Global Change Biology (2009) **15**, 1201–1213, doi: 10.1111/j.1365-2486.2008.01818.x

Gene expression profiling: opening the black box of plant ecosystem responses to global change

ANDREW D. B. LEAKEY*, ELIZABETH A. AINSWORTH*†, STEPHANIE M. BERNARD‡, R. J. CODY MARKELZ*, DONALD R. ORT*†, SARAH A. PLACELLA§, ALISTAIR ROGERS¶‖, MELINDA D. SMITH**, ERIKA A. SUDDERTH††, DAVID J. WESTON‡‡, STAN D. WULLSCHLEGER‡‡ and SHENGHUA YUAN**

*Department of Plant Biology, Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA, †Photosynthesis Research Unit, USDA-ARS, Urbana, IL 61801, USA, ‡Ecology Department, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, §Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA, ¶Department of Environmental Sciences, Brookhaven National Laboratory, Upton, NY 11973, USA, ‖Department of Crop Sciences, University of Illinois at Urbana–Champaign, 1207 West Gregory Drive, Urbana, IL 61801, USA, **Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA, ††Department of Integrative Biology, University of California, Berkeley, CA 94720, USA, ‡‡Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422, USA*

Abstract

The use of genomic techniques to address ecological questions is emerging as the field of genomic ecology. Experimentation under environmentally realistic conditions to investigate the molecular response of plants to meaningful changes in growth conditions and ecological interactions is the defining feature of genomic ecology. Because the impact of global change factors on plant performance are mediated by direct effects at the molecular, biochemical, and physiological scales, gene expression analysis promises important advances in understanding factors that have previously been consigned to the 'black box' of unknown mechanism. Various tools and approaches are available for assessing gene expression in model and nonmodel species as part of global change biology studies. Each approach has its own unique advantages and constraints. A first generation of genomic ecology studies in managed ecosystems and mesocosms have provided a testbed for the approach and have begun to reveal how the experimental design and data analysis of gene expression studies can be tailored for use in an ecological context.

Keywords: elevated CO_2, genomic, microarray

Received 7 July 2008 and accepted 14 September 2008

Introduction

The use of genomic techniques to address ecological questions is emerging as the important new field of genomic ecology (Jackson *et al*., 2002; Ouborg & Vriezen, 2007; Wullschleger *et al*., 2007; Roelofs *et al*., 2008; Shiu & Borevitz, 2008; Ungerer *et al*., 2008). Tools are now available to assess: (1) variation in genome sequence, (2) patterns of gene expression, and (3) gene function (Ouborg & Vriezen, 2007). The use of many of these tools, including quantitative trait loci analysis, association mapping, and genome sequencing has been reviewed previously (Lee *et al*., 2004; Straalen & Roelofs, 2006; Ouborg & Vriezen, 2007). This review focuses on how experiments investigating plant responses to elements of global change are becoming a testing ground for the use of transcript profiling, as a result of strategically targeted funding from U.S. Department of Energy's Program for Ecosystem Research (http://per.ornl.gov/PERprojects-current.html). Support for genomic ecology is timely because the new techniques available, and specifically gene expression analysis by transcript profiling, are ideal for addressing many of the major knowledge gaps in plant responses to global change. It is well recognized that our ability to predict the impact of global change on both ecosystem function and food supply is constrained by our limited

Correspondence: Andrew Leakey, tel. +1 217 244 0302, fax +1 217 244 2057, e-mail: leakey@life.uiuc.edu

understanding of plant responses to interacting elements of global change (e.g., drought × elevated CO_2), intra- and interspecific variation in response, nonlinear responses, and trophic interactions (Poorter, 1993; Wullschleger *et al.*, 2002; Fuhrer, 2003; Leakey *et al.*, 2006a; Long *et al.*, 2006; Bradley & Pregitzer, 2007; Delucia *et al.*, 2008). Because the impact of global change factors on plant performance are mediated by direct effects at the molecular, biochemical, and physiological scales, investigation of these processes promises understanding that has previously been consigned to the 'black box' of unknown mechanism (Fig. 1). This can be done in the traditional hypothesis-testing framework or in surveys designed to identify novel and unexpected aspects of response. In either case, there has been a move towards broader and more integrative thinking as transcript profiles are combined with high-throughput metabolite screening, physiological assessment, and automatic environmental data collection (Fig. 1).

Incorporation of global transcript profiling and other 'omic' approaches into ecological studies constitutes a major shift in philosophy compared with investigation of a few physiological and ecological parameters, and necessitates collaboration among scientists with diverse skill sets. An additional key feature of genomic ecology is experimentation under ecologically relevant treatments and conditions, unlike many molecular biology studies that have used shock treatments for the task of elucidating gene function. The new genomic ecology approach requires the physiologist and ecologist to learn new techniques and optimize the tools for use within the ecosystem context. This paper is the outcome of a workshop held at the University of Illinois in November 2007 to review the opportunities available for addressing important questions in global change biology using transcript profiling and associated technologies. We discuss the different approaches of studying model vs. nonmodel species, the opportunities and challenges in profiling ecologically relevant gene expression, and the value and interpretation of 'omic' data in an ecological context.

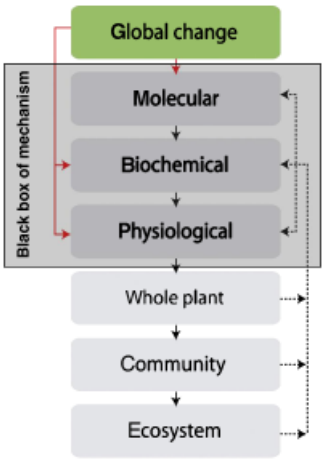

Fig. 1 Schematic describing the integration of plant, community, and ecosystem responses to an element of global change. Elements of global change directly impact molecular, biochemical, and physiological processes (red arrows), which combine to determine whole plant performance. Genotypic variation in whole plant responses drives ecological interactions that underlie community and ecosystem responses to global change. Feedbacks from larger scales of organization (dashed arrows) impact individual plant performance via effects on resource availability and disturbance that modify the direct effects on global change on plant function. Transcript profiling and high-throughput biochemical and physiological screening provide an opportunity to better understand the 'black box' of mechanisms driving plant responses to various elements of global change under field conditions.

Investigating model and nonmodel species

Busch & Lohmann (2007) classified the different methods for gene expression profiling in three categories: (1) PCR-based methods, such as quantitative real-time reverse-transcription PCR (qRT-PCR); (2) sequencing-based methods, such as cDNA-AFLP (amplified fragment length polymorphism), serial analyses of gene expression (SAGE), and massive parallel signature sequencing (MPSS); and (3) hybridization-based methods, such as microarrays. For this discussion, model species are defined as those for which a sufficiently large fraction of the genome has been sequenced to allow relatively easy transcript profiling of most or all genes by qRT-PCR or microarray analysis. Although real-time PCR can be high-throughput (Czechowski *et al.*, 2004), microarray analysis is currently the most common method of choice for transcript profiling. Microarrays are glass, plastic, or silicon chips with thousands of DNA oligonucleotides arrayed across their surface. Each oligonucleotide spot, or probe, corresponds to a specific target mRNA from a specific gene. The pool of RNA transcripts from sample tissue is extracted and labeled with a fluorescent tag before being washed over the microarray. Transcripts bind to their corresponding probes and the abundance of all transcripts is quantified by assessing the intensity of fluorescence associated with each probe. The result is information on the abundance of transcripts encoding a large fraction of the protein structures and enzymes in the sample tissue. A major assumption in interpretation of microarray data is that transcript abundance is related to protein synthesis and activity. The method does not directly assess the rate of gene expression or transcript degradation, but instead the pool size of

transcripts that is the result of the two processes. In addition, a number of posttranscriptional and posttranslation processes can disrupt the link between transcript abundance and enzyme activity. These assumptions influence the inferences that can be drawn from such datasets, but have not prevented the widespread use of this powerful technique.

To date, microarrays have been produced for at least 38 plant species (Table S1). Affymetrix is the largest commercial supplier of microarrays, and alone produces microarrays for Arabidopsis, barley, cotton, citrus species, grape, maize, *Medicago* spp., poplar, rice, soybean, sugarcane, tomato, and wheat (www.affymetrix.com). Other companies and research institutions manufacture microarrays for additional species, but these are also biased towards economically, rather than ecologically, significant species. As of February 2008, the National Center for Biological Information listed 37 land-plant species for which whole-genome sequencing was complete or in progress (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). Increasing numbers of ecologically and evolutionarily important species such as *Arabidopsis lyrata*, *Capsella rubella*, *Brachypodium distachyon*, *Mimulus lewisii*, and *Selaginella moellendorffii* are being sequenced. More species will rapidly become available for genomic investigation as techniques such as pyrosequencing allow smaller research groups to generate large amounts of sequence information and develop tools specifically for their own species of interest (Hudson, 2007). Alternatively, some researchers are using the technique of heterologous hybridization to profile transcripts of nonmodel species with microarrays designed for a closely related model species (e.g., Gong *et al*., 2005; Travers *et al*., 2007). These various tools and approaches for studying gene expression mean that one can choose between studying model and nonmodel species to address genomic ecology questions in global change biology; however, each approach has constraints that are important to consider.

Limitations to molecular and functional inference in model and nonmodel species

Generally, in model species that have been fully sequenced and for which microarrays have been specifically designed (e.g., Arabidopsis, *Populus* sp., and rice), data describing the abundance of nearly all transcripts can be attributed to the relevant genes with a high degree of confidence. Nonspecific binding of products from two or more genes to a single probe on the microarray, or cross-hybridization, can cause problems if genes share very high sequence similarity, but is relatively rare (Shiu & Borevitz, 2008). Even when a full genome sequence becomes available, it is not immediately possible to (1) identify all the genes capable of being expressed to produce proteins, and (2) assign all the RNA transcripts being profiled to specific genes. However, bioinformatic techniques to identify genes are becoming increasingly efficient, especially when sequences from multiple species are analyzed in parallel (Lin *et al*., 2007).

High-quality transcript profile data are also available for species for which microarrays have been produced from expressed sequence tag libraries but the full genome sequence is not available, such as maize and soybean (Wang *et al*., 2003; Vodkin *et al*., 2004). However, there is less certainty that (1) each probe sequence on the microarray is unique to a single gene, or (2) every functional gene is detected by the microarray. For example, the soybean genechip from Affymetrix probes expression of $\sim$ 38 000 unique genes, while the recent Joint Genome Institute (http://www.jgi.doe.gov/) release of the soybean genome suggests there are 58 556 loci containing protein-coding transcripts (http://www.phytozome.net/soybean). Further assembly and analysis is needed before it is known how much the disparity in these numbers is explained by partial polyploidy (Schlueter *et al*., 2007).

Using heterologous hybridization to study transcript profiles of nonmodel species causes greater uncertainty about cross-hybridization or missing genes. A preliminary analysis from hybridizing the genomic DNA of the study species to the microarray can be useful in identifying which probes have no corresponding gene and therefore can be subsequently ignored. Although this reduces the number of genes whose expression can be profiled, heterologous hybridization has been used to identify genes important to drought stress, cold stress, and heavy metal tolerance (Gong *et al*., 2005; Hammond *et al*., 2006; Sharma *et al*., 2007). In studies on a single nonmodel species, errors associated with heterologous hybridization should be common to all treatments, which limit some of the problems in interpretation. By comparison, if the transcript profiles of multiple species are assessed with a common microarray platform, then sequence divergence among species could impact the efficiency of hybridization and falsely suggest differential transcript abundance. Comparing among species the results of hybridizing genomic DNA with microarrays can help quantify the extent of this problem and again eliminate probes likely to cause problems (Shiu & Borevitz, 2008).

Functional interpretation of microarray data is dependent on correct annotation of gene function. As sequence data from plants accumulates, finding means to efficiently and effectively analyze the sequences and assign annotation remains a major challenge (Dong *et al*., 2005). Arabidopsis has been the primary subject

of studies determining gene function in plants and, therefore, more (though far from all) genes have been annotated in this species, and annotations are generally accepted with the greatest degree of confidence. Currently, ~60% of the 28 152 protein coding genes in Arabidopsis have been annotated to a Gene Ontology (GO) molecular function, with 50% annotated to a GO biological process (http://www.geneontology.org). The majority of annotations are based on a computational analysis of the gene sequence. Therefore, even in the most well studied plant species much work remains to be done to experimentally determine gene function. In other species the function of some genes may have been directly determined, but the annotation of the great majority of genes is inferred from sequence similarity to genes in Arabidopsis. An automated BLAST search (Altschul *et al*., 1997) against a protein database accomplishes this task. Top BLAST matches are typically assigned an expectation value along with a putative function and GO terms associated with similar protein sequences. The more evolutionarily distant from Arabidopsis the subject species is, the greater the likelihood the gene sequence will have diverged, which increases uncertainty in the annotation. Nonetheless, many genes are highly conserved and can be annotated with confidence in a large number of distantly related species (Frickey *et al*., 2008). The BLAST procedure has the inherent flaw of propagating annotation errors from one species to another (Gilks *et al*., 2002), but remains the most practical choice for sequence annotation. As more sequence data from various species becomes available, interspecific sequence analyses are also proving valuable for improvement of annotations, automation of annotation, and identification of novel coding regions (Windsor & Mitchell-Olds, 2006).

Limitations to ecological inference in model and nonmodel species

The vast majority of species for which substantial sequence information and transcript profiling tools are available have been selected because of their economic importance (Table S1). This has created enormous potential for investigating the mechanisms underlying the impacts of global change on crop yield and agroecosystem function. Transcript profiling can reveal changes in gene expression that drive physiological and ecological responses, and in doing so improve understanding of mechanism at all scales (Fig. 1). Managed ecosystems and mesocosms incorporating model species are an excellent test bed for genomic ecology because their low genetic and environmental heterogeneity increases the statistical power of field experiments and facilitates detection of subtle treatment differences (Ainsworth *et al*., 2006; Casteel *et al*., 2008; Leakey *et al*., 2009; Zavala *et al*., 2008). In addition, the current group of model species incorporates considerable diversity including angiosperms and gymnosperms, herbaceous plants and trees, C_3 and C_4 species, legumes and nonlegumes, and tropical and temperate species. This allows further fundamental biological questions to be asked regarding variation in response to global change of major functional and phylogenetic groups. However, these species are not always ideal subjects for addressing a number of important ecological and evolutionary questions in global change biology. The majority are crops bred for rapid growth and reproductive output on annual growth cycles. This means that the mechanisms underlying their responses to resource availability, disturbance, and competition may differ from those of other species adapted to diverse habitats in natural communities. Custom-made transcript profiling tools are not currently available for multiple plant species from even one natural community. This limits characterization of species-specific gene expression patterns and its contribution to driving the species interactions that control community and ecosystem responses (Fig. 1). One solution would be to accept the limitations and assumptions of heterologous hybridization in order to assess diversity of gene expression responses across a larger number of species (Travers *et al*., 2007). Alternatively, custom genomic tools could be developed for the species comprising a 'model' ecosystem or species possessing ecological traits of particular interest. Such an approach is becoming increasingly feasible with continued advances in the development of high-throughput sequencing technologies (Hudson, 2007).

Expectations, design, and analysis of ecologically relevant transcript profiling experiments

Expectations of gene expression responses to global change scenarios

Experimentation under environmentally realistic conditions to investigate the molecular response of plants to meaningful changes in growth conditions and ecological interactions is the defining feature of the genomic ecology approach. A typical laboratory-based microarray study aiming to elucidate the functions of genes will subject plants to an acute treatment that precipitates many-fold changes in the transcript abundance of thousands of genes. In contrast, results from a typical genomic ecology experiment will reveal markedly smaller magnitude changes in the abundance of transcripts from a smaller number of genes. This probably has two main causes: (1) the imposed treatments are less severe, and (2) the focus is often on plants that have

acclimated to the treatments, in many cases spending their entire lifecycle exposed to the given treatment. In field studies there are the additional distinguishing factors of greater noise in gene expression resulting from the variable growth conditions and the greater resilience of field grown plants than laboratory grown plants to perturbation.

Treatments in global change biology experiments are typically mild (e.g., a 40% difference in [CO_2]), because they aim to test the impact of changes between average field conditions today and those expected for later this century. By comparison, many molecular studies aiming to identify stress responsive genes have ensured significant treatment effects would be observed by imposing extreme conditions, such as supply of strong (200–500 mM) salt solutions (Bohnert *et al*., 2001), exposure to high (300 ppb) ozone concentrations (Tosti *et al*., 2006), and withholding water from plants in small pots of rapidly drying growth media (Talame *et al*., 2007). Important data have been generated from such experiments, but the results may not always inform us about the mechanisms controlling plant performance in the field. For example, a cell-death response leading to lesions on leaves has been identified as an important component of response in plants exposed to >300 ppb [O_3] (500% above background), but growth at <100 ppb [O_3] (60% above background) impairs productivity without causing visible damage to the plant (reviewed by Long & Naidu, 2002).

When plants experience a change in growing conditions (e.g., transfer from moderate to high temperature), they display a progression of responses. First, the altered condition is sensed, activating a signal transduction pathway, which typically drives metabolic adjustments and concludes with adoption of a new acclimated state. Well-studied examples are the time courses of cellular response to ozone exposure and attack by pathogens (Lamb & Dixon, 1997; Kangasjarvi *et al*., 2005). The changes in gene expression immediately and shortly after the change in condition are substantial in number and magnitude. Most studies aiming to understand the molecular basis of plant responses to abiotic and biotic stimuli have focused on characterizing the responses to short-term changes in conditions. This is very important for understanding the sensing and signaling processes that control the response. Also, in combination with strong treatments, the brief shock generates an easy-to-detect response. However, these short-term changes in gene expression do not reveal all the important controls of plant performance upon acclimation to the growth conditions. For example, when assessed using a high-density maize oligonucleotide array, far fewer ($<2\%$ vs. 27%) genes showed differential expression in maize ear tissue under a gradually developing stress than under a sudden stress (Campos *et al*., 2004).

Many genomic ecology studies are building on information from experiments employing acute treatments to determine gene function by characterizing the more subtle changes in gene expression that differentiate fully acclimated plant performance in different experimental treatments. This forces microarray studies to be designed and analyzed differently. For example, it is very logical to focus primarily on changes in transcript abundance of >1.5-fold if the objective is to identify components of a signal transduction pathway a specific number of hours following a stimulus (e.g., Tosti *et al*., 2006). Equally, fivefold changes in transcript abundance for metabolic genes are unlikely to be observed in plants that are fully acclimated to growth in two mildly different treatments. For example, in Free-Air Concentration Enrichment (FACE) experiments where, in many cases plants have been grown for their entire life cycles at current and elevated [CO_2], the largest fold changes in transcript abundance due to the CO_2 treatment are typically ca. twofold (Gupta *et al*., 2005; Taylor *et al*., 2005; Ainsworth *et al*., 2006; Leakey *et al*., 2009). Identifying these moderate changes can give considerable insight into alterations in metabolic pathways and allocation to biosynthetic pathways that occur over time in response to elements of global change. But the genomic ecologist is faced with the problem of balancing the cost of transcript profiling with the need for adequate replication to gain sufficient statistical power to detect small fold changes in transcript abundance.

By comparison with controlled environment facilities, field conditions can provide growing conditions for plants that are simultaneously more variable, more resource rich, and more stressful. For example, many habitats provide high light and unlimited rooting volume but also periods of water deficit and disease. This appears to reduce the sensitivity with which gene expression responds to stress treatments. For example, application of benzo(1,2,3)-thiadiazole-7-carbothioic acid *S*-methylester (BTH) to induce systemic resistance against pathogens in wheat caused substantial upregulation of defense-related genes in a greenhouse trial. However, when the experiment was repeated under field conditions, defense-related gene expression was constitutively high and did not increase further with the BTH treatment (Pasquer *et al*., 2005).

Experiments that investigate the response of plants to treatments simulating global change over long time periods are informative because they can generate understanding of (1) impacts over the entire life histories of the subject species, (2) slow ecological responses such as competition and succession, and (3) complex feedbacks

from ecological and ecosystem scale to whole plant performance (Fig. 1). Fewer space restrictions allow long-term experiments to be done in the field more successfully than under controlled environment conditions. However, plants in the field, and especially those in long-term studies, experience variable growth conditions on scales from minutes, hours, and days to months and seasons. Many of the parameters of ecological interest, for example, biomass, yield, and fecundity integrate these growth conditions over long periods of time. In contrast, transcript profiles in plants are known to respond rapidly and extensively to temperature (Seki *et al*., 2002) and light (Bertrand *et al*., 2005), show circadian rhythms (Michael & McClung, 2003; Blasing *et al*., 2005) and vary with development (Taylor *et al*., 2005; Ainsworth *et al*., 2006). Because a single sampling point only represents a snapshot view, it is important to distinguish responses of the transcriptome that are due to the experimental manipulation vs. time or weather-dependent changes (Miyazaki *et al*., 2004). This discrimination can be achieved to a significant extent by sampling at the same time each day, sampling on multiple occasions over the duration of an experiment and interpreting treatment effects on gene expression in the context of environmental data. In addition, efforts to sample homogenous tissue that is at the same developmental stage and growing under the same environmental conditions minimize unwanted variability that could prevent detection of treatment effects. In some cases the impact of natural variation in growth conditions on gene expression can provide novel understanding of the mechanisms underlying plant–environment interactions. For example, transcript profiling of pine trees grown in multiple field sites in Europe indicated that cold tolerance develops in response to combined photoperiodic and temperature cues (Joosen *et al*., 2006).

Design of experiments assessing gene expression responses to global change scenarios

Nettleton (2006) reviewed how the basic principles of experimental design apply to transcript profiling experiments, with emphasis on random assignment of experimental units to treatments, use of the maximum affordable replication and applying blocking. These issues are familiar to ecologists and physiologists, and have been extensively reviewed (e.g., Scheiner & Gurevitch, 2001). The more specific importance of understanding the distinction between, and value of, technical and biological replication in transcript profiling experiments has been highlighted by Allison *et al*. (2006) and Nettleton (2006). Technical replication provides multiple measures of a single sample from a single experimental unit. Biological replication involves measurements of multiple experimental units each of which is independently exposed to control or treatment conditions. Without biological replication it is not possible to statistically attribute observed changes in transcript abundance to the effects of a treatment. Most experiments are limited by the funds available for transcript profiling. The power to detect treatment effects will be maximized if the transcripts from each experimental unit at a given time are profiled with only one microarray (Nettleton, 2006). However, if the number of biological replicates is limited (e.g., at a Free-Air CO_2 Enrichment experiment) and there is significant measurement error, averaging across technical replicates can reduce variability and provide some gain in statistical power (Nettleton, 2006).

In ecological experiments and especially those in the field, variation in gene expression responses to experimental treatments over time are of great interest with respect to circadian/diel rhythms, interactions with climate, acclimation, and development. With a limited supply of microarrays, this creates both challenges and opportunities. If the primary aim of the experiment is to characterize treatment effects on gene expression at a single time point (e.g., a single development event such as flowering), then adding biological replicates will provide the most statistical power. If the primary aim of the experiment is to characterize the average treatment effects on gene expression (e.g., over a growing season), then it may be desirable to compromise technical or biological replication in order to allow additional sampling points over time. Of course, such trade-offs need to be determined on a case-by-case basis. Even for studies on the same species at a single field site, some experiments may necessitate technical replication (e.g., Ainsworth *et al*., 2006), while others benefit most from multiple measurements in time (e.g., Casteel *et al*., 2008).

Subsampling is often used to overcome the variation among individuals within a replicate plot in field experiments. For instance, averaging the rates of photosynthesis of four different sun leaves within individual plots of maize exposed to either ambient or elevated [CO_2] reduced variation among replicate plots and ensured there was sufficient statistical power to characterize a subtle, episodic treatment effect (Leakey *et al*., 2004, 2006a). In genomic ecology studies, one solution to the need for sampling variation within replicate plots without depleting microarray resources needed for sampling multiple biological replicates or time points is to pool mRNA from multiple samples collected within a single plot (Allison *et al*., 2006). Hybridizing this mixed mRNA sample to a single microarray will reduce between plot variance when biological variability is

high relative to measurement error (Kendziorski *et al*., 2005). This approach has been used successfully in transcript profiling studies of poplar and soybean responses to elevated [CO_2] in the field (e.g., Gupta *et al*., 2005; Taylor *et al*., 2005; Ainsworth *et al*., 2006).

Analysis of gene expression responses to global change treatments

One of the greatest challenges of transcript profiling is the data analysis. This is partly due to the large size of the datasets compared with most physiological or ecological experiments. Selecting from the large number of rapidly developing analysis tools and techniques that are available is also challenging. Although it is impossible to comprehensively discuss the advantages and disadvantages of all the available options here, it is worth briefly reviewing the major steps in the analysis process and highlighting a number of specialist reviews on the subject (e.g., Allison *et al*., 2006; Nettleton, 2006).

The first analysis step involves processing the images of the fluorescent spots on each microarray. Many approaches have been developed, and the service facilities that perform the hybridization and scanning of microarrays for most investigators can assist in making the appropriate choices. Before proceeding with data analysis, it is important to perform quality control steps and remove or replace data from defective slides or images. One simple method for eliminating poor quality data is to discount data from microarrays that do not meet threshold values of the Pearson correlation coefficient (e.g., 0.9) or kappa statistic (e.g., 0.75) when pairwise comparisons are made between microarrays from a given treatment and time point (Fleiss, 1981; McIntyre *et al*., 2006).

A number of microarrays include probes for more than one species. For example, the Affymetrix soybean genechip includes probes for genes from soybean, a nematode species and the phytopthera pathogen. If only transcripts from soybean are to be profiled, the data for probes specific to the other species should be disregarded. The Affymetrix genechip platform provides a statistic estimating whether each individual transcript is considered to have been present or absent from the sample (Affymetrix, 2002). This allows the investigator to discount data from probes for which a transcript was not considered present in a sufficient number of samples for meaningful replication to be achieved. This reduces the number of tests to be performed and prevents misinterpretation of results from probes for which there is not sufficient statistical power for meaningful testing.

Some methods can combine image analysis algorithms with the next analysis step, which is normalization. Normalization is the process that makes adjustments to minimize the influence of technical variability across different microarrays and experiments. The simplest approach involves normalizing the fluorescence intensity for individual genes by the median fluorescence intensity on an individual microarray basis (e.g., Ainsworth *et al*., 2006; Fung *et al*., 2008). This approach has the philosophical advantage of maintaining the independence of data from individual replicates, and the practical advantage of requiring a single normalization to be performed on a given microarray, even if the data are to be analyzed as part of more than one experiment. Alternatively, more complex procedures have also been developed, some of which incorporate information from all the chips in an experiment as part of the normalization process. Normalization is an area of ongoing research in which there is an unresolved debate about which method performs the best, and even how good performance should be defined (Irizarry *et al*., 2003; Bolstad *et al*., 2003; Choe *et al*., 2005; Allison *et al*., 2006). After normalization, log transformation of the data is performed in nearly all cases to ensure that the data are normally distributed.

The majority of published microarray studies use mixed-effects linear models to identify treatment effects on transcript abundance, with an independent analysis being performed for each probe in the dataset (Nettleton, 2006). This has the advantage of allowing the physiologist or ecologist to use familiar statistical tests and software packages. An additional reason for the approach is that different genes display different levels of variation in expression, creating heterogeneity that a single 'global' model has difficulty representing. However, it has been suggested that analyzing each gene independently is inefficient (Allison *et al*., 2006). Simulation studies have indicated that an intermediate approach, called variance shrinkage, which combines data from specific genes and all genes may perform better than gene-by-gene testing (Cui *et al*., 2005), although optimization of the technique is still required (Allison *et al*., 2006).

Because the analysis of most microarray experiments necessitates tens of thousands of statistical tests on individual probes, there is a greater likelihood of making type I errors (falsely identifying the abundance of transcripts as responsive to the treatment when in fact they are not) than in most physiological or ecological experiments. Consequently, techniques have been developed that quantify the false discovery rate (FDR) and allow it to be controlled (Benjamini & Hochberg, 1995; Storey & Tibshirani, 2003). Most commonly, this is done by adjusting the probability threshold at which treatment effects on transcript abundance are considered to be statistically significant, taking into account the

number of tests performed and the initial *P*-value returned for each transcript by the mixed-effects linear model. Importantly, while applying increasingly strict FDR corrections reduces the number of transcripts falsely identified to respond significantly to the treatment, it also increases the number of transcripts falsely identified *not* to respond significantly to the treatment (type II errors; Nettleton, 2006). In other words, there is a trade-off between identifying fewer genes than actually responded to the treatment, but with a high degree of confidence (strict FDR) vs. more genes that actually responded to the treatment, plus some that did not (relaxed FDR). During the experimental design and analysis processes, each researcher must select the FDR correction level that allows the most meaningful interpretation of the data.

In many global change biology experiments, where treatment effects can be small, applying strict FDR can result in few transcripts being identified as responding to the treatment. If more relaxed FDR are applied, other techniques are necessary to increase the confidence with which 'responsive' transcripts are identified. For instance, visualization of transcript data in the context of known metabolic pathways and signal transduction cascades can indicate when many transcripts associated with a common function or response display consistent responses to an experimental treatment (e.g. Leakey *et al*., 2009). If transcripts are identified as a result of random variation and not a true treatment effect, then positive and negative responses should be equal in number. However, for example, if the abundance of $\geq 50\%$ of all transcripts-encoding enzymes involved in the synthesis of flavonoids are greater when soybean grows at elevated [O_3], and no transcripts show the opposite result (Casteel *et al*., 2008), there is a good probability that the result is real rather than the result of random chance.

Difficulties associated with performing many tests can also be dealt with by putting transcripts into functional groups and performing a Fisher's exact test or chi-square test on each group. These two tests allow identification of groups within which a greater fraction of transcripts respond significantly to the treatment than on average across all transcripts, that is, functional groups of transcripts which disproportionately contribute to the overall transcriptional response. For example, the transcriptional response of soybean to growth at elevated [CO_2] was assessed by assigning each of the profiled transcripts into one of 32 functional groups (Leakey *et al*., 2009). A Fisher's exact test determined that the fraction of CO_2-responsive transcripts in functional groups related to respiration was significantly greater than the fraction of CO_2-responsive transcripts across all other functional groups.

The standard procedure of repeating an experiment can also be used to increase confidence in identification of 'responsive' transcripts. Transcripts whose abundance changes as a result of real treatment effects are more likely to display consistent changes in abundance of a similar magnitude and in the same direction. In contrast, false positives that have low *P*-values from the initial analysis of variance (ANOVA) as a result of random variation are equally likely to respond positively or negatively to the treatment in any given experiment. Varying the FDR threshold has a substantial impact on identification of transcripts which respond consistently in soybean grown at ambient and elevated [CO_2] over two consecutive growing seasons (Fig. 2). At an FDR of 0.2, 76 transcripts responded consistently in the 2 years, and no transcript displayed opposite responses in the 2 years. By contrast, applying an FDR of 0.5 to the same data identified 615 transcripts that responded consistently and 12 transcripts displaying opposite responses in the 2 years. The researcher has to choose between identifying treatment effects on 76 transcripts with a higher degree of confidence from a more conservative FDR correction or 615 transcripts from a less conservative FDR correction, plus the knowledge the transcript responded to the treatment in the same direction, and to a similar magnitude, in 2 consecutive years. Given the

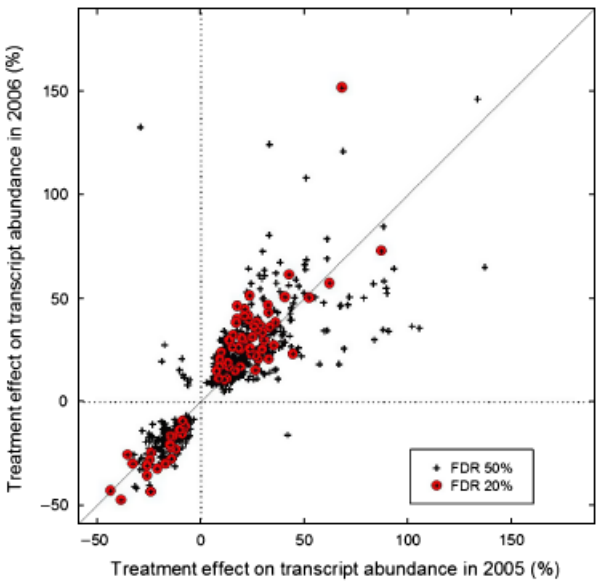

Fig. 2 Comparison of changes in transcript abundance in soybean leaves as a result of growth at ambient [CO_2] vs. elevated [CO_2] at SoyFACE during the 2005 and 2006 growing seasons. At a false discovery rate (FDR) of 0.2, 76 transcripts responded consistently in the 2 years and no transcript displayed opposite responses in the 2 years. By contrast, applying an FDR of 0.5 to the same data identified 615 transcripts that responded consistently and 12 transcripts displaying opposite responses in the 2 years. Data adapted from Leakey *et al*. (2009).

need to demonstrate that changes in transcript abundance have an impact on biochemical or physiological processes, there will be subsequent opportunity to eliminate false positives that have passed this initial analysis.

It is a common practice to validate the quantification of transcript abundance by microarrays using qRT-PCR on a subset of genes from the original experimental samples (Rajeevan *et al*., 2001). However, this practice has recently been suggested to be of little benefit because, while qRT-PCR probably provides a more accurate measure of transcript abundance, there is no reason to expect the new data will eliminate the types of errors that cause identification of false positives (Allison *et al*., 2006). This may not yet be a consensus view, but seems to be consistent with most physiological and ecological practices. For example, measurements of stomatal conductance using IRGA-based gas exchange systems are not typically validated with measurements using a porometer (e.g., Jones, 1999; Leakey *et al*., 2006b). On the other hand, qRT-PCR is more sensitive to changes in transcript abundance than microarrays and it is incredibly valuable and cost effective if transcripts identified in a microarray study are profiled by quantitative-PCR in samples from additional biological replicates, other tissues, or other time points. Such follow-up studies are vital to extend investigation from broad profiling analyses to detailed understanding of specific gene responses.

Linkages from gene expression to physiology and ecology

The ability to measure gene-specific and genome-wide patterns of transcript abundance provides a new opportunity to improve our understanding of how organisms and ecosystems respond to environmental change. Different elements of global change can elicit distinct changes in gene expression (e.g., drought vs. heat; Roelofs *et al*., 2008); therefore, the contribution of two simultaneous treatments in impacting physiological performance could start to be dissected by the molecular phenotypes. Because transcript profiling with microarrays potentially provides information on a large proportion of metabolic and signaling components, it is an ideal technique to broadly survey intra- and interspecific variation in response to a given treatment (e.g., Gong *et al*., 2005). Identifying different response pathways or magnitudes of response within a pathway at the molecular level identifies a smaller group of candidate mechanisms that can then be more easily examined at the physiological and ecological scale. For example, Leakey *et al*. (2009) used microarrays to characterize a transcriptionally driven acclimation of soybean to growth at elevated [CO_2], which led to stimulated foliar respiration. The transcript profiling also allowed a survey of biosynthetic metabolism to identify pathways that were transcriptionally upregulated coincident with the enhanced supply of energy and carbon skeletons from respiration. The ecological significance of these changes can now be evaluated in more detailed analyses. By comparison, previous methods would probably have involved laborious and less systematic investigation of individual biosynthetic pathways in different species by different research groups.

Interpretation of transcript abundance depends on assumptions about the relationship between the levels of transcripts and the functional activity of the proteins they encode. This is difficult to predict because post-transcriptional and posttranslational regulation can significantly alter the response predicted from transcript data alone (Scheible *et al*., 1997; Kaiser & Huber, 2001; Hendriks *et al*., 2003). In addition, the impact of changes in transcript abundance on a biological response depends on the turnover rate of the encoded proteins, their contribution to the control of metabolic pathways and the levels of metabolites associated with those pathways, which in turn can regulate the expression of the given gene. Genome-wide transcript profiling and analyses of enzyme activities have shown that transcript levels undergo marked and rapid changes during the diurnal cycles whereas changes in enzyme activities are often smaller and delayed, and appear to integrate changes in transcript levels over several diurnal cycles (Gibon *et al*., 2006; Morcuende *et al*., 2007; Stitt *et al*., 2007). Because transcripts, enzymes, and metabolites integrate information over different time scales, measuring their response provides a wider physiological snapshot than transcript abundance alone. Fortunately, unlike transcriptomics (and proteomics) which relies to a great extent on genomic information, metabolomics is widely applicable with only minimal time required to reoptimize protocols for a new species (Schauer & Fernie, 2006). High-throughput analysis of activity from >20 enzymes is now a reality (Gibon *et al*., 2004) and early indications suggest that these methods can also be transferred relatively easily among species (Rogers & Gibon, 2009). Although still a nascent field of investigation, techniques to model metabolic networks (Sweetlove & Fernie, 2005) and a diversity of bioinformatics tools are becoming available to aid in identifying genes that underlie important biological functions.

For transcriptomic and metabolomic data, visualization of the results in a biologically meaningful way is another challenge to functional interpretation. Thimm *et al*. (2004) introduced MAPMAN, a user-driven visualization tool for displaying transcript, metabolite, and enzyme activity datasets on plant-specific biological pathways. MAPMAN

is a flexible program that classifies genes into specific functional bins (e.g., photosynthesis, glycolysis, secondary metabolism), originally developed for Arabidopsis. It has since been extended to Solanaceous species (Urbanczyk-Wochniak *et al*., 2006) and legumes (Goffard & Weiller, 2006; Leakey *et al*., 2009) based on BLAST hits to the Arabidopsis proteome and the nonredundant protein database at NCBI. MAPMAN is but one example of a biologically relevant visualization tool. Such resources to interpret gene expression results are becoming ever more sophisticated and available for an increasing number of species. Many of the genes involved in photosynthesis, respiration, and nutrient acquisition can be identified using such software and results subsequently related to the response of plants to altered environmental conditions. Microarrays and bioinformatics, therefore, make a compelling combination to characterize mechanisms responsible for how plants respond to experimental manipulations of temperature, water, ozone, and CO_2 concentration (Watkinson *et al*., 2003; Ainsworth *et al*., 2006; Li *et al*., 2006; Weston *et al*., 2008).

Modeling gene expression data in the context of existing biochemical frameworks is useful, but requires that we understand *a priori* relationships between variables used to connect genes to physiology and beyond to ecosystem-scale processes. One challenge with this modeling approach is in reducing the dimensionality of the gene expression data before linking with the rest of the model. There are a number of approaches to accomplish this, such as the use of gene function ontologies to define a subset of genes whose expression values would subsequently be included in the model. Gene ontologies are insightful, but the functions of many genes are still not fully understood and extrapolation of model gene function to nonmodel genes is potentially problematic. Therefore, unsupervised approaches for delineating gene expression into functional clusters are promising. Weighted gene coexpression network analysis is encouraging in this regard because it is an unsupervised approach for clustering genes that share highly correlated expression patterns across treatment (Zhang & Horvath, 2005). Furthermore, the input data for this network approach are from normalized raw intensity values and thereby avoid multiple testing errors commonly associated with most expression array analytical techniques. Using this technique, Weston *et al*. (2008) were able to cluster Arabidopsis genes into functionally relevant stress responsive clusters (modules) that were then correlated to phenotypic characteristics. A similar statistical approach could be used to investigate module gene correlations with metabolites and enzyme activities of interest to strengthen understanding of the metabolic pathways governing phenotype.

Linking the large-scale datasets of genomic ecology to other predictors of plant responses to global change, including soil properties, biotic interactions, and climate conditions, presents several challenges. First, the spatial and temporal resolution of data collected across different levels of biological organization (i.e., molecular, organismal, community, and ecosystem) can vary significantly. In addition, accounting for the hierarchical structure of data can improve predictive accuracy when using multiple variables to explain observed plant responses. Statistical methods based on *probabilistic graphical models* provide a natural framework for modeling responses to environmental treatments. In this approach, the probabilistic relationships defining a complex system are specified via a sequence of nodes that represent random variables, and edges that encode direct physical or statistical dependencies (Jordan, 2004). The ability of graphical models to include *latent* or *hidden* variables to explicitly model unobserved relationships is particularly useful in biological research. A variety of computational methods developed by the statistics and machine-learning communities have been used to effectively analyze biological data with complex spatial and temporal structure. Directed graphical models, or Bayesian networks, are commonly used in systems biology to learn the structure of complex genetic networks (Blanchard, 2004; Friedman, 2004). Related multivariate modeling approaches such as structural equation modeling (SEM) have been used to identify the environmental and biotic predictors that influence plant response to various global change factors (Grace, 2006; Clark *et al*., 2007). Bayesian networks and SEM are only two examples of tools being used to analyze complex ecological responses to global change. These approaches provide powerful statistical tools that can be used to model plant responses to global change across levels of biological organization.

Conclusion

In summary, the technology to assess gene expression through transcript profiling is now available for model and nonmodel species. Managed ecosystems and mesocosms are proving to be good testbeds for the genomic ecology approach. Major advances in understanding natural communities are also promised by the increasing number of species for which transcript profiling tools are available and the accelerating advances in sequencing technology. This represents a significant new opportunity to assess the mechanisms underlying the responses of plants to elements of global change. The studies that have been performed to date have revealed some important distinctions between transcript profiling in ecological studies vs. molecular stu-

dies of gene function. This experience has allowed us to identify the strengths and weaknesses of various experimental design and analysis options available to the genomic ecologist. Possibly the biggest change resulting from the use of genomic tools is a new, integrative approach to investigating the abiotic and biotic interactions of plants. Using genomic ecology to understand the mechanisms currently consigned to the 'black box' of plant function will significantly advance analysis of future global change, its impacts on ecosystems and how we should respond to it.

Acknowledgements

We acknowledge support from the U.S. Department of Energy (DOE), Office of Science, Biological and Environmental Research program as part of its Program for Ecosystem Research. A. D. B. L., E. A. A., and D. R. O. were supported by grant no. DE-FG02-04ER63849. S. A. P., S. M. B., and E. A. S. were supported by contract no. DE-AC03-76SF00098 to Lawrence Berkeley National Laboratory. A. R. was supported by contract no. DE-AC02-98CH10886 to Brookhaven National Laboratory. D. J. W. and S. D. W. were supported by contract DE-AC05-00OR22725 to UT-Battelle, LLC, which manages Oak Ridge National Laboratory for the DOE.

References

Affymetrix (2002) *Statistical algorithms description document*. Technical Report. http://www.affymetrix.com/support/technical/whitepapers/sadd_whitepaper.pdf.

Ainsworth EA, Rogers A, Vodkin LO, Walter A, Schurr U (2006) The effects of elevated CO_2 concentration on soybean gene expression. An analysis of growing and mature leaves. *Plant Physiology*, **142**, 135–147.

Allison DB, Cui XQ, Page GP, Sabripour M (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nature Reviews Genetics*, **7**, 55–65.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B – Methodological*, **57**, 289–300.

Bertrand C, Benhamed M, Li YF *et al*. (2005) ArabidopsisHAF2 gene encoding TATA-binding protein (TBP)-associated factor TAF1 is required to integrate light signals to regulate gene expression and growth. *Journal of Biological Chemistry*, **280**, 1465–1473.

Blanchard JL (2004) Bioinformatics and systems biology, rapidly evolving tools for interpreting plant response to global change. *Field Crops Research*, **90**, 117–131.

Blasing OE, Gibon Y, Gunther M *et al*. (2005) Sugars and circadian regulation make major contributions to the global regulation of diurnal gene expression in Arabidopsis. *Plant Cell*, **17**, 3257–3281.

Bohnert HJ, Ayoubi P, Borchert C *et al*. (2001) A genomics approach towards salt stress tolerance. *Plant Physiology and Biochemistry*, **39**, 295–311.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Bradley KL, Pregitzer KS (2007) Ecosystem assembly and terrestrial carbon balance under elevated CO_2. *Trends in Ecology and Evolution*, **22**, 538–547.

Busch W, Lohmann JU (2007) Profiling a plant: expression analysis in Arabidopsis. *Current Opinion in Plant Biology*, **10**, 136–141.

Campos H, Cooper A, Habben JE, Edmeades GO, Schussler JR (2004) Improving drought tolerance in maize: a view from industry. *Field Crops Research*, **90**, 19–34.

Casteel CL, O'Neill BF, Zavala JA, Bilgin DD, Berenbaum MR, DeLucia EH (2008) Transcriptional profiling reveals elevated CO_2 and elevated O_3 alter resistance of soybean (*Glycine max*) to Japanese beetles (*Popillia japonica*). *Plant, Cell and Environment*, **31**, 419–434.

Choe SE, Boutros M, Michelson AM, Church GM, Halfon MS (2005) Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome Biology*, **6**, R16.

Clark CM, Cleland EE, Collins SL *et al*. (2007) Environmental and plant community determinants of species loss following nitrogen enrichment. *Ecology Letters*, **10**, 596–607.

Cui XG, Hwang JTG, Qiu J, Blades NJ, Churchill GA (2005) Improved statistical tests for differential gene expression by shrinking variance components estimates. *Biostatistics*, **6**, 59–75.

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK (2004) Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant Journal*, **38**, 366–379.

DeLucia EH, Casteel CL, Nabity PD, O'Neill BF (2008) Insects take a bigger bite out of plants in a warmer, higher carbon dioxide world. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 1781–1782.

Dong QF, Lawrence CJ, Schlueter SD, Wilkerson MD, Kurtz S, Lushbough C, Brendel V (2005) Comparative plant genomics resources at PlantGDB. *Plant Physiology*, **139**, 610–618.

Fleiss JL (1981) *Statistical Methods for Rates and Proportions*. John Wiley and Sons, New York.

Frickey T, Benedito VA, Udvardi M, Weiller G (2008) AffyTrees: facilitating comparative analysis of affymetrix plant microarray chips. *Plant Physiology*, **146**, 377–386.

Friedman N (2004) Inferring cellular networks using probabilistic graphical models. *Science*, **303**, 799–805.

Fuhrer J (2003) Agroecosystem responses to combinations of elevated CO_2, ozone, and global climate change. *Agriculture Ecosystems and Environment*, **97**, 1–20.

Fung RWM, Gonzalo M, Fekete C *et al*. (2008) Powdery mildew induces defense-oriented reprogramming of the transcriptome in a susceptible but not in a resistant grapevine. *Plant Physiology*, **146**, 236–249.

Gibon Y, Blaesing OE, Hannemann J *et al*. (2004) A robot-based platform to measure multiple enzyme activities in Arabidopsis

using a set of cycling assays: comparison of changes of enzyme activities and transcript levels during diurnal cycles and in prolonged darkness. *Plant Cell*, **16**, 3304–3325.

Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. *Genome Biology*, **7**, 76.

Gilks WR, Audit B, De Angelis D, Tsoka S, Ouzounis CA (2002) Modeling the percolation of annotation errors in a database of protein sequences. *Bioinformatics*, **18**, 1641–1649.

Goffard N, Weiller G (2006) Extending MapMan: application to legume genome arrays. *Bioinformatics*, **22**, 2958–2959.

Gong QQ, Li PH, Ma SS, Rupassara SI, Bohnert HJ (2005) Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative Arabidopsis. *Plant Journal*, **44**, 826–839.

Grace JB (2006) *Structural Equation Modeling and Natural Systems*. Cambridge University Press, Cambridge.

Gupta P, Duplessis S, White H, Karnosky DF, Martin F, Podila GK (2005) Gene expression patterns of trembling aspen trees following long-term exposure to interacting elevated CO_2 and tropospheric O_3. *New Phytologist*, **167**, 129–142.

Hammond JP, Bowen HC, White PJ *et al*. (2006) A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New Phytologist*, **170**, 239–260.

Hendriks JHM, Kolbe A, Gibon Y, Stitt M, Geigenberger P (2003) ADP-glucose pyrophosphorylase is activated by posttranslational redox-modification in response to light and to sugars in leaves of Arabidopsis and other plant species. *Plant Physiology*, **133**, 838–849.

Hudson ME (2007) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular Ecology Resources*, **8**, 3–17.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Research*, **31**, e15.

Jackson RB, Linder CR, Lynch M, Purugganan M, Somerville S, Thayer SS (2002) Linking molecular insight and ecological research. *Trends in Ecology and Evolution*, **17**, 409–414.

Jones HG (1999) Use of thermography for quantitative studies of spatial and temporal variation of stomatal conductance over leaf surfaces. *Plant, Cell and Environment*, **22**, 1043–1055.

Joosen RVL, Lammers M, Balk PA *et al*. (2006) Correlating gene expression to physiological parameters and environmental conditions during cold acclimation of *Pinus sylvestris*, identification of molecular markers using cDNA microarrays. *Tree Physiology*, **26**, 1297–1313.

Jordan MI (2004) Graphical Models. *Statistical Science*, **19**, 140–155.

Kaiser WM, Huber SC (2001) Post-translational regulation of nitrate reductase: mechanism, physiological relevance and environmental triggers. *Journal of Experimental Botany*, **52**, 1981–1989.

Kangasjarvi J, Jaspers P, Kollist H (2005) Signalling and cell death in ozone-exposed plants. *Plant, Cell and Environment*, **28**, 1021–1036.

Kendziorski C, Irizarry RA, Chen KS, Haag JD, Gould MN (2005) On the utility of pooling biological samples in microarray experiments. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 4252–4257.

Lamb C, Dixon RA (1997) The oxidative burst in plant disease resistance. *Annual Review of Plant Physiology and Plant Molecular Biology*, **48**, 251–275.

Leakey ADB, Bernacchi CJ, Dohleman FG, Ort DR, Long SP (2004) Will photosynthesis of maize (*Zea mays*) in the US Corn Belt increase in future [CO_2] rich atmospheres? An analysis of diurnal courses of CO_2 uptake under Free-Air Concentration Enrichment (FACE). *Global Change Biology*, **10**, 951–962.

Leakey ADB, Bernacchi CJ, Ort DR, Long SP (2006b) Long-term growth of soybean at elevated [CO_2] does not cause acclimation of stomatal conductance under fully open-air conditions. *Plant, Cell and Environment*, **29**, 1794–1800.

Leakey ADB, Uribelarrea M, Ainsworth EA, Naidu SL, Rogers A, Ort DR, Long SP (2006a) Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO_2 concentration in the absence of drought. *Plant Physiology*, **140**, 779–790.

Leakey ADB, Xu F, Gillespie KM, McGrath JM, Ainsworth EA, Ort DR (2009) The genomic basis for stimulated respiratory carbon loss to the atmosphere by plants growing under elevated [CO_2]. *Proceedings of the National Academy of Sciences*, in press.

Li PH, Mane SP, Sioson AA, Robinet CV, Heath LS, Bohnert HJ, Grene R (2006) Effects of chronic ozone exposure on gene expression in Arabidopsis ecotypes and in *Thellungielia halophila*. *Plant, Cell and Environment*, **29**, 854–868.

Lin MF, Carlson JW, Crosby MA *et al*. (2007) Revisiting the protein-coding gene catalog of *Drosophila melanogaster* using 12 fly genomes. *Genome Research*, **17**, 1823–1836.

Long SP, Ainsworth EA, Leakey ADB, Nosberger J, Ort DR (2006) Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. *Science*, **312**, 1918–1921.

Long SP, Naidu SL (2002) Effects of oxidants at the biochemical, cell and physiological levels, with particular reference to ozone. In: *Air Pollution and Plant Life* (eds Bell JNB, Treshow M), pp. 69–88. John Wiley and Sons, Chichester.

McIntyre LM, Bono LM, Genissel A *et al*. (2006) Sex-specific expression of alternative transcripts in *Drosophila*. *Genome Biology*, **7**.

Michael TP, McClung CR (2003) Enhancer trapping reveals widespread circadian clock transcriptional control in Arabidopsis. *Plant Physiology*, **132**, 629–639.

Miyazaki S, Fredricksen M, Hollis KC *et al*. (2004) Transcript expression profiles of Arabidopsis grown under controlled conditions and open-air elevated concentrations of CO_2 and of O_3. *Field Crops Research*, **90**, 47–59.

Morcuende R, Bari R, Gibon Y *et al*. (2007) Genome-wide reprogramming of metabolism and regulatory networks of Arabidopsis in response to phosphorus. *Plant, Cell and Environment*, **30**, 85–112.

Nettleton D (2006) A discussion of statistical methods for design and analysis of microarray experiments for plant scientists. *Plant Cell*, **18**, 2112–2121.

Ouborg NJ, Vriezen WH (2007) An ecologist's guide to ecogenomics. *Journal of Ecology*, **95**, 8–16.

Pasquer F, Isidore E, Zarn J, Keller B (2005) Specific patterns of changes in wheat gene expression after treatment with three antifungal compounds. *Plant Molecular Biology*, **57**, 693–707.

Poorter H (1993) Interspecific variation in the growth-response of plants to an elevated ambient CO_2 concentration. *Vegetatio*, **104**, 77–97.

Rajeevan MS, Ranamukhaarachchi DG, Vernon SD, Unger ER (2001) Use of real-time quantitative PCR to validate the results of cDNA array and differential display PCR technologies. *Methods*, **25**, 443–451.

Roelofs D, Aarts MGM, Schat H, van Straalen NM (2008) Functional ecological genomics to demonstrate general and specific responses to abiotic stress. *Functional Ecology*, **22**, 8–18.

Rogers A, Gibon Y (2009) Enzyme kinetics: theory and practice. In: *Plant Metabolic Networks* (ed. Schwender J) Springer, Berlin (in press)

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science*, **11**, 508–516.

Scheible WR, GonzalezFontes A, Morcuende R *et al*. (1997) Tobacco mutants with a decreased number of functional nia genes compensate by modifying the diurnal regulation of transcription, post-translational modification and turnover of nitrate reductase. *Planta*, **203**, 304–319.

Scheiner SM, Gurevitch J (2001) *Design and Analysis of Ecological Experiments*, 2nd edn. Oxford University Press, Oxford.

Schlueter JA, Lin JY, Schlueter SD *et al*. (2007) Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. *BMC Genomics*, **8**.

Seki M, Narusaka M, Ishida J *et al*. (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant Journal*, **31**, 279–292.

Sharma N, Cram D, Huebert T, Zhou N, Parkin IAP (2007) Exploiting the wild crucifer *Thlaspi arvense* to identify conserved and novel genes expressed during a plant's response to cold stress. *Plant Molecular Biology*, **63**, 171–184.

Shiu SH, Borevitz JO (2008) The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity*, **100**, 141–149.

Stitt M, Gibon Y, Lunn JE, Piques M (2007) Multilevel genomics analysis of carbon signalling during low carbon availability: coordinating the supply and utilisation of carbon in a fluctuating environment. *Functional Plant Biology*, **34**, 526–549.

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.

Straalen NM van, Roelofs D (2006) *An Introduction to Ecological Genomics*. Oxford University Press, pp. 299.

Sweetlove LJ, Fernie AR (2005) Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. *New Phytologist*, **168**, 9–23.

Talame V, Ozturk NZ, Bohnert HJ, Tuberosa R (2007) Barley transcript profiles under dehydration shock and drought stress treatments: a comparative analysis. *Journal of Experimental Botany*, **58**, 229–240.

Taylor G, Street NR, Tricker PJ *et al*. (2005) The transcriptome of *Populus* in elevated CO_2. *New Phytologist*, **167**, 143–154.

Thimm O, Blasing O, Gibon Y *et al*. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, **37**, 914–939.

Tosti N, Pasqualini S, Borgogni A, Ederli L, Falistocco E, Crispi S, Paolocci F (2006) Gene expression profiles of O_3 treated Arabidopsis plants. *Plant, Cell and Environment*, **29**, 1686–1702.

Travers SE, Smith MD, Bai JF *et al*. (2007) Ecological genomics: making the leap from model systems in the lab to native populations in the field. *Frontiers in Ecology and the Environment*, **5**, 19–24.

Ungerer MC, Johnson LC, Herman MA (2008) Ecological genomics: understanding gene and genome function in the natural environment. *Heredity*, **100**, 178–183.

Urbanczyk-Wochniak E, Usadel B, Thimm O *et al*. (2006) Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. *Plant Molecular Biology*, **60**, 773–792.

Vodkin LO, Khanna A, Shealy R *et al*. (2004) Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. *BMC Genomics*, **5**.

Wang H, Miyazaki S, Kawai K, Deyholos M, Galbraith DW, Bohnert HJ (2003) Temporal progression of gene expression responses to salt shock in maize roots. *Plant Molecular Biology*, **52**, 873–891.

Watkinson JI, Sioson AA, Vasquez-Robinet C *et al*. (2003) Photosynthetic acclimation is reflected in specific patterns of gene expression in drought-stressed loblolly pine. *Plant Physiology*, **133**, 1702–1716.

Weston DJ, Gunter LE, Rogers A, Wullschleger SD (2008) Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Systems Biology*, **2**, 16.

Windsor AJ, Mitchell-Olds T (2006) Comparative genomics as a tool for gene discovery. *Current Opinion in Biotechnology*, **17**, 161–167.

Wullschleger SD, Leakey ADB, St Clair SB (2007) Functional genomics and ecology – a tale of two scales. *New Phytologist*, **176**, 735–739.

Wullschleger SD, Tschaplinski TJ, Norby RJ (2002) Plant water relations at elevated CO_2 – implications for water-limited environments. *Plant, Cell and Environment*, **25**, 319–331.

Zavala JA, Casteel CL, DeLucia EH, Berenbaum MR (2008) Anthropogenic increase in carbon dioxide compromises plant defense against invasive insects. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5129–5133.

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, **4**, 17.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Currently available resources for microarray analysis of plant gene expression.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

