

# 20 The Potential of Genomics and Genetics to Understand Plant Response to Elevated Atmospheric [CO<sub>2</sub>]

G. TAYLOR, P.J. TRICKER, L.E. GRAHAM, M.J. TALLIS, A.M. RAE, H. TREWIN, and N.R. STREET

## 20.1 Introduction

### 20.1.1 What We Know and What We Need to Know

There is now a pressing need to understand more about long-term adaptation and genetic changes in future CO<sub>2</sub> concentrations, particularly for adaptive traits that are relevant to plant productivity and ecological characteristics that determine survival, fitness, yield and interaction with pests and pathogens (Ward and Kelly 2004). We wish to identify the genes that determine ecological success in future CO<sub>2</sub> environments (Feder and Mitchell-Olds 2003) and plant yields in crop systems (Martin 1989) – a subject defined as *ecological and environmental genomics*. We have an unprecedented opportunity to utilise new genomic and genetic techniques to address these questions in relation to elevated CO<sub>2</sub> using current FACE facilities. We have already quantified changes in the major characteristics determining function, productivity, yield and fitness that are sensitive to elevated CO<sub>2</sub> (Ainsworth and Long 2005), including increased photosynthesis, Rubisco acclimation, decreased plant water use and altered plant canopy architecture and leaf quality, as reported in this volume. The advantages of FACE experiments are clear – they provide realistic environmental conditions to study both short- and long-term responses of a wide range of crop, managed and natural ecosystems. The potential of new technologies is also clear. We are now able to consider the expression of many thousands of genes simultaneously, using microarrays, a technology originally developed for human disease screening (Schena et al. 1995) that has become routine in laboratory studies, but with very few field experiments on plants reported. Plant microarrays were first developed for

the model *Arabidopsis* (Schaffer et al. 2000) and this initially restricted their use, but arrays are now available for a wide range of crop plants, including maize (Fernandes et al. 2002), rice (Wasaki et al. 2003) and soybean (Vodkin et al. 2004), and for a tree, poplar (Andersson et al. 2004). Microarrays allow us to consider *gene expression* – previously only possible with mRNA blots and other forms of difficult differential expression.

Complementary to genomic approaches are those of high-throughput protein identification – proteomics and metabolic profiling (metabolomics). These technologies may be considered together in an approach that has been defined as *systems biology* (Ideker et al. 2005), or *integrative biology*, the guiding principal of which is that all constituents of a cell should be studied at once, in order to obtain a meaningful understanding of networks and controls that are operating at any given time or in response to altered conditions (Blanchard 2004). However, in its widest sense, the science of ecological and environmental genomics may also be used to consider other technologies that allow us to elucidate aspects of the genome responsible for adaptive traits (Cronk 2005). This includes the use of natural genetic variation to identify quantitative trait loci (QTL) explaining a wide range of developmental and adaptive changes that occur in response to an altered environment, as summarised in Table 20.1.

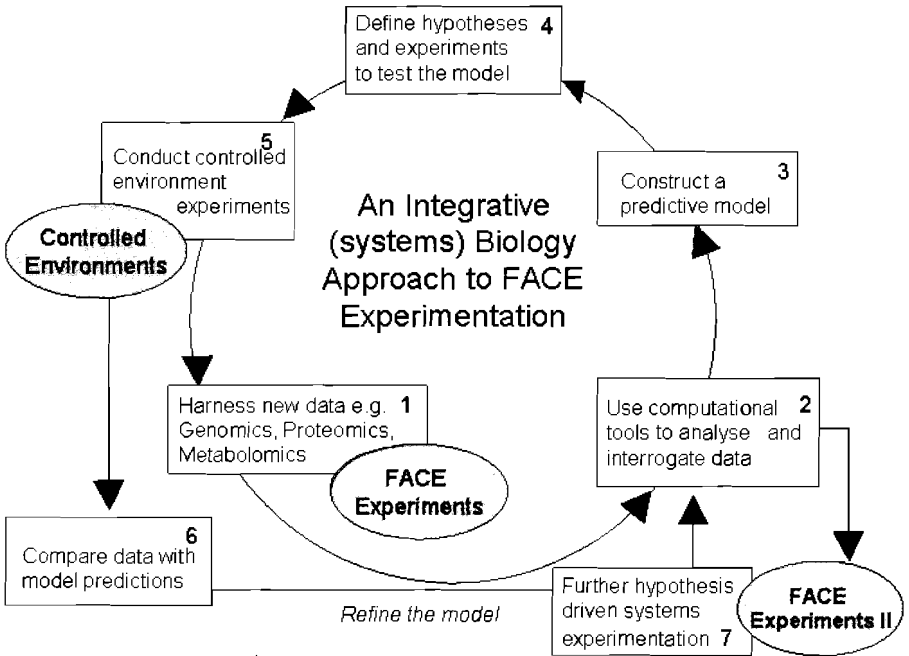
### 20.1.2 Can an Integrative (Systems) Biology Approach be Useful?

Complex biological systems should be considered as units where billions of molecules interact together to transform energy into life. Such complexity cannot easily be broken down in a reductionist approach, by studying only a single part of the system – one gene and its regulation, for example – but must be considered together and understood by integrating information from gene, protein and metabolite, as illustrated in Fig. 20.1.

Using an integrative biology approach, the FACE experiment may be seen as central to the production of large quantitative datasets at gene, protein and metabolite level (Fig. 20.1, step 1). Linking these to the development of computational models and analysis approaches (step 2), the output (step 3) is then used for hypothesis-driven experiments in controlled environments (step 4), to undertake short-term and rapid experimentation (step 5). Using experimental data (step 6), the hypothesis can then be refined with further second-stage experimentation using FACE facilities (step 7). The ability to manage this large and complex amount of information is at the forefront of current developments in bioinformatics and this is what currently limits our understanding (Thimm et al. 2004). In the remainder of this brief review, we document the major high-throughput technologies available to future FACE scientists, reviewing on-going activity in FACE experimentation and providing a glimpse of future experiments that may be possible in FACE facilities.

**Table 20.1** Genomic, genetic and other high-throughput technologies available for, but as yet largely unexploited, in FACE experiments

Technology	Description	Use in FACE experiments?
<i>Genomics</i>		
1 Microarrays – global transcript profiling	Glass-based cDNA spotted arrays or oligonucleotide-synthesised arrays, with several thousand probes representing a significant portion of the genes of an organism. Limited in the past by sequence information from species of interest. Cross-species hybridisations may be possible.	Limited use. First data available from soyFACE and POPFACE for cDNA microarrays with several thousand ESTs showing only small numbers of genes appeared sensitive to elevated CO <sub>2</sub> , with variability and reproducibility an issue (Miyazaki et al. 2004; Taylor et al. 2005)
2 Natural genetic variation and QTL discovery	The identification of areas of the genome responsible for complex traits (those generally determined by several rather than single genes). Utilises a molecular genetic map saturated with markers, e.g. SSRs and a segregating population (Rae et al. 2005).	No reported use in FACE experiments, although plans currently underway. Open-top chamber study of <i>Populus</i> (Ferris et al. 2002; Rae et al. 2005).
3 Association genetics	The utilisation of genetic variation (such as SNPs) in a natural population to find associations with phenotypic variation.	No reported use. Difficulty of placing natural populations into a FACE facility. Possible for <i>Ara-bidopsis</i> ecotypes.
<i>Proteomics</i>		
	Protein profiling. The use of 2-D gels, ICAT and AQUA for protein. Identification using mass spectrometry and isotopic signatures.	No reported use, but POPFACE samples under analysis. Controlled-environment study identified 13 proteins sensitive to elevated CO <sub>2</sub> (Bae and Sicher 2004).
<i>Metabolomics</i>		
	Gas chromatography mass spectrometry-based metabolic profiling for the identification of metabolites in tandem with PCA and hierarchical clustering techniques can reveal informative biochemical phenotypes (Fiehn et al. 2000).	No reported studies in FACE.



**Fig. 20.1** An integrated (systems) biology approach to future experimentation in FACE facilities utilising the high through-put technologies of genomics, proteomics and metabolomics

## 20.2 Genomics in Field-Grown Plants

### 20.2.1 Transcript Profiling

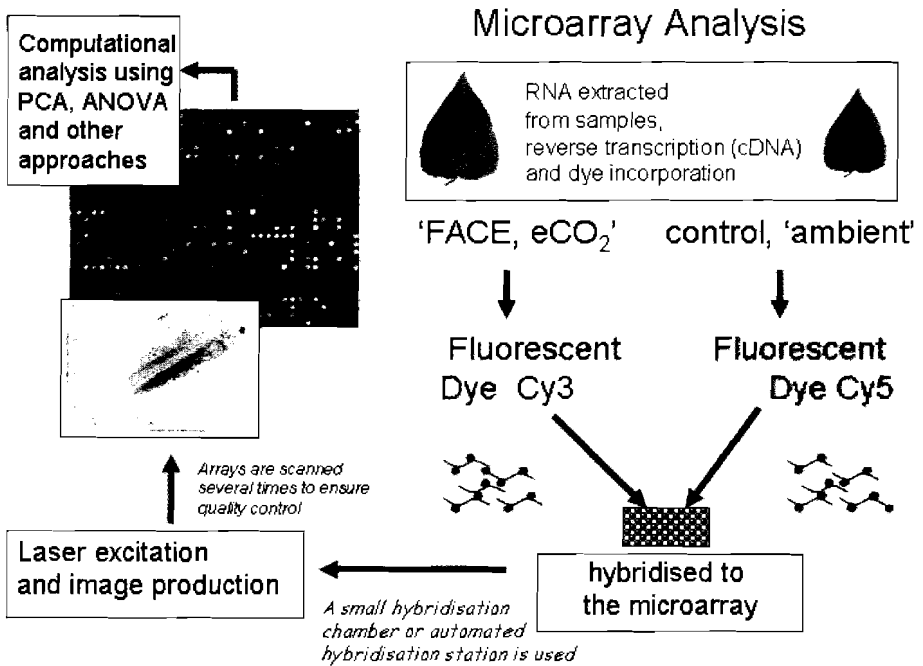
The sequencing of plant genomes and the development of genomic techniques has increased our fundamental understanding of plant growth and function. To date, genome studies have been used to discover and classify gene functions with an increasing number of plant *environmental genomic* studies completed, including plant response to stresses such as ozone (Matsuyama et al. 2002).

The introduction of the microarray has allowed realistic *global transcript profiling* to be undertaken in many replicate biological samples. Microarrays allow the parallel screening of gene expression for (potentially) all genes from an organism either at a particular time or in response to a treatment. Microarrays allow speculative investigation of gene expression in the absence of hypothesis-testing (Schrader et al. 2004). This technology is now firmly established as a valuable research tool and its use is becoming routine

(Alba et al. 2004). There are a number of variations in the production of microarrays, primarily related to how the array is spotted or synthesised, as reviewed by Deyholos and Galbraith (2001), but here we focus on the use of cDNA glass-based microarrays, since currently these offer much potential to FACE scientists, being relatively cheap to produce and use, with more and more species of interest available for spotting. The production of cDNA microarrays involves spotting probes onto a solid support. Physically, glass slides are favoured above material filters because they are solid, transparent and have low fluorescence, allowing the target direct access to the probe, and are easy to visualise. The major limitation to field biologists has been the availability of sequence information for the organisms of interest. Ideal probes are fragments of cDNA that have been sequence-validated, annotated and are unique, which show minimal cross-hybridisation to related sequences and collectively represent a comprehensive portion of the expressed genome. Typically gene expression is profiled by the competitive hybridisation of cDNA from two targets, each labelled with a different fluorescent dye (Fig. 20.2). A ratio of expression for a target at a particular probe is the desired result; and therefore the nature of the probe and the amount of the target are key considerations. In order to obtain reliable results, targets must hybridise to probes with a high degree of specificity and sensitivity. In practice, researchers are limited in their choice of probes by the EST libraries available, as the sequencing, resequencing for validation and production of cDNA clones is expensive and time-consuming.

This genome-wide approach has advantages. By assuming that genes with related functions may be co-regulated, candidate genes likely to be involved in the same processes are discovered (Eisen et al. 1998) and classes of functional genes may be identified (Golub et al. 1999). For example, cDNA microarray analysis of poplar leaves during senescence revealed that overall gene expression was up-regulated, demonstrating that senescence involves active processes rather than a down-regulation of earlier activity and that specific groups of functional genes associated with senescence may be identified (Bhalerao et al. 2003; Andersson et al. 2004).

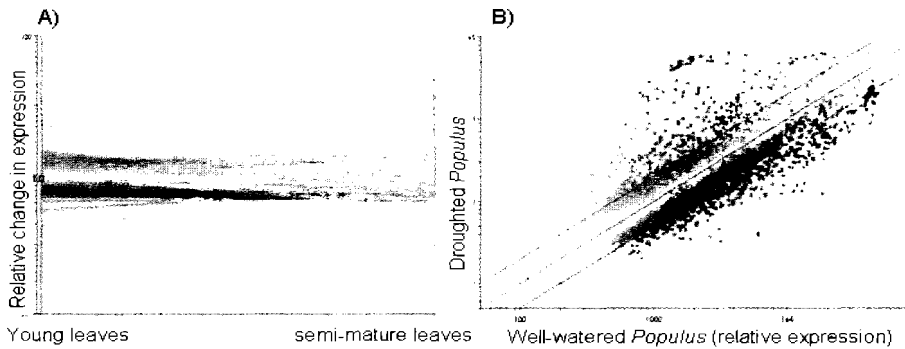
Bioinformatics – “conceptualizing biology in terms of macromolecules and then applying ‘informatics’ techniques to understand and organize the information associated with these molecules, on a large scale” (Luscombe et al. 2001) – uses information from large microarray datasets to cluster gene expression profiles into broad patterns of biological behaviour and to filter specific genes of interest (Wu 2001), as depicted in Fig. 20.2.



**Fig. 20.2** Performing a microarray hybridisation requires the extraction of high quality RNA from field material that is reverse transcribed to form cDNA. Fluorescent dyes, contrasting between treatments ('targets') are incorporated and hybridised to the 'probes' on the array, often a glass-based cDNA array, although other arrays with shorter-sequence probes exist. Following hybridisation in controlled conditions, laser excitation is used to quantify the relative incorporation of the dye with each spot, giving a measure of relative RNA expression in the two samples from *FACE* and *ambient* CO<sub>2</sub>. Large datasets are produced that require proprietary software or specialist programming. Parametric statistics, if used, must incorporate routines to avoid type 1 and type 2 statistical errors

### 20.2.2 Use of Expression Arrays in FACE Experiments

Few attempts have been made to use microarrays in current FACE facilities. In the POPFACE experiment, transcript profiling was completed in years 2, 3, 5 and 6 of exposure to elevated CO<sub>2</sub> (Taylor et al. 2005) and this approach revealed several novel results (summarised in Fig. 20.3). A consistent result was the finding in that the effect of elevated CO<sub>2</sub> on relative transcript expression depended strongly on stage of leaf development. For young leaves, differential expression suggested that genes were up-regulated in elevated compared to expression in current CO<sub>2</sub>, whilst the opposite was true for older leaves, as illustrated in Fig. 20.3. A second finding was that rather few transcripts appeared sensitive to elevated CO<sub>2</sub> (a few dozen), but this may be a



**Fig. 20.3** Transcript profiling in *Populus* trees grown in either current (350 ppm) or elevated (550 ppm) CO<sub>2</sub> for 6 years (A), or trees exposed to drought (B). A *Young leaves* depicts the changes in gene expression for elevated compared to current CO<sub>2</sub>, where each *line* is a single spot (transcript) on the array, with *red* for differential up-regulated in elevated CO<sub>2</sub> (>1.0), *green* for down-regulation in elevated CO<sub>2</sub> (<1.0) and *yellow* for no change (relative expression ~1.0). The pattern of response to elevated CO<sub>2</sub> differed depending on leaf age. B Each spot on the array is represented by a spot on the figure, with colour notations as in (A). A one-to-one and two-fold change in gene expression is depicted by the *parallel lines*. Those genes up-regulated in drought are shown in *red*

reflection of environmental variation. Transcripts of note included that for the small sub-unit of Rubisco, which was most likely to be up- and down-expressed in elevated relative to current CO<sub>2</sub> in young and semi-mature leaves respectively, a result confirmed by quantitative RT-PCR (Taylor et al. 2005). Most studies of Rubisco expression using Northern blot analysis have revealed a down-regulation of this transcript in elevated CO<sub>2</sub>, although Moore et al. (1998) also showed that up-regulation was possible, particularly in young leaves. The transcript profiling from POPFACE also revealed significant changes in the expression of xyloglucan endotransglycosylase/hydrolase (XTH) transcripts for a cell wall-loosening enzyme, known to show increased activity in elevated CO<sub>2</sub> (Ranasinghe and Taylor 1996; Ferris et al. 2001), an effect associated with increased leaf cell expansion and leaf area (Taylor et al. 2003). Interestingly, at the ASPENFACE site, transcripts for this enzyme were also up-regulated in elevated CO<sub>2</sub>, following a transcript-profiling study using Nylon membrane arrays (Gupta et al. 2005a, b). It would appear that this is an important growth mechanism in elevated CO<sub>2</sub>, but perhaps that was predictable before these hybridisations were performed. The real value of transcript profiling is in revealing previously unsuspected genes such as the calcium-dependent protein kinase that appeared to be up-regulated in elevated compared to CO<sub>2</sub>; and the RAS-related GTP-binding protein was consistently down-regulated in elevated compared to current CO<sub>2</sub> for semi-mature leaves (Taylor et al. 2005). At the soyFACE site, a preliminary transcript-profiling

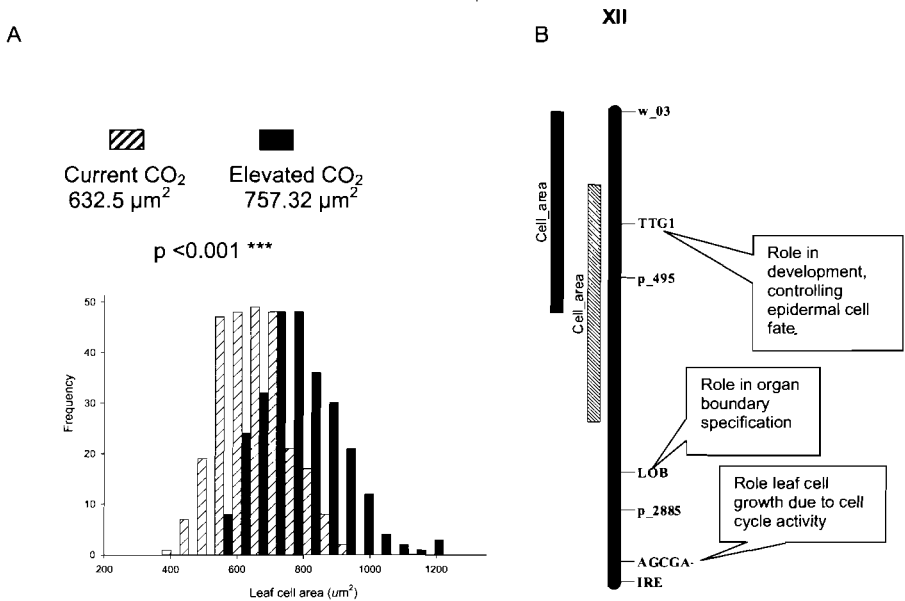
experiment using *Arabidopsis* was also reported by Miyazaki et al. (2004) and revealed that, for *Arabidopsis* at least, larger differences in gene expression were apparent between controlled environment versus field than between current versus elevated CO<sub>2</sub>. Taken together, these three preliminary studies suggest that field-grown plant material is subjected to large environmental variation on both a seasonal and daily basis and it is possible that this, combined with small changes in the expression of many genes rather than large changes in the expression of few genes, means that microarrays are difficult systems to work with in the field.

### 20.2.3 QTL Discovery for Responsive Traits

Because patterns of plant development, growth and productivity are controlled by many rather than single genes, they can be resolved at the genomic level using population and quantitative genetics, through the identification of QTL. Elucidation of QTL is not only useful for analysing important agronomic traits in crops, but also for understanding fundamental aspects of genetic control in plants, particularly model species, grown under differing conditions. It can provide evidence that a plant characteristic of interest has a genetic component and is a good start-point for future studies on individual genes and genomic regions, or in focusing on the inheritance and evolution of specific traits of interest. To our knowledge, few studies of QTL identification in elevated CO<sub>2</sub> have been published (Ferris et al. 2002). This is surprising, because the approach has yielded valuable insight into plant response to a range of environmental changes, prompting gene-cloning strategies, and identified *Arabidopsis* as a valuable model to understand the ecological significance of genetic variation (Alonso-Blanco et al. 1998).

There is considerable evidence that past changes in atmospheric elevated CO<sub>2</sub> have acted as a selection pressure, leading to altered plant development and adaptation. For example, stomatal numbers have declined since pre-industrial and across geological time-scales (Hetherington and Woodward 2003) – an effect attributed to rising atmospheric CO<sub>2</sub>. These adaptive changes are likely to have an effect on plant competitive ability and fitness. In order to use the power of quantitative genetics to unravel genetic response to global change, we need first to develop a well characterised segregating population, such as *Arabidopsis* (Lister and Dean 1993) and also *Populus* (Bradshaw et al. 1994), which must be genotyped; and, coupled to this, the whole population (in replicate) is then subjected to the conditions of interest. However, there are some technical difficulties, since the numbers involved in any mapping study tend to be large (Lister and Dean 1993).

By exposing a mapping population to elevated CO<sub>2</sub>, we have revealed these responses as well as detected the underlying QTL determining growth and development traits. Elevated [CO<sub>2</sub>] resulted in the production of larger trees,



**Fig. 20.4** The use of quantitative genetics to elucidate the genomic regions sensitive to elevated CO<sub>2</sub>, as yet a technology not utilised in FACE facilities. **A** Here, an open-top chamber experiment on *Populus* reveals that the F<sub>2</sub> progeny of an interspecific cross (*P. trichocarpa* × *P. deltoides*) had significantly larger leaf epidermal cells in elevated CO<sub>2</sub>. This information was used to determine several QTL (areas of the genome) responsible for this trait. **B** On linkage group XII of *Populus*, a QTL for cell area was found in both current and elevated CO<sub>2</sub> treatments and at least one candidate gene, taken from the physical sequence of *Populus* (TTG1 – known to determine cell patterning and fate), is found to co-locate to this region. This would be a good candidate gene for further study. The physical sequence of *Populus* may be interrogated at <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>

which is in keeping with other literature reporting increased above- and below-ground plant growth and biomass accumulation (Norby et al. 1999). Figure. 20.4 shows how leaf epidermal cell expansion was stimulated by elevated CO<sub>2</sub>, identifies a QTL for this trait (several others were present) and co-locates a putative candidate gene that may be contributing to this response. Similar responses to elevated CO<sub>2</sub> were found when the RILs of *Arabidopsis* were exposed to this treatment in four separate experiments over the course of 3 years (Rae et al. 2005).

### 20.2.4 Association Genetics

QTL discovery in segregating pedigrees relies on linkage between a QTL and a marker generated by hybridisation between the parents of a segregating

pedigree. This means that recombination, and therefore mapping resolution, is limited by the number of generations involved to produce the segregating pedigree. In general, these studies have positioned the identified QTL within chromosomal regions spanning 10–40 cM (Kearsey and Farquhar 1998). Although this level of precision may be sufficient for some applications of marker-assisted selection, to more accurately identify candidate genes for the trait of interest, the theory of association mapping has been put forward.

Association genetics utilises linkage disequilibrium (LD) to identify QTL in natural populations. LD is the non-random association between alleles, usually at linked loci (Weir 1990). This occurs when alleles at different loci occur together more often than expected; LD is a statistical measure that quantifies the non-independence of genotypes at several loci.

Interest in the study of LD has increased dramatically in recent years due to genomic technologies enabling rapid identification of haplotypes at many genetic loci, either by DNA sequencing or by high-throughput single nucleotide polymorphisms (SNP) analysis. Two main approaches can be taken in association mapping. One is the whole genome scan. In the presence of significant LD, of the order of tens of kilobases or more, it can be possible to identify genetic regions that are associated with a particular trait of interest by a high-density genome scan of individuals from an existing population. Alternatively, if LD declines rapidly in the population, a candidate gene approach may be taken where variation in genes, that are presumed to be of importance, are analysed and associated with the variation in the trait, i.e. genes are identified that may be responsible for the trait of interest by screening a limited number of candidate genes. Individual SNP haplotypes within a candidate gene are systematically tested for association with the phenotype of interest. In some cases, it should be possible to identify a polymorphism within a gene that is responsible for the difference in alternative phenotypes (Palaisa et al. 2004), sometimes termed quantitative trait nucleotides (QTN; Rafalski and Morgante 2004).

LD is expected to vary greatly in different populations because of the randomness of history, but the average rate of decay of LD (i.e. the genetic or physical distance over which LD can be measured) depends on the demographic history of the population. In particular, the extent of selfing versus outcrossing in plant populations can have a strong effect. It has been shown that LD is extensive in the mainly inbreeding species *A. thaliana*, but that it is far from genome-wide (Nordborg 2000). Association-mapping studies have been successful in human studies (e.g. Cardon et al. 1994; Fullerton et al. 2002) and livestock (Kirkpatrick and Jarne 2000), but its use is just beginning in plants (Gupta et al. 2005a, b).

Association studies can be used in a similar manner, in FACE experiments, as linkage analysis in segregating pedigrees, but the process differs in that it examines a set of presumably unrelated genotypes containing more allelic diversity and recombination events than in a typical controlled cross, therefore

reducing the time and cost of producing the pedigree and increasing the resolution and applicability. The predicted effect of the QTL should not be specific to a single family or mapping pedigree. The major limitation of LD mapping is that it provides little insight into the mechanistic basis of the LD detected (e.g. LD may not be due to linkage, but population subdivision and admixture), so that genomic localisation and cloning of genes based on LD may not always be successful. Therefore joint linkage and LD-mapping strategies have been devised (Wu et al. 2002). At present, there are no reported association studies in FACE experiments, but this should be a useful tool for future work.

### 20.3 Proteomics and Metabolomics in Field-Grown Plants

The term *genomics* is used increasingly to encompass a range of high-throughput investigative methods that examine all components of a biological state at various scales, principally transcriptomics, proteomics, and metabolomics. The term *functional genomics* describes the use of genomics to ascribe functional roles to genes and their products. Proteomics is an ever-developing technology now widely used in a range of scientific disciplines in order to analyse the proteome (the *protein* component of the *genome*) with the potential to elucidate responses to biotic and abiotic stresses, including elevated CO<sub>2</sub>. Such results complement transcriptome data and further aid the understanding of the functional identity of all plant genes existing in the genome. Microarrays are confounded by the fact that they only provide information on changes occurring at the level of transcription and do not take into account post-translational modifications that may occur before the protein is fully functional. Using proteomics in conjunction with information gained from the transcriptome further aids the understanding of biological processes and mechanisms (Tao and Aebersold 2003). In order to identify and quantify proteins correctly and accurately, appropriate methods are required that are both sensitive and powerful. There are several different approaches that have been developed for quantitative protein profiling. The two-dimensional gel electrophoresis technique (hereafter referred to as 2-DE) was primarily the most commonly used proteomic procedure. However, it is a limited technology and, like most proteomic technologies, does not identify post-translational modifications that may be of regulatory significance. This is a major limitation. The 2-DE has been used in long-term exposure of *Arabidopsis* to elevated carbon dioxide levels. In this experiment, 13 proteins were found to differ significantly in response to elevated CO<sub>2</sub> (Bae and Sicher 2004). Six of the proteins were identified by mass spectrometry (MS) and were involved in plant development or stress and photosynthesis (Bae and Sicher 2004). However, the 2-DE approach to proteomics was much criticised, since the technique requires a large amount of technical expertise to produce ade-

quate gels (Quadroni and James 1999) and is also extremely time-consuming. Furthermore, in the experiments conducted by Bae and Sicher (2004), it was difficult to identify hydrophobic proteins or those that were hard to solubilise. It is also apparent that this technique is not appropriate for detecting proteins present in low abundance. In addition to the gel-based 2-DE, isotopically labelling peptides may also be used to identify proteins. One such technique, isotope-coded affinity tags (ICAT) can be used to determine the relative amounts of proteins from cells subjected to two different conditions. The stable isotope tags used in the technique are chemically identical and simply differ in mass. The proteins from one treatment are labelled with 'heavy' tags, whilst those from treatment two are labelled with 'light' tags. The two samples are combined and, after protein digestion, they are run through a column where the tagged peptides are captured. After fractionation, tandem MS is used to identify the labelled peptides. The ICAT approach has been used for quantification and has produced some good results (Ranish et al. 2003). There have however been problems with the reagents used in the ICAT method, which are currently being further developed. Recently, a similar approach was developed, termed iTRAQ (Applied Biosystems; [www.appliedbiosystems.com](http://www.appliedbiosystems.com)). Similarly to ICAT, it is a non-gel-based approach and, following protein digestion, the peptides are tagged and analysed by tandem MS. There are a number of other non-gel-based approaches, including combined fractional diagonal chromatography (COFRADIC), which was developed in Professor Joël Vandekerckhove's laboratory (University of Ghent). The technique involves separating peptides based upon chromatographic differences and using MS for identification. Appropriate analytical methods are essential for unravelling the complex results produced from proteomic experiments. There are a number of databases constructed to aid the interpretation of the results from MS data, including MASCOT (Perkins et al. 1999), PepFrag (Qin et al. 1997) and MS-Tag (Clauser et al. 1999). Furthermore, a centralized proteomics database was recently created, with the aim of sharing experimental data with other researchers (<http://bioinformatics.icmb.utexas.edu/OPD>). Similar schemes had already been set up for transcript analysis of microarrays, such as the Stanford Microarray Database. Such apparatus will further encourage the use of proteomics for future research purposes.

Another technique involving isotopic labelling is known as absolute quantification of proteins (AQUA; <http://www.proteome.soton.ac.uk/aqua.htm>) and measures absolute protein expression (in terms of number of molecules per cell). It is a highly sensitive method which involves using stable isotopically labelled peptides (e.g.  $^{13}\text{C}$ ) as a reference. Tandem MS is again used to quantify the protein of interest by comparison with the level of the corresponding reference. Applications of this technique however are not well documented.

Proteomic approaches have been used in *Arabidopsis* and alfalfa (*Medicago sativa*) to study cell wall proteins (Chivasa et al. 2002; Watson et al. 2004).

Whilst many named cell wall proteins with known biological function have been identified using such techniques, including expansins, glucanases and peroxidases, a number of previously unnamed proteins have also been identified (Chivasa et al. 2002), thus illustrating the potential for protein discovery. Although not yet utilised extensively for protein profiling in FACE experiments, there is huge potential for this application which may be realised in the next few years. Since it has already been shown that 2-DE can be successfully used to detect in proteins in response to ozone in rice seedlings (*Oryza sativa*; Agrawal et al. 2002) and elevated CO<sub>2</sub> in *Arabidopsis* (Bae and Sicher 2004), it seems likely that future potential of protein profiling is large.

Metabolomics is becoming an important component of the holistic era of genomic investigations, although it has currently received less attention than genomic approaches. In order to understand and model complex plant responses to environment, it is essential to quantify and understand changes in the metabolome, alongside those of genome and proteome. The plant metabolome consists of the low molecular weight molecules present within cells and the study of these metabolites is termed *metabolomics* (Fiehn et al. 2000). Both primary metabolites (e.g. amino acids, fatty acids, carbohydrates) and secondary metabolites (e.g. flavonoids, terpenoids) are present within plant cells and represent end-products of gene expression. Their study alongside transcriptomics and proteomics is therefore critical to attaining an integrated biology understanding of adaptation and development. Many metabolites have critical functions in resistance and stress responses of plants and, commercially, are important as they are constituents of the taste, smell and colour of edible crops and flowers (Bino et al. 2004). The composition of cellular metabolites defines the biochemical state of a cell and the metabolome is tightly linked to the biological functioning of cells. This close relationship between the metabolome and biological function makes metabolomics an essential approach for understanding biochemical adaptation to an altered environment. In the same manner as transcriptomics and proteomics, metabolomic analysis allows patterns of co-expression to be identified and these can indicate regulons that fall under the control of single genes.

Metabolomics can be considered the biological interpretation of data derived from chemical data through the application of complex mathematical techniques. Metabolic profiling requires the identification of low molecular weight compounds, which can be achieved through gas chromatography (GC), high-performance liquid chromatography (HPLC), and nuclear magnetic resonance (NMR). Compounds are chromatographically separated and compared to a well defined sets of calibration standards. Metabolomics aims to identify metabolites present in crude extracts, using NMR, MS (including quadrupole time-of-flight MS; QTOF) and Fourier-transformed infrared spectroscopy (FT-IR). Metabolites are then identified by comparison of spectral data to that available in databases (Wagner et al. 2003) such as NIST ([www.nist.gov](http://www.nist.gov)), Wiley ([www.wileyregistry.com](http://www.wileyregistry.com)) Sigma-Aldrich (<http://www>).

[sigmaaldrich.com/Area\\_of\\_Interest/Equipment\\_Supplies\\_Books/Key\\_Resources/Spectral\\_Viewer.html](http://sigmaaldrich.com/Area_of_Interest/Equipment_Supplies_Books/Key_Resources/Spectral_Viewer.html)) and other research databases, such as the Max Planck Institute of Molecular Plant Physiology ([www.mpimp-golm.mpg.de/mms-library/index-e.html](http://www.mpimp-golm.mpg.de/mms-library/index-e.html)). At present, these databases contain only a fraction of the total metabolomic pool, representing less than 30% of identified spectral peaks (Bino et al. 2004). It is estimated that a typical *Arabidopsis* leaf will contain in the order of 5000 primary and secondary metabolites, with only 10% of these currently having functional annotation data available (Wagner et al. 2003). These 5000 represent only a fraction of the 200 000 metabolites thought to exist within the plant kingdom (Pichersky and Gang 2000). In order to attain information about the entire metabolome, combinations of the above technologies are required (Hall et al. 2002). The combination of technologies used represents a balance between speed, accuracy and resolution (Sumner et al. 2003).

## 20.4 The Importance of Experimental Design and Sampling Strategy in FACE Facilities

A crucial component of genomics is that of experimental design, data handling and data analysis. For such extensive and complicated datasets as those produced by genomics, the complexity, learning curve and essential role of data analysis cannot be understated. For FACE experiments, an added consideration is that of environmental heterogeneity, which may be considerable across both small and large temporal and spatial scales. We already know that gene expression can be affected by time of day (Michael and McClung 2003), temperature (Seki et al. 2002) and light environment (Bertrand et al. 2005) and that all these act to confound treatment effects unless a relatively stringent sampling strategy is applied, for example by 'time of day' or 'season' (including daylength). It is also critical to determine which type of parameters or 'traits' are likely to be responsive to CO<sub>2</sub> and tractable using these technologies. Most FACE experiments have some differences between rings within each treatment – 'block effects' – and correct statistical treatment of these data is essential. Biological replicates (individual plants) may be pooled in some approaches to remove some of this variation, although this should always be checked against individual replicates. Analysis is the visualisation and identification of responsive components, followed by the functional annotation of genes, proteins and metabolites and their placement within biochemical pathways. These last two steps are essential to the formation of a biological understanding of the functional contribution of metabolites. Another aspect to analysis involves the handling of quantitative data. Statistical analysis of the data is required to determine the probability that gene expression, for example, is significantly affected by the condition of interest or between the

tissues, organs, developmental stages, or genotypes under consideration. Many issues arising from the handling of large-scale datasets have been resolved or are currently being researched in relation to transcriptomics data analysis. Due to the large number of profiles that can be obtained from high-throughput ‘omics’, analytical methods to reduce the dimensionality of the data are often employed. The most common of these are principal components analysis (PCA), hierarchical cluster analysis (HCA) and K-means clustering. Sumner et al. (2003) offers a brief overview of these. As previously mentioned, examination of the functional classification of genes and their metabolites is of critical importance in order to inform a biological understanding of metabolome responses and their role in plant–environment interactions. To this end, many resources have been made available recently that enable visual interpretation of data from a range of genomics data. One such resource is a software package called MapMan (Thimm et al. 2004; <http://gabi.rzpd.de/projects/MapMan/>). Figure 20.5 shows an example image from the elevated CO<sub>2</sub> dataset that is freely available from the MapMan website. It gives an overview of cellular response mechanisms. Each block within the categories represents a gene and the expression level of that gene is colour-coded (blue for down-regulation in response to elevated CO<sub>2</sub> and red for up-regulation

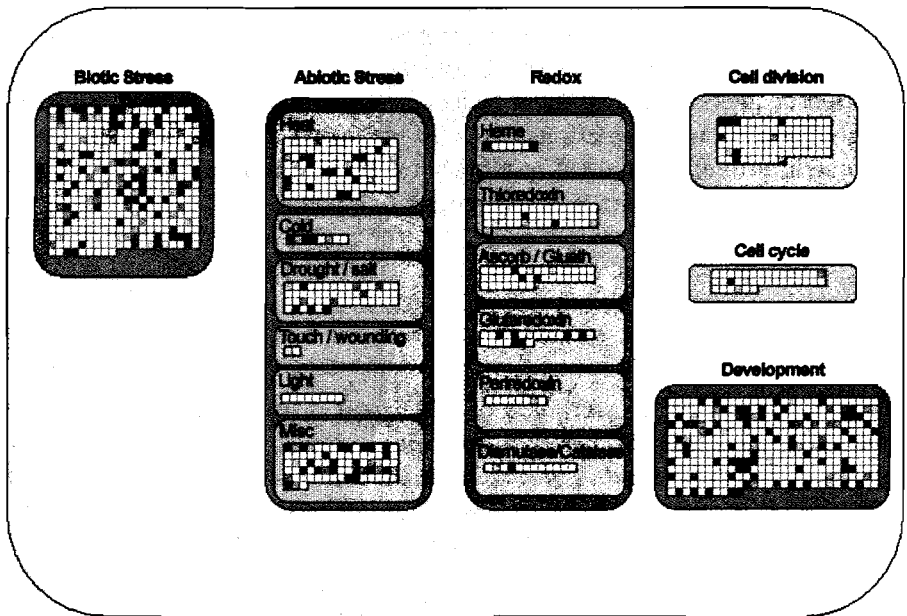


Fig. 20.5 Categories involved in cellular response mechanisms. Each *block* represents genes and the expression of the gene in response to elevated CO<sub>2</sub> is colour-coded, where *blue* represents down-regulation and *red* up-regulation. Source: <http://gabi.rzpd.de/projects/MapMan/data.shtml>

tion). Other such tools include Aracyc (<http://www.arabidopsis.org/>), Kyoto encyclopedia of genes and genomes (KEGG; <http://www.genome.jp/kegg/pathway.html>) and Gene ontology (GO; [www.geneontology.org](http://www.geneontology.org)).

As metabolomics represents an end-result of gene expression and proteome expression, its functioning and response are highly coupled to the biological state of the organism. Resolution at the metabolome level may prove to be higher than that of the proteome or the transcriptome. As identification and annotation of previously uncharacterised metabolites and their inclusion in public databases increases, metabolomics will play a central role in contributing to an integrated biology level understanding of plant–environment interactions.

## 20.5 The Future

The potential of genomics has yet to be fully realised in FACE experiments. There is no doubt that an integrative (systems) Biology approach will be pertinent over the coming years. Crop production will be altered in future climates and there is a need to link Biology to molecular crop breeding and improvement programmes, with a strong emphasis on field as opposed to controlled environment studies. In addition, FACE experiments provide a realistic environment in which to study mechanisms of genetic adaptation to future CO<sub>2</sub> concentrations. Few such studies are as yet available and the combination of genomic and genetic techniques will allow us to gain powerful insights into future adaptations. FACE experiments provide the ideal large-scale facility to allow multi-disciplinary teams to focus on important questions in ecological genomics in future high CO<sub>2</sub> conditions.

## 20.6 Conclusions

This chapter presents a summary of all current research effort on genomic and other 'omic' approaches that are being utilized in FACE experiments. Surprisingly few data are as yet available, but we describe activity and likely activity in transcriptomics, metabolomics, proteomics and molecular genetics (QTL analysis). New research to identify the genes that underlie adaptation to elevated CO<sub>2</sub> combines the transcriptomic approach with that of QTL discovery and, in future, this promises to yield new and exciting data that can only be collected at field-scale, since large replicated populations are necessary.

- New technologies and resources in molecular genetics and genomics have to date been largely unused in large-scale ecosystem manipulative experi-

ments such as FACE. Few transcriptome, proteome and metabolome studies have been undertaken. Similarly, the use of natural genetic variation to isolate QTL and genomic regions linked to adaptation to elevated CO<sub>2</sub> have been little considered, despite their potential.

- Model species, including poplar, arabidopsis, maize, rice and soybean, can currently be used to elucidate genomic response to elevated CO<sub>2</sub> because they have wide-ranging resources, including physical DNA sequence, microarrays, protein databases and molecular genetic maps. In the future, genomic resources will be extended to a wide range of ecologically relevant species. This promises to provide exciting new insights into long-term ecosystem responses to elevated CO<sub>2</sub> and ozone.
- The first transcriptome studies have revealed a small set of genes that may be sensitive to long-term exposure to elevated CO<sub>2</sub>, including genes involved in the control of growth of the plant cell wall and cell size and shape. Microarrays, however, present inherent difficulties for field-grown material which may be highly variable and where experimental design and sampling are critical in obtaining high quality data.
- QTL for adaptive responses to elevated CO<sub>2</sub> have been identified in poplar and arabidopsis; and these provide the first clues to genetic adaptation to elevated CO<sub>2</sub>. The genes underlying these QTL should be determined as a matter of priority.
- A systems biology approach to genomics should enable full integration of the environmental factors conferring a given phenotype, overcoming limitation centred on gene expression studies alone and providing novel insight into growth, development and adaptation in elevated CO<sub>2</sub>.

## References

- Agrawal GK, Rakwal R, Yonekura M, Kubo A, Saji H (2002) Proteome analysis of differentially displayed proteins as a tool for investigating ozone stress in rice (*Oryza sativa* L.) seedlings. *Proteomics* 2:947–959
- Ainsworth EA, Long SP (2005) What have we learned from 15 years of free-air CO<sub>2</sub> enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO<sub>2</sub>. *New Phytol* 165:351–371
- Alba R, Fei ZJ, Payton P, Liu Y, Moore SL, Debbie P, Cohn J, D'Ascenzo M, Gordon JS, Rose JKC, Martin G, Tanksley SD, Bouzayen M, Jahn MM, Giovannoni J (2004) ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. *Plant J* 39: 697–714
- Alonso-Blanco C, El-Assal SE, Coupland G, Koornneef M (1998) Analysis of natural allelic variation at flowering time loci in the landsberg erecta and Cape Verde Islands ecotypes of *Arabidopsis thaliana*. *Genetics* 149:749–764
- Andersson A, Keskitalo J, Sjodin A, Bhalerao R, Sterky F, Wissel K, Tandré K, Aspeborg H, Moyle R, Ohmiya Y, Bhalerao R, Brunner A, Gustafsson P, Karlsson J, Lundeberg J, Nilsson O, Sandberg G, Strauss S, Sundberg B, Uhlen M, Jansson S, Nilsson (2004) A transcriptional timetable of autumn senescence. *Genome Biol* 5:R24

- Bae H, Sicher R (2004) Changes in soluble protein expression and leaf metabolite levels in *Arabidopsis thaliana* grown in elevated carbon dioxide. *Field Crop Res* 90:61–73
- Bertrand C, Benhamed M, Li YF, Ayadi M, Lemonnier G, Renou JP, Delarue M, Zhou DX (2005) *Arabidopsis* HAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1, is required to integrate light signals to regulate gene expression and growth. *J Biol Chem* 280:1465–1473
- Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Jonsson Birve S, Karlsson J, Gardstrom P, Gustafsson P, Lundeberg J, Jansson S (2003) Gene expression in autumn leaves. *Plant Physiol* 131:1–13
- Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9:418–425
- Blanchard JL (2004) Bioinformatics and systems biology, rapidly evolving tools for interpreting plant response to global change. *Field Crops Res* 90:117–131
- Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theor Appl Genet* 89:167–178
- Cardon LR, Fulker DW (1994) The power of interval mapping of quantitative trait loci using selected sib pairs. *Am J Hum Genet* 55:825–833
- Chivasa S, Ndimba BK, Simon WJ, Robertson D, Yu XL, Knox JP, Bolwell P, Slabas AR (2002) Proteomic analysis of the *Arabidopsis thaliana* cell wall. *Electrophoresis* 23:1754–1765
- Clauser, K.R., Baker, P., and Burlingame, A.L. (1999) Role of accurate mass measurement ( $\pm 10$  ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal Chem* 71:2871–2882
- Cronk QCB (2005) Plant eco-devo: the potential of poplar as a model organism. *New Phytol* 166:39–48
- Deyholos MK, Galbraith DW (2001) High-density microarrays for gene expression analysis. *Cytometry* 43:229–238
- 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
- Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature* 4:649–655
- Fernandes J, Brendel V, Gai XW, Lal S, Chandler VL, Elumalai P, Galbraith DW, Pierson EA, Walbot V (2002) Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. *Plant Physiol* 128:896–910
- Ferris R, Sabatti M, Miglietta F, Mills RF, Taylor G (2001) Leaf area is stimulated in *Populus* by free air CO<sub>2</sub> enrichment (POPFACE), through cell expansion and production. *Plant Cell Environ* 24:305–315
- Ferris R, Long L, Bunn SM, Robinson KM, Bradshaw HD, Rae AM, Taylor G (2002) Leaf stomatal and epidermal cell development: identification of putative quantitative trait loci in relation to elevated carbon dioxide concentration in poplar. *Tree Physiol* 22:633–640
- Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. *Nat Biotechnol* 18:1157–1161
- Fullerton J, Cubin M, Bhomra A, Davidson S, Miller S, Turn M, Dolby C, Mott R, Wang C, Tiwari H, Allison D, Neale M, Fairburn C, Goodwin G, Flint J (2002) Linkage analysis of extremely discordant and concordant sibling pairs identifies QTL that influence variation in a human personality trait. *Am J Hum Genet* 71[Suppl]:263
- Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES (1999) Molecular classification

- of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537
- Gupta P, Duplessis S, white H, Karnosky DF, Martin F, Podila GK (2005a) Gene expression patterns of trembling aspen trees following long-term exposure to interacting elevated CO<sub>2</sub> and tropospheric O<sub>3</sub>. *New Phytol* (in press)
- Gupta PK, Rustgi S, Kulwal PL (2005b) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485
- Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R (2002) Plant metabolomics: the missing link in functional genomics strategies. *Plant Cell* 14:1437–1440
- Hetherington AM, Woodward FI (2003) The role of stomata in sensing and driving environmental change. *Nature* 424:901–908
- Ideker T, Galitski T, Hood L (2005) A new approach to decoding life: Systems biology. *Annu Rev Genomics Hum Genet* 2:343–372
- Kearsey MJ, Farquhar AGL (1998) QTL analysis in plants; where are we now? *Heredity* 80:137–142
- Kirkpatrick M, Jarne P (2000) The effects of a bottleneck on inbreeding depression and the genetic load. *Am Nat* 155:154–167
- Lister C, Dean C (1993) Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. *Plant J* 4:745–750
- Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods Inform Med* 40:346–358
- Martin GB (1998) Gene discovery for crop improvement. *Curr Opin Biotechnol* 9:220–226
- Matsuyama T, Tamaoki M, Nakajima N, Aono M, Kubo A, Moriya S, Ichihara T, Suzuki O, Saji H (2002) cDNA microarray assessment for ozone-stressed *Arabidopsis thaliana*. *Environ Pollut* 117:191–194
- Michael TP, McClung CR (2003) Enhancer trapping reveals widespread circadian clock transcriptional control in *Arabidopsis*. *Plant J* 31:279–292
- Miyazaki S, Fredricksen M, Hollis KC, Poroyko V, Shepley D, Galbraith DW, Long SP, Bohnert HJ (2004) Transcript expression profiles of *Arabidopsis thaliana* grown under controlled conditions and open-air elevated concentrations of CO<sub>2</sub> and O<sub>3</sub>. *Field Crop Res* 90:47–59
- Moore BD, Cheng SH, Rice J, Seemann JR (1998) Sucrose cycling, Rubisco expression, and prediction of photosynthetic acclimation to elevated atmospheric CO<sub>2</sub>. *Plant Cell Environ* 21:905–915
- Norby RJ, Wullschlegel SD, Gunderson CA, Johnson DW, Ceulemans R (1999) Tree responses to rising CO<sub>2</sub> in field experiments: implications for the future forest. *Plant, Cell Environ* 22:683–714
- Nordborg M (2000) Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* 154:923–929
- Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc Natl Acad Sci USA* 101:9885–9890
- Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20:3551–3567
- Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci* 5:439–445
- Quadroni M, James P (1999) Proteomics and automation. *Electrophoresis* 20:664–677
- Qin J, Fenyo D, Zhao YM, Hall WW, Chao DM, Wilson CJ, Young RA, Chait BT (1997) A strategy for rapid, high confidence protein identification. *Anal Chem* 69:3995–4001

- Rae AM, Graham LE, Street NR, Hughes J, Hanley ME, Tucker J, Taylor G (2005) QTL for growth and development in elevated carbon dioxide in two model plant genera: a novel approach for understanding adaptation to climate change? *Global Change Biol* (in press)
- Rafalski A, Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet* 20:103–111
- Ranasinghe S, Taylor G (1996) Mechanism for increased leaf growth in elevated CO<sub>2</sub>. *J Exp Bot* 47:349–358
- Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R (2003) The study of macromolecular complexes by quantitative proteomics. *Nat Genet* 33:349–355
- Schaffer R, Landgraf J, Perez-Amador M, Wisman E (2000) Monitoring genome-wide expression in plants. *Curr Opin Biotechnol* 11:162–167
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science* 270:467–470
- Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M, Sandberg G (2004) A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *Plant Cell* 16:2278–2292
- Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002) Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J* 31:279–292
- Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62:817–836
- Tao WA, Aebersold R (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr Opin Biotechnol* 14:110–118
- Taylor G, Tricker PJ, Zhang FZ, Alston VJ, Miglietta F, Kuzminsky E (2003) Spatial and temporal effects of free-air CO<sub>2</sub> enrichment (POPFACE) on leaf growth, cell expansion, and cell production in a closed canopy of poplar. *Plant Physiol* 131:177–185
- Taylor G, Street NR, Tricker PJ, Sjödin A, Graham L, Skogström O, Calfapietra C, Scarascia-Mugnozza, Janssen S (2005) The transcriptome of *Populus* in elevated CO<sub>2</sub>. *New Phytol* 167:143–154
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (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
- Vodkin LO, Khanna A, Shealy R, Clough SJ, Gonzalez DO, Philip R, Zabala G, Thibaud-Nissen F, Sidarous M, Stromvik MV, Shoop E, Schmidt C, Retzel E, Erpelding J, Shoemaker RC, Rodriguez-Huete AM, Polacco JC, Coryell V, Keim P, Gong G, Liu L, Pardinias J, Schweitzer P (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. *Genomics* 5:73
- Wagner C, Sefkow M, Kopka J (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry* 62:887–900
- Ward JK, Kelly JK (2004) Scaling up evolutionary responses to elevated CO<sub>2</sub>: lessons from *Arabidopsis*. *Ecol Lett* 7:427–440
- Wasaki J, Yonetani R, Shinano T, Kai M, Osaki M (2003) Expression of the OsPI1 gene, cloned from rice roots using cDNA microarray, rapidly responds to phosphorus status. *New Phytol* 158:239–248

- Watson BS, Lei ZT, Dixon RA, Sumner LW (2004) Proteomics of *Medicago sativa* cell walls. *Phytochemistry* 65:1709–1720
- Weir BS, Basten CJ (1990) Sampling strategies for distances between DNA-sequences. *Biometrics* 46:551–572
- Wu R, Ma CX, Casella G (2002) Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics* 160:779–792
- Wu TD (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J Pathol* 195:53–65