is very important for functional genomics in plants because direct gene knockout by homologous recombination occurs only at marginal frequencies in flowering plants (Puchta and Hohn 1996; Reski 1998a; Mengiste and Paszkowski 1999). This fact has up to now urged researchers to study higher plant gene function by using overexpression or antisense RNA suppression. Other methods that are used to produce reduction-of-function mutants take advantage of silencing-related phenomena such as co-suppression (Meins 2000), virus-induced gene silencing (Baulcombe 1999) and double-stranded RNA-mediated interference (Chuang and Meyerowitz 2000). Manipulations of known plant genes were also recently reported to be possible using chimeric oligonucleotides (Zhu et al. 2000). Plant genetic diversity is also being increased by using gene trap and activation tagging technologies (Springer 2000; Weigel et al. 2000). Furthermore, natural allelic variation among the several hundred different ecotypes of *Arabidopsis*, for example, is recognized as a rich source for the identification of new gene functions (Bressan et al. 2001).

In very many cases it has been found that *Arabidopsis* insertional knockout mutants do not show an informative phenotype. The reason for this is speculated to be functional redundancy among genes or the fact that mutations are conditional and do not readily change plant morphology (Bouche and Bouchez 2001). In such cases biological function has to be inferred from measurements of gene expression and activity. In addition, the analysis of orthologous gene knockouts in *Physcomitrella* can contribute to deciphering gene function (see below). These findings underscore the fact that functional genomic analysis depends on an array of novel methods that are producing more genetic diversity among plants and which use new techniques to accelerate analysis of transcripts, proteins and metabolites. In the following paragraphs we discuss the latest developments and applications of these novel high-throughput analysis tools.

Transcriptomics – gene expression analysis by mRNA profiling

High-throughput analysis of differential gene expression is a powerful tool for discovering novel genes or for gaining additional information about certain biological processes on a genomic scale. Transcriptional profiling using microarrays has developed into the most prominent tool for functional genomics and has convincingly demonstrated how information from raw sequence data can be converted into a broad understanding of gene function. mRNA synthesis is the first step towards protein synthesis and, for many genes, changes in mRNA abundance are related to changes in protein levels. Changes in mRNA steady state levels are mostly accomplished by changing the transcriptional rate of genes. Such fluctuations in relative mRNA amounts are indicative of changes in environment and developmental program or reflect responses to all kinds of stimuli. To properly understand a gene's function it is not only critical to know when, where, and to what extent a gene is expressed, but it is also essential to discover other genes which are co-regulated with the gene of interest. Monitoring the transcriptome, i.e. the complement of all transcribed mRNAs of an organism, by measuring mRNA concentrations of defined genes in a multiparallel and quantitative way allows to assign function to a multitude

of unknown genes. Taking the high number of scientific reports as a measure, transcriptional profiling using microarrays currently appears as the most powerful and versatile of the technologies used for functional genomics. Constant improvement of the technology will offer even more perspectives for the fast and comprehensive analysis of plant gene function.

The advantage of microarray-based expression analysis is the large number of different genes, be they of anonymous or defined sequence, that can be monitored at the same time using a small amount of biological sample material. Such genes are immobilized on a solid support and interrogated in a single experiment. The technique enables the parallel mass hybridization of labelled DNA or RNA pools from two or more samples to thousands of immobilized and arrayed DNA probes and the subsequent read-out of the number of bound target molecules. The basic principle behind transcriptional profiling with microarrays is the reverse Northern blot method. This method has been adapted for large-scale applications. In a typical array experiment the probe consists of labelled cDNA molecules which are hybridized to an array of distinct immobilized target molecules. An important technical specification of microarrays is their sensitivity. Microarray analysis of transcript levels allows the identification of low-abundance mRNAs at a copy number of only a few molecules per cell (Schena et al. 1996; Ruan et al. 1998). The different methods for the production and hybridization of microarrays have been extensively discussed in a number of reviews (Lemieux et al. 1998; Baldwin et al. 1999; Brown and Botstein 1999; Duggan et al. 1999; Eisen and Brown 1999; Lockhart and Winzeler 2000). Basically two types of microarray formats are currently used. These differ in the way the target molecules are produced and deposited on the chip surface, the nature and chemistry of the solid support, and the number of target molecules per area.

The first format uses DNA-fragment-based target molecules (Schena et al. 1995). These are either gene-specific PCR products (e.g. EST clones), plasmids or oligonucleotides. Such molecules are laid down on to the solid support surface such that spots are arranged into a highly ordered pattern. High-density arrays are produced using a gridding robot, resulting in spot densities of up to 1,000 targets per square centimetre. The solid support may either be a conventional nylon membrane or a chemically-coated microscope slide. Those using the second format are oligonucleotide-based microarrays. The target molecules (so-called “features”) are synthesized in situ on the surface of the chip using a photolithography process (Fodor et al. 1993; Lipshutz et al. 1999). The oligonucleotide sequence is selected to be gene-specific. Oligonucleotide chips can be produced with a very high density (up to 100,000 features per chip). Crucial to the success of a profiling experiment is not only the high quality of the microarray but also the experimental set-up and sampling of the tissue material for making the hybridization probe. The classical experiment involves the simultaneous measurement of the relative concentration of a given transcript in two different samples by competitive two-colour hybridization. Most experiments use reverse transcription for labelling the cDNAs of both samples. cDNA molecules are labelled with different fluorophores (e.g. fluorescein, Cy5, Cy3). After hybridization to the arrayed target molecules the relative measure of gene expression for each gene analysed is determined with the help of a high-resolution laser scanning device. Scanning is conducted using two different wavelengths, giving a quantitative fluorescence image for the two different probe populations (Lockhart and Winzeler 2000).

As more and more nucleotide sequences are being generated by an increasing number of genomic sequencing projects, the need for deposition of a growing number of genes on a given surface area becomes necessary. The apparent need for miniaturization is spurring the development of new chip formats. Meanwhile it is possible to put the entire set of genes from a fully sequenced organism on to a chip. Such expression patterns on a truly genomic scale have been obtained for yeast and other microorganisms (DeRisi et al. 1997; Wodicka et al. 1997). Since the early establishment of basic methods, this technology has been refined with respect to higher spot densities and sensitivity. Microarray technology is also being continuously developed for various genomics applications other than expression profiling (Blohm and Guiseppi-Elie 2001). Expression profiling experiments employing microarrays can be used to address diverse biological problems. By monitoring the changing expression patterns in response to a particular treatment or stimulus, one can get a complete picture of a defined biological process. Accordingly, new insights into the orchestration of certain plant responses or developmental processes can be gained. Also the complete set of genes (cluster) can be identified that is characteristic of a certain biological process. Characteristic patterns of gene expression which are derived from such clustering analyses may serve as a transcriptional fingerprint of a particular transition phase of the cell. Defining such sets of genes, which are co-regulated under defined conditions, also allows for the identification of “marker genes” that are diagnostic for certain developmental or environmental processes. The clustering of groups of genes that are co-expressed further enables the assignment of a putative function to previously uncharacterized genes by assuming that genes which can be grouped on the basis of similar expression patterns are likely to be related in function – a concept referred to as “guilt by association” (Oliver 2000).

Microarray technology stands out from other techniques which can be employed to profile transcript abundances. A variety of innovative techniques have recently been developed which alternatively offer great potential in analysing and cataloguing the identity and relative frequencies of mRNA transcripts in a given polyA(+) RNA preparation and may also serve as gene discovery tools. Such methods include differential display (DD) (Liang and Pardee 1992), serial analysis of gene expression (SAGE) (Velculescu et al. 1995), cDNA-amplified fragment length polymorphism (cDNA AFLP) display

Table 2 Websites relevant to plant functional genomics

Plant genomic databases	
The Arabidopsis Information Resource (TAIR)	http://www.arabidopsis.org/agi.html
Rice Genome Project (RGP)	http://rgp.dna.affrc.go.jp/
Arabidopsis Transposon Insertion Database	http://formaggio.cshl.org/~h-liu/attdb/index.html
Arabidopsis Knockout Facility	http://www.biotech.wisc.edu/Arabidopsis
Torrey Mesa Research Institute T-DNA collection	http://tmri.org/pages/collaborations/garlic_files/
Arabidopsis SNP-site	http://www.arabidopsis.org/cereon/index.html
Project 2010	http://www.arabidopsis.org/workshop1.html
Microarray databases	
EBI: The ArrayExpress database	http://www.ebi.ac.uk/arrayexpress/
Affymetrix	http://www.affymetrix.com/products/Arabidopsis_content.html
TIGR Arabidopsis arrays	http://atarrays.tigr.org/
AFGC	http://afgc.stanford.edu/
GARNet	http://www.york.ac.uk/res/garnet/may.htm
Rice transcriptional database	http://microarray.rice.dna.affrc.go.jp
Proteomics databases	
Proteome Analysis at EBI	http://www.ebi.ac.uk/proteome/
Arabidopsis Membrane Protein Library	http://www.cbs.umn.edu/arabidopsis/
Database of *A. thaliana* Annotation	http://luggagefast.Stanford.EDU/group/arabprotein/
ExPASy *A. thaliana* 2D-proteome database	http://www.expasy.ch/cgi-bin/map2/def?ARABIDOPSIS
PlantsP: Functional Genomics of Plant Phosphorylation	http://PlantsP.sdsc.edu/
PPMdb: Plant Plasmamembrane DataBase	http://sphinx.rug.ac.be:8080/ppmdb/index.html

(Bachem et al. 1996), total gene expression analysis (TOGA) (Sutcliffe et al. 2000), and gene calling (GC) (Shimkets et al. 1999). These protocols are largely based on methodologies that yield differentially amplified cDNA fragments by employing the polymerase chain reaction (PCR) combined with restriction digest of cDNA molecules and sequence identification of such fragments. Other methods use high-throughput sequencing approaches such as comparative expressed sequence tag (EST) sequencing of randomly chosen cDNA clones from different cDNA libraries (Adams et al. 1995; Audic and Claverie 1997), or monitor gene expression by massively parallel signature sequencing (MPSS) of cDNAs cloned in vitro on microbead arrays to obtain signature sequences for each cDNA clone of a given library (Brenner et al. 2000). Combinations of these methods have also been described (Nacht et al. 1999; Ivashuta et al. 1999). The above-mentioned techniques have successfully been applied to solve diverse problems of plant biology (Ewing et al. 1999; Matsumura et al. 1999; Bruce et al. 2000; Breyne and Zabeau 2001; Vercauteren et al. 2001). However, difficulties with reliability and reproducibility of such methods remain. Clearly, such techniques are the method of choice for plant systems for which relatively little genomic sequence information is available (Breyne and Zabeau 2001). In the future more completely sequenced genomes and large plant EST collections will be available. This fact will further increase the need for technologies that allow the analysis of very large numbers of genes simultaneously. Using microarrays for transcript profiling on a genome-wide scale will certainly become the standard approach. To date, research on the yeast and human transcriptomes has already boosted the technological progress of this method and demonstrated the general applicability of this technique. Yeast microarrays which contain all 6,200 predicted ORFs have been used to record mRNA profiles under conditions such as cell-cycle progression, diauxic shift, sporulation, alkylating treatment, and starvation (Chu et al. 1998; Spellman et al. 1998; Brown and Botstein 1999; Jelinsky and Samson 1999). Human disease- and tissue-specific gene-chips have also been investigated (Perou et al. 1999, 2000).

The usefulness of microarray technology for transcript profiling approaches in plants has been the focus of recent reviews (Baldwin et al. 1999; Richmond and Somerville 2000). For *Arabidopsis* and other plant species, different sets of chips are currently available from commercial suppliers. These chips may comprise up to 10,000 plant genes. Several publicly funded projects such as AFGC and GARNet also offer microarray services (Wisman and Ohlrogge 2000; Table 2). Early plant microarray studies were designed to demonstrate the general applicability of the multiparallel profiling technology for addressing various plant molecular biological questions (Schena et al. 1995; Desprez et al. 1998). More recent microarray analyses in plants have approached various biological questions in different plant species by interrogating an increasing number of genes (Ruan et al. 1998; Aharoni et al. 2000; Girke et al. 2000; Maleck et al. 2000; Schenk et al. 2000; Wang et al. 2000; Kawasaki et al. 2001; Schaffer et al. 2001; Seki et al. 2001; Zhu et al. 2001). To date a large number of reports have demonstrated the usefulness of microarray analysis for addressing biological questions on a global scale. However, conducting array experiments is not a trivial task as not all technical problems have been solved yet. This is why data verification is crucial. An important issue in this respect is tissue sampling. Tissue samples are often com-

posed of different cell types, making the sample a complex mixture of different transcriptional states. Moreover, the relative content of the different cell types may vary between samples. making measurements of steady state mRNA levels difficult to interpret. Another point of concern is the apparent biological variation between samples which have been harvested under the same conditions. As a consequence the sampling and probe preparation processes need to be highly standardized.

Generally, transcriptional profiling experiments have to be interpreted with caution because some biological processes lack a clear correlation between mRNA and protein levels. Instead of analysing the corresponding protein levels, however, it may help to investigate mRNA subfractions of the total mRNA pool of the cell to fine-tune microarray expression studies. The specific translational state of mRNAs can, for example, be identified by preparing specific subpools of well-translated and under-translated mRNAs and mRNAs may also be fractionated into membrane-bound and cytosolic polysomes in order to analyse the expression level of secreted and membrane-associated genes (Zong et al. 1999; Diehn et al. 2000). An essential part of microarray analysis is the way that data are collected and evaluated with image analysis software. Multiple array read-outs may generate an enormous amount of multidimensional information which has to be extracted and processed for normalization and interpretation. This can only be achieved if large-scale relational databases are connected with robust and reliable data mining tools and specialized pattern-recognition software. Although some of these tools for data mining, integration, and clustering analysis have already been developed (Eisen et al. 1998; Spellman et al. 1998; Hughes et al. 2000; Kim et al. 2000; Kerr et al. 2000) new programs are still needed in order to refine data analysis. Data interpretation problems may also arise on the basis of the DNA targets that are chosen. EST microarrays, for example, will not give a complete picture of the transcriptional state of the cell because genes which are expressed at a low level are most probably underrepresented in EST libraries. Also some degree of unspecific cross-hybridization between different PCR fragments cannot be excluded (Richmond and Somerville 2000), as in plants many genes belong to a multi-member gene family and a high incidence of gene duplications is found in *Arabidopsis* (The Arabidopsis Genome Initiative 2000). One of the biggest obstacles for comprehensive microarray data analysis is still the quality control and normalization of gene expression data between two or more experiments. Several methods are currently used to reduce the noise in array experiments and improve the quantitative evaluation process (Eickhoff et al. 1999; Beissbarth et al. 2000; Hedge et al. 2000; Schuchhardt et al. 2000). The scientific question under investigation and the experimental set-up determine the method for normalization which has to be chosen. The above-mentioned sources of variation necessitate that more than one chip experiment should be performed in order to be able to reliably interpret the data. After microarray experiments have been performed and analysed, the next analytical step must involve verification of the data on a gene-by-gene basis by using conventional methods such as quantitative PCR-based methods or RNA and protein blot analysis.

Transcriptional profiling experiments are not done fast and need a great deal of data verification to be done. Moreover, the equipment for microarray production, hybridization and read-out is still expensive. Transcriptomics, however, is already a frequently used tool for simultaneous analysis of a large number of genes. New technologies for spotting and various other technical innovations are being developed, some of which will eventually lead to considerable cost reductions in the near future. This will make the microarray technology more consumer-friendly and accessible to many more plant biologists.

Proteomics – analysing the plant proteome

The proteome has been defined as the entire complement of proteins present in a cell, organism, or tissue under defined conditions. Accordingly, proteomics is the study of this complement expressed at a given time or under certain environmental conditions (Wilkins et al. 1997). Proteomics can be divided into three main areas: (1) large-scale identification of proteins by linking structural information, such as N-terminal protein or internal peptide sequences, amino acid composition, isoelectric point, molecular mass, peptide mass fingerprints or sequence tags of selected peptides to DNA or protein databases; (2) comparison of variations in protein patterns due to stimulation by hormones, environmental changes and/or genetic mutations, followed by identification of proteins which show changes in concentration or post-translational modifications; (3) studies of protein–protein interactions using techniques such as mass spectrometry, the yeast two-hybrid system or other molecular genetic tools (Pandey and Mann 2000).

Large-scale identification of proteins involves three basic steps, beginning with the purification of small amounts of proteins from complex mixtures, digestion with a specific protease followed by peptide mass spectrometry analysis and finally computer analysis of the resulting peptide data against predicted proteolytic peptides from protein-sequence databases (Gevaert and Vandekerckhove 2000). The most popular approach is the purification of cellular fractions through two-dimensional gel electrophoresis followed by excision of individual spots, proteolytic *in gel* digestion of the isolated protein and further purification of the resulting peptides. The mass spectrometry analysis begins with peptide mapping, that is, the accurate mass measurement of the proteolytic products of the individual protein (Shevchenko et al. 1996). Each protein sequence that is available from a DNA or protein database can be virtually digested according to the specificity of the used protease and the masses of the resulting peptides are calculated and compared to the experimental data, i.e. the peptide mass profile of each protein from an individual gel spot. If the

identification is not possible because of poor recovery of peptides from the gel matrix, the analysis can be completed by mass spectrometry sequencing of selected peptides (Yates 2000). Depending on the type of mass spectrometer, a few amino acids can be depicted (sequence tag) or complete sequence can be acquired (de novo sequencing). N-terminal protein or peptide sequences can also be determined using the automated Edmann sequencing method after transfer onto a PVDF-blotting membrane (Hewick et al. 1981). Different software tools are available which allow high selectivity and specificity for the protein identification process (Fenyö 2000), provided that DNA and protein databases exist.

In plants, although database resources are limited, numerous efforts are being undertaken to explore the corresponding proteomes by using specific subcellular fractions (Table 2). The completion of the genome sequence of *Arabidopsis* allows the systematic protein identification in this organism. The proteome database issued from the translation of the genomic data can already be downloaded from the Internet. Proteomic analyses of organelles from different ecotypes are currently being organized. The Arabidopsis Plant Plasma Membrane Project is a consortium involving several laboratories which are organized with the help of the European Community. The goal of this network is to map and identify the transmembrane proteins, which represent a set of proteins particularly difficult both to extract and to resolve on 2-D gels (Santoni et al. 1998; Seigneurin-Berny et al. 1999). Up to 50% of the membrane proteins visible on 2 D gels can already be linked to gene sequences, many of which of unknown function. Another advanced project concerns the proteome of rice. Based on the partial genome data available, but also on a huge collection of EST sequences, subcellular protein mixtures are analysed using a two-dimensional liquid-chromatography system ending with online digestion directly coupled to a nanospray mass spectrometer (Yates et al. 1999). This method allows complete automation of the process, a higher sensitivity than 2-D electrophoresis and thus a high throughput. To date, more than 1,000 proteins have been identified but the resulting proteome database is proprietary. Recently, the lumenal and peripheral thylakoid proteomes from pea have been systematically investigated (Peltier et al. 2000). Both fractions contain about 200 different protein species and although protein sequence data from pea are limited, a combination of mass spectrometry analysis and Edmann sequencing has allowed the identification of 61 proteins. Other examples of plant-specific proteomes are illustrated by the construction of proteome maps of the maritime pine (Costa et al. 1999; Plomion et al. 2000), maize (Touzet et al. 1996), and the peribacteroid membrane of soybean root nodules (Panter et al. 2000). Large collections of standardized 2-D gels are available from these projects; unequivocal identification of protein, however, remains difficult due to the small set of DNA sequence data available. The same technological tools, i.e. 2-D mapping associated with mass spectrometry analysis, are used to detect variations in protein patterns between two defined states, for example between a hormone-stimulated and a control tissue, or between a wild-type and a mutated sample. The mass spectrometry analysis of *in gel*-digested proteins is not a quantitative method, and thus quantitative comparisons of different protein samples have to be done within the 2-D separation step. Multiple approaches are available: comparison of protein patterns on conventional 2-D gel electrophoresis followed by staining (Jung et al. 2000; for discussion see Gygi et al. 2000), Western blotting, or other immunoaffinity (Pandey et al. 2000; Rigaut et al. 1999) and liquid-chromatography-based methods (Nilsson and Davidsson 2000). Relative quantification of proteins or peptides using mass spectrometry analyses is possible for samples labelled with a stable isotope such as deuterium and compared with a non-labelled sample (Gygi et al. 2000).

In plants, comparative analysis of 2-D gel protein patterns is an important tool for studying the genetic and physiological differences between variants of cultivated species (Thiellement et al. 1999) and thus contributes to the definition of quantitative trait loci (QTLs). Comparative proteomic analyses have been made between green and etiolated rice shoots (Komatsu et al. 1999), and different tissues stimulated by hormone or environmental factors (Rossignol 2001). Symbiotic interactions have also been investigated (Natera et al. 2000), as well as the effect of the sink–source shift on the proteins present in the phloem (Kehr et al. 2001). In principle, variations between protein patterns are either due to differences in gene expression levels modulating protein concentration, or may result from post-translational modifications, which change the structural properties of the protein of interest. Identification and quantification of post-translational modifications can be analysed by mass spectrometry through the characteristic mass shift of the modified protein after cleavage of the modifier radicals (Kaufmann et al. 2001). Identification of phosphorylated proteins is particularly successful in plants, and a database devoted to the plant phosphorylome is already available (Table 2). Specific antibodies against phosphorylated tyrosine, serine and threonine are available and widely used to detect phosphorylated proteins (Kameyama et al. 2000; Rokka et al. 2000). The most promising method allowing us to assess the stoichiometry of in vivo protein phosphorylation in a complex sample is based on the separation of phosphorylated peptides from proteolytic digest by affinity chromatography coupled to mass spectrometry (Vener et al. 2001). The specificity of the immobilized metal affinity chromatography (IMAC) method allows not only the purification of peptides phosphorylated at tyrosine, serine, threonine, but also at histidine or asparagine residues. This is particularly important for the investigation of phosphorylation events related to plant signal transduction pathways as, unlike in mammals, phosphorylation preferentially occurs at histidine or asparagine residues (Inoue et al. 2001). As exemplified by reversible protein phosphorylation in photosynthetic membranes, the method is rapid and accurate and thus opens the possibility for precise investigations of the time-course of protein phosphorylation.

To assess protein function, the most direct tool is to study the interaction partner of the protein of interest. The proteomic approach has boosted this field and numerous new high-throughput techniques have been developed. Mass spectrometry analysis of large protein complexes such as nuclear pore (Rout et al. 2000) or spliceosome (Neubauer et al. 1998) is a powerful approach, which has led to the functional identification of numerous proteins. To date, in plants only the chloroplast RNA polymerase complex from mustard has been investigated (Pfannschmidt et al. 2000). The critical point is the availability of purification methods for the complex under study. Some complexes are stable enough to be directly purified (e.g. ribosomes), for others it is possible to use specific antibodies. Over recent years, several different teams have tried to develop so-called protein-chips in analogy with the DNA chips used in transcriptomics (Pandey and Mann 2000). Different generic methods based on the production of large libraries of fusion proteins allow for the standardized purification or expression of unknown proteins which are themselves immobilized or tested against immobilized known ligands (Rigaut et al. 1999; Rappsilber et al. 2000). The read-out of the chips can be done by MALDI-based mass spectrometry (Davies et al. 1999) or different ligand specific detection methods. Applications of the protein-chip approach are large-scale functional assays (Emili and Cagney 2000) as well as the automated yeast two-hybrid system (Ito et al. 2000) or phage display screening (Zozulya et al. 1999). In the plant field, the development of protein or peptide-fusion libraries is just beginning and essentially concerns *Arabidopsis*. The well-known green-fluorescent protein gene (GFP) was used to produce a random GFP::cDNA fusion library that has been introduced en masse by *Agrobacterium*-mediated transformation in *Arabidopsis*. The transgenic plants have been screened for the subcellular localization of the fusion proteins using the fluorescent properties of the GFP (Cutler et al. 2000). The goal is both the characterization of peptide motifs responsible for the specific targeting of proteins and the collection of protein-fusion markers allowing the study of the dynamic properties of subcellular structures. Schauser et al. (2001) have built a library of peptides constituted of peptamer bound to an inactive staphylococcal nuclease. After generation of a collection of transgenic plants expressing the peptide library, it is planned to screen for the inhibition potential of the peptides on the hypersensitive response.

In conclusion, substantial technological progress with respect to proteome analyses has been made by the plant research community as demonstrated by the numerous reports published during the last 5 years. However, while tackling the plant proteome, researchers have struggled with the same difficulties as in other biological systems. The possibility of addressing the entire complexity of protein complements of cells or tissue and of detecting the differences in the protein complements between different kinds of samples is currently still limited by the separation methods used, in particular 2-D electrophoresis. Currently high-throughput methods are available for the preparation of protein digests and mass spectrometric analyses. The next technological challenge will be to speed up and standardize the separation steps, at first through liquid chromatography processes and, in the near future, through protein-chip technology.

Metabolomics – profiling the composition of metabolites

The term metabolomics describes recent high-throughput approaches in the field of metabolic genomics that aim to identify gene function on the basis of analysing the metabolome, the full complement of metabolites of an organism. By simultaneously profiling the chemical constituents of the cell it is possible to obtain a full non-biased description of the metabolome status of an organism under defined conditions, at a given time in a given genetic background. This technology has the potential to visualize relationships that occur via regulation at the metabolic level. Unlike transcriptomic and proteomic strategies measuring changes in mRNA and protein levels, metabolic profiling allows to get closer to the specific role that a certain gene performs in a biochemical process. Additionally, novel links within the metabolic network of the cell may become apparent and novel products of side-reactions could be identified. Changed levels of certain compounds are recorded which can imply the presence or absence of gene products which are involved in synthesis or breakdown. The goal of such analyses is to determine the biochemical function of plant genes.

High-throughput metabolic screening approaches involve automated multiparallel analyses of a wide range of compounds by one single measurement. The development of reliable protocols and the number of published reports on large-scale metabolite profiling is clearly lagging behind the efforts already made in the fields of transcriptomics and proteomics. However, the undoubted potential of metabolomics is clearly recognized and currently being transformed into a new tool for comparative display of gene function in plants and other systems (Trethewey et al. 1999; Glassbrook et al. 2000; Glassbrook and Ryals 2001; Trethewey 2001). Metabolome, transcriptome and proteome analyses are similar in that they are context-dependent, i.e. the composition of proteins and metabolites varies with different states of the cell. However, metabolites, like proteins, are functional entities rather than mere transmitters of gene activity such as messenger RNAs. Estimates as to the maximum number of metabolites in the plant kingdom range between 90,000 and 200,000 different chemicals. *Arabidopsis* alone is believed to have more than 5,000 different low-molecular-weight compounds, with the majority of these compounds being products of secondary metabolism (Roessner et al. 2001). Analysing a large subset of defined or undefined metabolites is attractive because a biochemical phenotype can be detected directly by coupling chemical analysis with genetic analysis. Direct informa-

tion about a change in biological function can be linked to a mutation event. Different chemical analysis techniques can be used to rapidly and economically record the profile of small molecules and screen for relative changes in certain compounds in a plant mutant collection. Metabolite profiles are typically recorded by using techniques such as gas chromatography coupled with mass spectrometry (GC/MS) and high performance liquid chromatography (HPLC). Also the combination of GC and HPLC with time-of-flight (TOF) mass spectrometers is currently being developed for biochemical profiling applications (Glassbrook and Ryals 2001). Apart from GC/MS and HPLC, nuclear magnetic resonance (NMR) spectroscopy-based techniques have also recently been applied for the profiling of small compounds (Bligny and Douce 2001; Raamsdonk et al. 2001). The key to the applicability of these techniques for profiling approaches is the development of rapid and simple extraction and separation protocols. The fractionation methods need to guarantee that the catabolic and anabolic activities of the cells are immediately stopped at the beginning of the extraction to avoid further modification of the metabolite pool. The identification of a wide range of metabolites is done against a database of profiles of known purified compounds. To be able to find even small differences in metabolite composition between samples, different pattern recognition methods such as hierarchical cluster analysis and principal component analysis (Jurs 1986) can be applied to calculate an individual metabolite profile and compare it to other metabolite profiles. Profiles of samples which group into a defined cluster can then be used to define a "metabolic phenotype". Furthermore, correlation analysis allows to determine linkages and metabolic interactions between all compounds of a cluster. At this point the exact chemical nature of the analysed compounds need not be known, but the precise identification of compounds may become necessary in follow-up experiments. This step of analysis may then be more labour-intensive and needs to employ techniques such as MS or NMR. As with other functional genomics approaches, the interpretation of results may become problematic because of the sheer mass of data generated. The fact that biochemical pathways make up highly regulated networks adds to the complexity of the analysis. The application of new data-mining tools is therefore necessary in order to reduce the complexity of metabolite profiling read-outs.

Metabolic profiling of biological systems has not been used to a great extent in the past. Applications have been reported in human systems (Duez et al. 1996; Ning et al. 1996), but only recently has this technology been employed for studies in plants (Katona et al. 1999; Weichert et al. 1999, 2000; Fraser et al. 2000; Fiehn et al. 2000a, b; Roessner et al. 2000, 2001). The feasibility of metabolome studies for plant functional genomics has, for instance, been demonstrated by recording distinct metabolic profiles from *Arabidopsis* and potato extracts (Fiehn et al. 2000a; Roessner et al. 2001). In an attempt to demonstrate the usefulness of new metabolite profiling and extraction protocols, Roessner et al. (2001) analysed a wide range of hydrophilic compounds in potato tubers. Four independent potato transgenic genotypes were studied. These transgenic lines, through expression of heterologous enzymes, showed modifications in sucrose metabolism and were analysed for the presence of around 80 hydrophilic compounds. It was possible to assign defined clusters of compounds to each phenotype and the most important components of the clusters were identified. Correlation analysis revealed close linkages between certain metabolites in the clusters. Fiehn et al. (2000a) presented an unbiased metabolite analysis of more than 300 distinct compounds in *Arabidopsis*. The profiles allowed for the reproducible assignment of defined "metabolic phenotypes". A challenge to this technology is the variability inherent in biological systems rather than the variability inherent in the analytical procedure itself. Fiehn et al. (2000a) found a high degree of variability between samples even when genetically identical *Arabidopsis* plants were grown under identical parameters. These findings mirror the metabolic flexibility of plants in general – a phenomenon that this technology has to cope with. It is, however, clearly possible to find significant metabolic differences in different plant genotypes, enabling to distinguish defined metabolic mutants from parental wild-type background plants (Fiehn et al. 2000a).

The reported examples demonstrate the general usefulness of metabolite profiling as a novel tool for plant biology that has the potential to significantly extend and enhance the power of the already existing transcriptomic and proteomic approaches. Despite the prevailing problem of high metabolic variability among genetically identical plants, promising technological advances have been made to date. Providing the development of robust protocols and standardization methods, metabolomics will enable scientists to comprehensively characterize the complement of low-molecular-weight chemicals in plants. Plant metabolite profiling will allow the comprehensive analysis of the metabolic consequences of genetic modifications and as a result will help towards a better understanding of biochemical networks and their regulation – a prerequisite for improving the metabolic engineering of crop plants.

Phenomics – phenotypic profiling of mutant collections

The term phenomics stands for the large-scale analysis of plant diversity with respect to plant morphology. Classical genetics is largely based on the screening of collections of mutant plants followed by isolation of the mutated gene. Such screens so far have been hypothesis-driven, i.e. researchers focused primarily on the discovery of a mutant phenotype that was believed to mirror the lesion in a gene of interest. With novel global-scale technologies currently being developed for the toolbox of plant functional genomics, conventional mutant screening is now at the edge of the transition from small-scale to non-biased large-scale phenotype monitoring of mutant collections.

Such phenotypic profiling approaches aim to measure the physical and chemical properties of an organism at specific time points during its life cycle. Collecting data on a certain set of morphological and developmental traits allows researchers to put together a comprehensive picture of a mutant's phenome (Boyes et al. 2001). Such a mutant phenotype profile can then be compared to the profile of the wild-type plant, which is recorded simultaneously. As with all new global-scale analysis tools, such profiling methodologies will generate vast amounts of data which will serve as the basis for systemically studying the relationship between a mutant genotype and its respective phenotype. For plants one can imagine a whole lot of phenotypic characteristics that can be analysed. These may include growth rate, growth behaviour, timing of flowering and seed set, seed shape, leaf shape, colour changes, root density, nutrient utilization, and other deviations from wild type. The concept of phenomics is nicely illustrated by the analysis of the natural variation in light sensitivity among 141 different *Arabidopsis* accessions. Monitoring the response patterns to four light conditions, two hormone conditions, and darkness was used to identify the genes behind changes in different signalling pathways (Maloof et al. 2001). This example exemplifies phenomics on a rather small scale. Large-scale phenomic screens with thousands of mutant plants are currently being implemented by companies (Boyes et al. 2001). Once an interesting mutant is identified, the integration of phenomics data with data sets from transcriptome, proteome and metabolome analyses will help to obtain a clear picture of mutant gene function.

Physcomitrella – a tool for plant functional genomics

Bryophytes, non-vascular multicellular land plants, have been shown to be versatile tools for plant genetic research since the beginning of the century. It was recognized early on that the haploid moss protonema has great potential for genetically dissecting plant differentiation processes. Moss research, therefore, has contributed greatly to understanding plant genetics and the evolution of plant developmental processes (Reski 1998a, 1999).

While early moss studies mainly employed physiological and cytological approaches, more recent studies have used molecular genetic tools to answer biological questions. Molecular research has, for example, yielded evidence for conserved synteny of plastid DNA gene order among *Physcomitrella patens* and higher plants (Calie and Hughes 1987). Lately *Physcomitrella* has convincingly been demonstrated to be a system amenable to modern molecular and genetic analysis (Schaefer et al. 1991; Reski et al. 1994, 1998; Kruse et al. 1995; von Schwartzenberg et al. 1998; Cho et al. 1999; Kranz et al. 2000; Schulz et al. 2000). The use of moss as an appropriate model for plant biology has been underscored by analysis of a set of moss EST clones which revealed a high degree of sequence conservation at the nucleotide level between *Physcomitrella* and all types of seed plants. Furthermore, no significant difference in codon usage was found between *Physcomitrella* genes and seed plants (Reski et al. 1998; Machuka et al. 1999). Novel insights into the physiology and molecular genetics of the moss *Physcomitrella* have accumulated over recent years (Reski 1998a, 1999). The moss genome has been estimated to be only three times larger than the *Arabidopsis* genome (Reski 1999). The already known favourable characteristics of the moss, such as the small genome size and the dominant haploid gametophytic stage, have recently been turned into a striking advantage over other plant systems, when it was demonstrated that *Physcomitrella* is able to integrate transforming DNA at high frequency by way of homologous recombination – a feature that is unique among plants (Schaefer and Zryd 1997). The homologous recombination rates in moss are several orders of magnitude higher than in seed plants (Reski 1998, 1999). Accordingly, *Physcomitrella* has emerged as the only land plant that allows us to perform targeted knockout with high precision. Knockout at the DNA level correlates with a mutant phenotype that reveals the biological function of the disrupted gene (Strepp et al. 1998; Girke et al. 1998; Girod et al. 1999). The fact that the moss spends most of its life-cycle in the haploid phase allows us to directly study and find phenotypic alterations as a consequence of gene mutation. Much as in yeast, this potential can now be exploited for true reverse genetic approaches which have to date not been feasible in other plants (Reski 1998; Schaefer 2001).

Currently *Physcomitrella* is being established as a new model system for studying plant genes of unknown function. Mosses are known to contain a whole range of secondary metabolites which cannot be found in higher plants. The moss is, for instance, able to synthesize high proportions of arachidonic acid, a 20:4 fatty acid unusual in seed plants. This fact facilitates the discovery of novel plant genes from mosses such as novel desaturases and elongases (Girke et al. 1998; Sperling et al. 2000; Zank et al. 2000). Girke et al. (1998) demonstrated the use of *Physcomitrella* to discover and functionally characterize a new plant desaturase gene by knocking out the wild-type locus. In addition, moss has the advantage of being an attractive system for studying plant morphogenesis. Morphogenesis in moss protonemata can be pinpointed to single cell events. This property enables researchers to gain novel insights into the establishment of cellular polarity in plants (Bhatla et al. 2002). *Physcomitrella* also allows us to assign specific functions to genes that play roles in cytological processes. Using transient protoplast transformation, Kiessling et al. (2000) visualized a cytoskeleton-like network in chloroplasts by imaging of the FtsZ gene product using a green fluorescent protein (GFP) tag. Using the same approach, U. Richter et al. (unpublished) demonstrated dual targeting of *Physcomitrella* phage-type polymerases into mitochondria and plastids.

In collaboration with our industrial partner, BASF Plant Science, we are using *Physcomitrella* as a tool to understand plant gene function on a genomic scale. Our

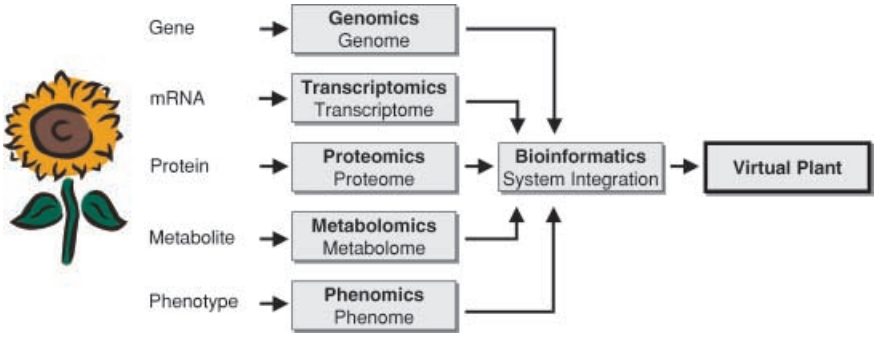

Fig. 1 Scheme showing how the integration of results from different technological levels of functional genomics leads to construction of a virtual plant

approach is based on large-scale production of moss protonema tissue in bioreactor cultures (Hohe and Reski 2002) and a proprietary EST library comprised of 110,000 clones from different tissues. Analysis of the ESTs revealed that *Physcomitrella* genome can be estimated to contain around 25,000 genes, among which are 11,000 novel plant genes (S.A. Rensing et al., unpublished). We take advantage of the system's high rate of homologous recombination to follow two routes to analyse all *Physcomitrella* genes. First, a saturated mutant collection (several thousands of individual mutant plants) is generated by using a large-scale insertional mutagenesis approach. These mutants contain moss cDNA transgenes which are disrupted by a selection cassette. The moss ESTs are mutagenized using a novel transposon shuttle mutagenesis system. This highly efficient system allows the disruption of each cDNA of the EST library by way of inserting a neomycin-phosphotransferase gene conferring kanamycin resistance to moss mutants. The aim is to transform moss with the entire complement of its genes and thereby mutate the whole moss genome. This will allow the economical, genome-wide screening for moss gene disruption phenotypes (T. Egener et al., unpublished). Second, targeted knockout mutations are generated from certain genes of interest which are derived from the EST database. Both the forward and reverse genetic approach will serve to rapidly isolate novel plant genes with the potential to improve crop plant performance. Rather than using cDNA, genomic DNA also offers itself as a basis for large-scale production of knockout constructs. Nishiyama et al. (2000) used shuttle mutagenesis of genomic library material to perform tagged mutagenesis and gene-trapping in *Physcomitrella*. A similar approach has been used to mutagenize the yeast genome, leading to the biggest collection of defined and targeted yeast mutants to date (Ross-MacDonald et al. 1999). Other groups gather new sequences from expressed moss genes. A publicly available collection of 30,000 *Physcomitrella* ESTs is currently being generated by a collaboration between researchers at Washington University and Leeds University (Table 1). The laboratory of Mitsuyasu Hasebe, Okazaki, Japan, has generated 50,000 *Physcomitrella* ESTs which are currently released into public databases.

In the future, moss functional genomics will take advantage of the whole spectrum of techniques that are currently available to study the plant transcriptome, proteome and metabolome. *Physcomitrella* will be used as a routine tool to systematically address the function of novel plant genes by reverse genetics. Studies in *Physcomitrella* will certainly complement functional analysis in other species and will add considerable value to plant genomics as a whole.

Conclusions

Plant genomics research has entered the phase of fast functional characterization of all plant genes. For efficient gene function analysis, researchers can choose from a multitude of different methods from the toolbox of plant functional genomics. Throughout recent years it has become increasingly clear, however, that each method has its inherent limitations and none of them alone suffices to unequivocally assign a function to a gene of interest. The strength of functional genomics is that it enables researchers to synergistically bring together complementary approaches. If one wants to take full benefit of the available genomic information on plant genes, only the multidisciplinary integrated approach will allow the functional characterization of plant genes. Furthermore the vast amount of data from various approaches has to be interconnected and organized into central databases in order to allow easy extraction and comparison of data for meaningful analysis. Such resources need to link information as to the sequence, genomic context, expression and mutant phenotype of a gene. All this necessitates novel bioinformatic tools to be developed. Currently the methodological groundwork is being laid out for the functional analysis of all plant genes. The goal of the ambitious "Project 2010" (Somerville and Dangl 2000) is to take full advantage of the knowledge generated by the *Arabidopsis* genome project and employ state-of-the-art functional genomic techniques to assign a function to each of the estimated 25,000 *Arabidopsis* genes. System integration of all data from different levels of functional analysis will lead to a complete blueprint of the complex network of gene activities of a plant (Vidal 2001). The vision is that in the near future it will be possible to construct a virtual plant (see Fig. 1), which will enable researchers to take a look at all gene activities at every stage of plant development.

Work on other plant species which are of immediate commercial interest will certainly benefit from future *Arabidopsis* studies. Also targeted gene knockouts in *Physcomitrella* will permit a better understanding of the resulting downstream changes. Such studies will provide plant biologists with increased knowledge about the regulation of agronomically interesting traits.

References

Adams MD, et al (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature 377[Suppl]:3–174

Aharoni A, et al (2000) Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 12:647–662

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 3:208–218

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7:986–995

Bachem CW, Hoeven RS van der, Bruijn SM de, Vreugdenhil D, Zabeau M, Visser RG (1996) Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant J 9:745–753

Baldwin D, Crane V, Rice D (1999) A comparison of gel-based, nylon filter and microarray techniques to detect differential RNA expression in plants. Curr Opin Plant Biol 2:96–103

Baulcombe DC (1999) Fast forward genetics based on virus-induced gene silencing. Curr Opin Plant Biol 2:109–113

Beissbarth T, et al (2000) Processing and quality control of DNA array hybridization data. Bioinformatics 16:1014–1022

Bell CJ, et al (2001) The *Medicago* genome initiative: a model legume database. Nucleic Acids Res 9:114–117

Bhatla SC, Kiessling J, Reski R (2002) Observation of polarity induction by cytochemical localization of phenylalkylamine-binding sites in regenerating protoplasts of the moss *Physcomitrella patens*. Protoplasma 219:99–105

Bligny R, Douce R (2001) NMR and plant metabolism. Curr Opin Plant Biol 4:191–196

Blohm DH, Guiseppi-Elie A (2001) New developments in microarray technology. Curr Opin Biotechnol 12:41–47

Bouche N, Bouchez D (2001) *Arabidopsis* gene knockout: phenotypes wanted. Curr Opin Plant Biol 4:111–117

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. Plant Cell 13:1499–1510

Brenner S, et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol 18:630–634

Brent R (2000) Genomic biology. Cell 100:169–183

Bressan RA, Zhang C, Zhang H, Hasegawa PM, Bohnert HJ, Zhu JK (2001) Learning from the *Arabidopsis* experience: the next gene search paradigm. Plant Physiol 127:1354–1360

Breyne P, Zabeau M (2001) Genome-wide expression analysis of plant cell cycle modulated genes. Curr Opin Plant Biol 4: 136–142

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. Nat Genet 21:33–37

Bruce W, Folkerts O, Garnaat C, Crasta O, Roth B, Bowen B (2000) Expression profiling of the maize flavonoid pathway genes controlled by estradiol-inducible transcription factors CRC and P. Plant Cell 12:65–79

Calie PJ, Hughes KW (1987) The consensus land plant chloroplast gene order is present with two alterations in the moss *Physcomitrella patens*. Mol Gen Genet 208:335–341

Camilleri C, et al (1998) A YAC contig map of *Arabidopsis thaliana* chromosome 3. Plant J 14:633–642

Cho SH, Chung YS, Cho SK, Rim YW, Shin JS (1999) Particle bombardment mediated transformation and GFP expression in the moss *Physcomitrella patens*. Mol Cells 9:14–19

Chu S, DeRisi J, Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I (1998) The transcriptional program of sporulation in budding yeast. Science 282:699–705

Chuang CF, Meyerowitz EM (2000) Specific and heritable genetic interference by double-stranded RNA in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 97:4985–4990

Costa P, Pionneau C, Bauw G, Dubos C, Kremer A, Frigerio JM, Plomion C (1999) Separation and characterization of needle and xylem maritime pine proteins. Electrophoresis 20:1098–1108

Cutler SR, Ehrhardt DW, Griffitts JS, Somerville CR (2000) Random GFP::cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. Proc Natl Acad Sci USA 97:3718–3723

Davies H, Lomas L, Austen B (1999) Profiling of amyloid beta peptide variants using SELDI protein chip arrays. Biotechniques 27:1258–1261

DeRisi JL, Iyer VR, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680–686

Desprez T, Amselem J, Caboche M, Hofte H (1998) Differential gene expression in *Arabidopsis* monitored using cDNA arrays. Plant J 14:643–652

Diehn M, Eisen MB, Botstein D, Brown PO (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. Nat Genet 25:58–62

Duez P, Kumps A, Mardens Y (1996) GC-MS profiling of urinary organic acids evaluated as a quantitative method. Clin Chem 42:1609–1615

Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. Nat Genet 21:10–14

Eckardt NA (2000) Sequencing the rice genome. Plant Cell 12: 2011–2017

Eickhoff B, Korn B, Schick M, Poustka A, Bosch J van der (1999) Normalization of array hybridization experiments in differential gene expression analysis. Nucleic Acids Res 27:e33

Eisen MB, Brown PO (1999) DNA arrays for analysis of gene expression. Methods Enzymol 303:179–205

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868

Emili AQ, Cagney G (2000) Large-scale functional analysis using peptide or protein arrays. Nat Biotechnol 18:393–397

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie JM (1999) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9:950–959

Fenyö D (2000) Identifying the proteome: software tools. Curr Opin Biotechnol 11:391–395

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000a) Metabolite profiling for plant functional genomics. Nat Biotechnol 18:1157–1161

Fiehn O, Kopka J, Trethewey RN, Willmitzer L (2000b) Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. Anal Chem 72:3573–3580

Fodor SP, Rava RP, Huang XC, Pease AC, Holmes CP, Adams CL (1993) Multiplexed biochemical assays with biological chips. Nature 364:555–556

Fraser PD, Pinto ME, Holloway DE, Bramley PM (2000) Technical advance: application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids. Plant J 24:551–558

Gevaert K, Vanderkerckhove J (2000) Protein identification methods in proteomics. Electrophoresis 21:1145–1154

Girke T, Schmidt H, Zahringer U, Reski R, Heinz E (1998) Identification of a novel delta 6-acyl group desaturase by targeted gene disruption in *Physcomitrella patens*. Plant J 15:39–48

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing *Arabidopsis* seeds. Plant Physiol 124:1570–1581

Girod PA, Fu H, Zryd JP, Vierstra RD (1999) Multiubiquitin chain binding subunit MCB1 (RPN10) of the 26S proteasome is essential for developmental progression in *Physcomitrella patens*. Plant Cell 11:1457–1472

Glassbrook N, Ryals J (2001) A systematic approach to biochemical profiling. Curr Opin Plant Biol 4:186–190

Glassbrook N, Beecher C, Ryals J (2000) Metabolic profiling on the right path. Nat Biotechnol 18:1142–1143

Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R (2000) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. Proc Natl Acad Sci USA 97: 9390–9395

Hedge P, et al (2000) A concise guide to cDNA microarray analysis. BioTechniques 29:548–562

Hewick RM, Hunkapiller MW, Hood LE, Dreyer WJ (1981) High-sensitivity sequencing with a gas-phase sequenator. J Biol Chem 256:7990–7997

Hohe A, Reski R (2002) Establishment of a semicontinuous bioreactor culture of *Physcomitrella patens* for mass production of protoplasts. Acta Hortic 56:425–428

Hughes TR, et al (2000) Functional discovery via a compendium of expression profiles. Cell 102:109–126

Inoue T, et al (2001) Identification of CRE1 as a cytokinin receptor from *Arabidopsis*. Nature 409:1060–1063

Ito T, et al (2000) Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. Proc Natl Acad Sci USA 97:1143–1147

Ivashuta S, Imai R, Uchiyama K, Gau M (1999) The coupling of differential display and AFLP approaches for nonradioactive mRNA fingerprinting. Mol Biotechnol 12:137–141

Jelinsky SA, Samson LD (1999) Global response of *Saccharomyces cerevisiae* to an alkylating agent. Proc Natl Acad Sci USA 96:1486–1491

Jung E, Hoogland C, Chiappe D, Sanchez JC, Hochstrasser DF (2000) The establishment of a human liver nuclei two-dimensional electrophoresis reference map. Electrophoresis 21:3483–3487

Jurs P (1986) Pattern recognition used to investigate multivariate data in analytical chemistry. Science 232:1219–1224

Kameyama K, Kishi Y, Yoshimura M, Kanzawa N, Sameshima M, Tsuchiya T (2000) Tyrosine phosphorylation in plant bending. Nature 407:37

Katona ZT, Sass P, Nolnar-Perl I (1999) Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography-mass spectrometry. Anal Biochem 266:77–84

Kaufmann H, Bailey JE, Fussenegger M (2001) Use of antibodies for detection of phosphorylated proteins separated by two-dimensional gel electrophoresis. Electrophoresis 22:194–199

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889–906

Kehr J, Haebel S, Blechschmidt-Schneider S, Willmitzer L, Steup M, Fisahn J (2001) Analysis of phloem protein patterns from different organs of *Cucurbita maxima* Duch. by matrix-assisted laser desorption/ionization time of flight mass spectroscopy combined with sodium dodecyl sulfate polyacrylamide gel electrophoresis. Planta 207:612–619

Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression micro-array data. J Comput Biol 7: 819–837

Kiessling J, Kruse S, Rensing SA, Harter K, Decker E, Reski R (2000) Visualization of a cytoskeleton-like FtsZ network in chloroplasts. J Cell Biol 151:945–950

Kim S, Dougherty ER, Chen Y, Sivakumar K, Meltzer P, Trent JM, Bittner M (2000) Multivariate measurement of gene expression relationships. Genomics 67:201–209

Komatsu S, Muhammad A, Rakwal R (1999) Separation and characterization of proteins from green and etiolated shoots of rice (*Oryza sativa* L.): towards a rice proteome. Electrophoresis 20:630–636

Kranz H, Scholz K, Weisshaar B (2000) c-MYB oncogene-like genes encoding three MYB repeats occur in all major plant lineages. Plant J 21:231–235

Kruse S, Quast S, Ipsen A, Reutter K, Reski R (1995) Cytokinin affects stability of complex plastid transcripts in cytokinin-sensitive moss mutants. Plant Sci 112:187–196

Krysan PJ, Young JC, Sussman MR (1999) T-DNA as an insertional mutagen in *Arabidopsis*. Plant Cell 11:2283–2290

Lemieux B, Aharoni A, Schena M (1998) DNA chip technology. Mol Breed 4:277–289

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971

Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. Nat Genet 21:20–24

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405:827–836

Machuka J, Bashiardes S, Ruben E, Spooner K, Cuming A, Knight C, Cove D (1999) Sequence analysis of expressed sequence tags from an ABA-treated cDNA library identifies stress response genes in the moss *Physcomitrella patens*. Plant Cell Physiol 40:378–387

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. Nat Genet 26:403–410

Maloof JN, et al (2001) Natural variation in light sensitivity of *Arabidopsis*. Nat Genet 29:441–446

Martienssen RA (1998) Functional genomics: probing plant gene function and expression with transposons. Proc Natl Acad Sci USA 95:2021–2026

Matsumura H, Nirasawa S, Terauchi R (1999) Technical advance: transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J 20:719–726

Meins F (2000) RNA degradation and models for post-transcriptional gene-silencing. Plant Mol Biol 43:261–273

Mengiste T, Paszkowski J (1999) Prospects for the precise engineering of plant genomes by homologous recombination. Biol Chem 380:749–758

Nacht M, et al. (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res 59:5464–5470

Natera HA, Guerreiro N, Djordjevic MA (2000) Proteome analysis of differentially displayed proteins as a tool for the investigation of symbiosis. Mol Plant Microbe Interact 13:995–1009

Neubauer G, et al. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. Nat Genet 20:46–50

Nilsson CL, Davidsson P (2000) New separation tools for comprehensive studies of protein expression by mass spectrometry. Mass Spectrometry Rev 19:390–397

Ning C, et al (1996) Gas chromatographic–mass spectrometric metabolic profiling of patients with fatal infantile mitochondrial myopathy with de Toni-Fanconi-Debre syndrome. Acta Paediatr Jpn 38:661–666

Nishiyama T, Hiwatashi Y, Sakakibara I, Kato M, Hasebe M (2000) Tagged mutagenesis and gene-trap in the moss, *Physcomitrella patens* by shuttle mutagenesis. DNA Res 7:9–17

Oliver S (2000) Guilt-by-association goes global. Nature 403: 601–603

Pandey A, Mann M (2000) Proteomics to study genes and genomes. Nature 405:837–846

Pandey A, Andersen JS, Mann M (2000) Use of mass spectrometry to study signaling pathways.http://www.stke.org/content/full/OC_sigtrans;2000/37/pl1

Panter S, Thomson R, Bruxelles G de, Laver D, Trevaskis B, Udvardi M (2000) Identification with proteomics of novel proteins associated with the peribacteroid membrane of soybean root nodules. Mol Plant Microbe Interact 13:325–333

Parinov S, Sundaresan V (2000) Functional genomics in *Arabidopsis*: large-scale insertional mutagenesis complements the genome sequencing project. Curr Opin Biotechnol 11:157–161

Parinov S, Sevugan M, Ye D, Yang WC, Kumaran M, Sundaresan V (1999) Analysis of flanking sequences from dissociation insertion lines: a database for reverse genetics in *Arabidopsis*. Plant Cell 11:2263–2270

Peltier JB, Friso G, Kalume DE, Roepstorff P, Nilsson F, Adamska I, Wijk KJ van (2000) Proteomics of the chloroplast: systematic identification and targeting analysis of lumenal and peripheral thylakoid proteins. Plant Cell 12:319–341

Perou CM, et al (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 96:9212–9217

Perou CM, et al (2000) Molecular portraits of human breast tumours. Nature 406:747–752

Pfannschmidt T, Orgrzewalla K, Baginsky S, Sickmann A, Meyer HE, Link G (2000) The multisubunit chloroplast RNA polymerase A from mustard (*Sinapis alba* L.): integration of a prokaryotic core into a larger complex with organelle-specific functions. Eur J Biochem 267:253–261

Plomion C, Pionneau C, Brach J, Costa P, Baillières H (2000) Compression wood-responsive proteins in developing xylem of maritime pine (*Pinus pinaster* Ait.). Plant Physiol 123: 959–969

Puchta H, Hohn B (1996) From centimorgans to base pairs – homologous recombination in plants. Trends Plant Sci 1: 340–348

Raamsdonk LM, et al (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol 19:45–50

Ramachandran S, Sundaresan V (2001) Transposons as tools for functional genomics. Plant Physiol Biochem 39:243–252

Rappsilber J, Siniossoglou S, Hurt EC, Mann M (2000) A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. Anal Chem 72:267–275

Reski R (1998a) Development, genetics and molecular biology of mosses. Bot Acta 111:1–15

Reski R (1998) *Physcomitrella* and *Arabidopsis*: the David and Goliath of reverse genetics. Trends Plant Sci 3:209–210

Reski R (1999) Molecular genetics of *Physcomitrella*. Planta 208:301–309

Reski R, Faust M, Wang XH, Wehe M, Abel WO (1994) Genome analysis of the moss *Physcomitrella patens* (Hedw.) BSG. Mol Gen Genet 244:352–359

Reski R, Reynolds S, Wehe M, Kleber-Janke T, Kruse S (1998) Moss (*Physcomitrella patens*) expressed sequence tags include several sequences which are novel for plants. Bot Acta 111: 145–151

Richmond T, Somerville S (2000) Chasing the dream: plant EST microarrays. Curr Opin Plant Biol 3:108–116

Rigaut G, Shevchenvko A, Rutz B, Wilm M, Mann M, Sépharin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotech 1:1030–1032

Roessner U, Wagner C, Kopka J, Trethewey RN, Willmitzer L (2000) Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. Plant J 23:131–142

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie AR (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13:11–29

Rokka A, Aro EM, Herrmann RG, Andersson B, Vener AV (2000) Dephosphorylation of photosystem II reaction center proteins in plant photosynthetic membranes as an immediate response to abrupt elevation of temperature. Plant Physiol 123:1525–1536

Ross-MacDonald P, et al (1999) Large-scale analysis of the yeast genome by transposon tagging and gene disruption. Nature 402:413–418

Rossignol M (2001) Analysis of the plant proteome. Curr Opin Biotech 12:131–134

Rounsley S, Briggs S (1999) The paradigm shift of genomics – a complement to traditional plant science. Curr Opin Plant Biol 2:81–82

Rout MP, Aitchison JD, Suprapto A, Hjertaas K, Zhao Y, Chait BT (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. J Cell Biol 148:635–651

Ruan Y, Gilmore J, Conner T (1998) Towards *Arabidopsis* genome analysis: monitoring expression profiles of 1,400 genes using cDNA microarrays. Plant J 15:821–833

Salanoubat M, et al (2000) Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. Nature 408:820–822

Santoni V, et al (1998) Use of a proteome strategy for tagging proteins present at the plasma membrane. Plant J 16:633–641

Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr Opin Plant Biol 3:138–141

Schaefer DG (2001) Gene targeting in *Physcomitrella patens*. Curr Opin Plant Biol 4:143–150

Schaefer DG, Zryd JP (1997) Efficient gene targeting in the moss *Physcomitrella patens*. Plant J 11:1195–1206

Schaefer DG, Zryd JP, Knight CD, Cove DJ (1991) Stable transformation of the moss *Physcomitrella patens*. Mol Gen Genet 226:418–424

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13:113–123

Schauser I, Smoker M, Jones JDG (2001) Genetic selection of inhibitors of biological pathways in plants. In 2001 Keystone symposia abstract book. Systems approach to plant biology. (B2) Keystone Symposia, Silverthorne, Colo. p 41

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1,000 genes. Proc Natl Acad Sci USA 93:10614–10619

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in arabidopsis revealed by microarray analysis. Proc Natl Acad Sci USA 97:11655–11660

Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H (2000) Normalization strategies for cDNA microarrays. Nucleic Acids Res 28:e47

Schulz P, Reski R, Maldiney R, Laloue M, Schwartzenberg K von (2000) Kinetics of cytokinin production and bud formation in *Physcomitrella*: Analysis of a wild type, a developmental mutant and two of its *ipt* transgenics. J Plant Physiol 156: 768–774

Schwartzenberg K von, Kruse S, Reski R, Moffatt B, Laloue M (1998) Cloning and characterization of an adenosine kinase from *Physcomitrella* involved in cytokinin metabolism. Plant J 13:249–257

Seigneurin-Berny D, Rolland N, Garin J, Joyard J (1999) Differential extraction of hydrophobic proteins from chloroplast envelope membranes: a subcellular-specific proteomic approach to identify rare intrinsic membrane proteins. Plant J 19: 217–228

Seki M, et al (2001) Monitoring the expression pattern of 1,300 *Arabidopsis* genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell 13:61–72

Shevchenko A, et al (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two-dimensional gels. Proc Natl Acad Sci USA 93: 14440–14445

Shimkets RA, et al (1999) Gene expression analysis by transcript profiling coupled to a gene database query. Nat Biotechnol 17:798–803

Somerville C, Dangl J (2000) Genomics: plant biology in 2010. Science 290:2077–2078

Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 9: 3273–97

Sperling P, Lee M, Girke T, Zahringer U, Stymne S, Heinz E (2000) A bifunctional delta-fatty acyl acetylenase/desaturase from the moss *Ceratodon purpureus*: a new member of the cytochrome *b*5 superfamily. Eur J Biochem 267:3801–3811

Springer PS (2000) Gene traps: tools for plant development and genomics. Plant Cell 12:1007–1020

Strepp R, Scholz S, Kruse S, Speth V, Reski R (1998) Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. Proc Natl Acad Sci USA 95:4368–4373

Sussmann MR, Amasino RM, Young JC, Krysan PJ, Austin-Phillips S (2000) The *Arabidopsis* Knockout facility at the University of Wisconsin-Madison. Plant Physiol 124:1465–1467

Sutcliffe JG, Foye PE, Erlander MG, Hilbush BS, Bodzin LJ, Durham JT, Hasel KW (2000) TOGA: an automated parsing technology for analyzing expression of nearly all genes. Proc Natl Acad Sci USA 97:1976–1981

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Thiellement H, et al (1999) Proteomics for genetic and physiological studies in plants. Electrophoresis 20:2013–2026

Tissier AF, Marillonnet S, Klimyuk V, Patel K, Torres MA, Murphy G, Jones JD (1999) Multiple independent defective suppressor-mutator transposon insertions in *Arabidopsis*: a tool for functional genomics. Plant Cell 11:1841–1852

Touzet P, Vienne D de, Huet JC, Ouali C, Bouet F, Zivy M (1996) Amino acid analysis of proteins separated by two-dimensional electrophoresis in maize: isoform detection and function identification. Electrophoresis 17:1393–1401

Trethewey RN (2001) Gene discovery via metabolic profiling. Curr Opin Biotechnol 12:135–138

Trethewey RN, Krotzky AJ, Willmitzer L (1999) Metabolic profiling: a Rosetta Stone for genomics? Curr Opin Plant Biol 2:83–85

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science 270:448–487

Vener AV, Harms A, Sussman MR, Vierstra RD (2001) Mass spectrometric resolution of reversible protein phosphorylation in photosynthetic membranes of *Arabidopsis thaliana*. J Biol Chem 276:6959–6966

Vercauteren I, Van Der Schueren E, Van Montagu M, Gheysen G (2001) *Arabidopsis thaliana* genes expressed in the early compatible interaction with root-knot nematodes. Mol Plant Microbe Interact 14:288–299

Vidal M (2001) A biological atlas of functional maps. Cell 104: 333–339

Wang R, Guegler K, LaBrie ST, Crawford NM (2000) Genomic analysis of a nutrient response in *Arabidopsis* reveals diverse expression patterns and novel metabolic and potential regulatory genes induced by nitrate. Plant Cell 12:1491–1509

Weichert H, Stenzel I, Berndt E, Wasternack C, Feussner I (1999) Metabolic profiling of oxylipins upon salicylate treatment in barley leaves–preferential induction of the reductase pathway by salicylate(1). FEBS Lett 464:133–137

Weichert H, Kohlmann M, Wasternack C, Feussner I (2000) Metabolic profiling of oxylipins upon sorbitol treatment in barley leaves. Biochem Soc Trans 28:861–862

Weigel D, et al (2000) Activation tagging in *Arabidopsis*. Plant Physiol 122:1003–1013

Wilkins MR, Williams KL, Appel RD, Hochstrasser DF (1997) Proteome research: new frontiers in functional genomics. Springer, Berlin Heidelberg New York

Wisman A, Ohlrogge J (2000) *Arabidopsis* microarray service facilities. Plant Physiol 124:1468–1471

Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol 15:1359–1367

Yates JR (2000) Mass spectrometry from genomics to proteomics. Trends Genet 16:5–8

Yates JR, Carmack E, Hays L, Link AJ, Eng JK (1999) Automated protein identification using microcolumn liquid chromatography–tandem mass spectrometry. Methods Mol Biol 112: 553–569

Young CJ, Krysan PJ, Sussmann MR (2001) Efficient screening of *Arabidopsis* T-DNA insertion lines using degenerate primers. Plant Physiol 125:513–518

Zank TK, Zahringer U, Lerchl J, Heinz E (2000) Cloning and functional expression of the first plant fatty acid elongase specific for delta-6-polyunsaturated fatty acids. Biochem Soc Trans 28:654–658

Zhu T, Mettenburg K, Peterson DJ, Tagliani L, Baszczynski CL (2000) Engineering herbicide-resistant maize using chimeric RNA/DNA oligonucleotides. Nat Biotechnol 18:555–558

Zhu T, Budworth P, Han B, Brown D, Chang H-S, Zou G, Wang X (2001) Toward elucidating the global gene expression patterns of developing *Arabidopsis*: parallel analysis of 8,300 genes by a high-density oligonucleotide probe array. Plant Physiol Biochem 39:221–242

Zong Q, Schummer M, Hood L, Morris DR (1999) Messenger RNA translation state: the second dimension of high-throughput expression screening. Proc Natl Acad Sci USA 96: 10632–10636

Zozulya S, Lioubin M, Hill RJ, Abram C, Gishizky ML (1999) Mapping signal transduction pathways by phage display. Nat Biotechnol 17:1193–1198