Studying interaction transcriptomes: coordinated analyses of gene expression during plant-microorganism interactions

Paul R.J. Birch and Sophien Kamoun

Many microorganisms establish intimate associations with plants. Whether pathogenic or symbiotic, such interactions involve complex recognition events between plant and microorganism, leading to signalling cascades and regulation of countless genes required for, or associated with, the interaction. Here we introduce the concept of the ‘interaction transcriptome’: the combined components of the transcriptomes of both plant and microorganism that are expressed during the interaction. We discuss how RT-PCR, genomics and bioinformatics approaches are combined to meet the daunting challenge of identifying and characterizing transcriptional changes during infection or symbiosis.
The induction of biochemical-response pathways or the genetic components involved, and the transcriptional changes that occur in the interaction transcriptome can be dissected into three host and three microorganism sub-components. The green portions represent transcripts from genes that are constitutively expressed in plant or microorganism tissue. The relative portions of the three host segments and the three microorganism segments vary depending on the organisms involved. The differentially regulated portions of the interaction transcriptome are most relevant to the interaction and are more likely to be mechanistically involved in its outcome. The corresponding genes offer an excellent prospect for biotechnological manipulation to achieve improved resistance and enhanced symbiosis, and can be readily identified using a series of reverse transcriptase polymerase chain reaction-based methods, genomics and bioinformatics techniques. The constitutively transcribed portions of the interaction transcriptome consist mainly of transcripts from housekeeping genes and other genes involved in basic cellular processes that might not be directly involved in the interaction. However, some of this class of transcripts might correspond to genes that are not differentially regulated but are still relevant to the interaction, such as genes encoding receptors or components of signal transduction pathways. In this case, post-translational modifications might occur that alter the activity of the protein and thus affect the outcome of the interaction. The study of this component of the interaction transcriptome and its corresponding ‘interaction proteome’ offers interesting technical challenges and requires advanced biochemical and proteomics tools.

Methods to study interaction transcriptomes

Differential gene expression profiling

A major challenge in analysing plant-microorganism interactions is often the small amount of biological material available. This limitation has been overcome using PCR-based methods, initially developed for DNA fingerprinting, which allow profiles of gene expression (following conversion of mRNA to cDNA) to be readily visualized. Most notable is differential display (DD) RT-PCR, the use of which has been recently reviewed in both eukaryotes and prokaryotes. Briefly, after cDNA synthesis using reverse transcriptase and an oligo dT primer that anneals to the 3’ polyA tail of mRNA, subsets of cDNA populations for comparison are PCR amplified with short, non-specific oligonucleotide primers, in combination with oligo dT primers, and visualized on polyacrylamide gels.

In the context of plant pathology, DD RT-PCR was first used to isolate both plant and pathogen genes specifically expressed in the interaction between tomato and the fungal pathogen Botrytis cinerea. More recently, it has allowed the isolation of genes up-regulated during the pathogenic interactions between plant and virus, plant and bacterium and plant and nematode, and during the symbiotic associations between plant and arbuscular mycorrhiza and plant and growth-promoting rhizobacterium (PGPR) (Ref. 11). In all cases, key plant genes induced in these interactions were identified. At the physiological level, the association of the PGPR Pseudobacteriolyza with Arabidopsis thaliana resulted in higher levels of plant resistance to the pathogen Erwinia carotovora (biotic stress) and in greater tolerance to drought (abiotic stress). DD RT-PCR identified an Arabidopsis gene, ERD15, which was induced by P. polymyxa. This Arabidopsis gene can also be induced by drought, further demonstrating the activation of common biochemical pathways in response to different stimuli.

Isolation of differentially expressed genes from both host and pathogen or symbiont during their interaction might be dependent on the method of cDNA synthesis. Of the
studies referred to above, only two reported identification of both host and microorganism cDNAs (Refs 9,10). Both these reports involved eukaryote-eukaryote interactions in which both host and microorganism mRNAs possess 3’ polyA tails and were thus converted to cDNA using an oligo dT primer. However, prokaryotic organisms do not generate mRNAs with 3’ polyA tails, and thus synthesis of bacterial CDNA cannot be generated using an oligo dT primer. Recently, using a mixture of 11-mer primers designed to anneal to conserved regions in the 3’ ends of enterobacterial genes, representative cDNAs have been synthesized from the bacterial plant pathogen E. carotovora. Differential gene expression in E. carotovora grown in different media was profiled using cDNA-amplified fragment length polymorphism (cDNA-AFLP).

Like DD RT-PCR, cDNA-AFLP is derived from a DNA fingerprinting method and again involves selective PCR amplification of sub-sets of cDNA populations for comparison on polyacrylamide gels. However, cDNA-AFLP is an improvement on DD RT-PCR in that amplification is specific, using primers with higher annealing temperatures that bind to adaptors ligated to the ends of double-stranded cDNA molecules following restriction digestion.

The work of Alia Dellagi et al. offers the possibility of distinguishing between differentially expressed bacterial and plant genes during the E. carotovora-potato interaction by using different strategies for cDNA synthesis. Synthesis of cDNA from the interaction using an oligo dT primer should produce cDNA specifically derived from eukaryotic mRNA (i.e. from the plant), whereas priming with the 11-mer oligonucleotides should efficiently synthesize cDNA from the bacterium. The two cDNA populations can then be compared by profiling them using either cDNA-AFLP or DD RT-PCR.

The cDNA-AFLP technique has also been effectively used to identify tomato cDNAs that are up-regulated in the resistance response to Cladosporium fulvum, when R protein Cf-9 is activated by the Avr9 protein from the pathogen.

Several plant genes that were rapidly induced by Avr9 elicitation were also up-regulated by wounding, again indicating that common pathways might be activated in both defence and stress. The authors did not report clear identification of pathogen cDNAs in this analysis.

The gene expression profiling approach of either DD RT-PCR or cDNA-AFLP allows all components of the interaction transcriptome (genes that are up- and down-regulated or constitutively expressed; Fig. 1) to be identified simultaneously. Nevertheless, in spite of their documented success, these techniques do not allow rapid, high-throughput generation of cDNA sequence data, because large numbers of PCR primer combinations are required to profile all transcripts within an infected eukaryotic cell.

Large-scale cDNA sequencing (ESTs)
The emergence of low-cost, high-throughput DNA sequencing methods has allowed plant biology to enter the

Figure 2. Approaches to study interaction transcriptomes
Several methods are currently being employed to study interaction transcriptomes. RNA prepared from a particular interaction, and from biological material for comparison (such as host and microorganism cultured separately), is converted initially to cDNA using appropriate primers for synthesis. Populations of cDNAs can be directly compared by profiling on polyacrylamide gels using RNA fingerprinting methods, or can be ‘sampled’ by sequencing cDNAs to yield ESTs (expressed sequence tags). RNA fingerprinting methods include differential display reverse transcriptase polymerase chain reaction (DD RT-PCR) and cDNA-AFLP and can be used to identify cDNAs that are constitutively expressed in either host or microorganism, or are down-regulated during the interaction in either host or microorganism, or are up-regulated during the interaction in either host or microorganism. ESTs generated by direct sequencing of cDNA libraries, including enriched libraries produced by suppression subtractive hybridization (SSH), can be analysed by measuring, for example, GC content to discriminate between cDNAs derived from either host or pathogen. Sequences from cDNAs generated by the above methods can be used either to make microarrays, or to generate 9–11 bp sequence-specific tags for SAGE (serial analyses of gene expression) to produce patterns of gene expression temporally across a range of physiological and biochemical conditions.
‘genomics era’. In particular, projects involving large-scale sequencing of cDNAs [expressed sequence tags (ESTs)] are on-going for a wide variety of crop plants. Such cDNA sample sequencing provides an overview of the genes expressed in a particular cell type or developmental stage. This approach, when combined with bioinformatic analyses, also allows rapid and exhaustive sampling of transcripts that are up-regulated or constitutively expressed in interaction transcriptomes.

An example of the power of ‘interaction’ ESTs is provided by a limited study of Medicago truncatula, a model legume plant and a model for plant–microorganism symbioses.15 A central aspect of the nitrogen-fixing symbiosis with rhizobia is the development of a structure called the root nodule. ESTs were generated from a cDNA library constructed from emerging nodules induced in M. truncatula by the bacterium Sinorhizobium meliloti. Because oligo-dT primers were used for cDNA synthesis, the library contained solely plant cDNAs. Of 389 ESTs, 117 were shown to be up-regulated in the root nodule by reverse northern hybridization15. Among the cDNAs that showed similarities to sequences in databases were 48 clones likely to encode novel nodulin genes, and 33 cDNAs were clones of already known nodulin genes. Thus, generation of EST sequences, and analysis of their expression, is a powerful way to identify genes involved in the nodulation process.

Recently, EST information has emerged from two related plant pathogens, the oomycetes Phytophthora infestans, which causes late-blight on tomato and potato16, and Phytophthora sojae, which causes root and stem rot on soybean17. ESTs generated from cDNA libraries constructed from Phytophthora-infected plant tissue could be of either pathogen or host origin; the challenge is to distinguish between the plant and Phytophthora EST populations using bioinformatic analyses. In these particular cases, it appears that plant and Phytophthora ESTs have markedly different GC content; thus most ESTs can be easily distinguished on this basis. For example, percentage GC content was assessed for sequences from cDNA libraries derived solely from P. sojae and soybean. Both sets of sequences produced distinct, slightly overlapping, normal distribution curves, with the pathogen ESTs averaging 58% GC content and the host ESTs 46% GC content16. A similar analysis of sequences from a P. sojae infected soybean cDNA library revealed ESTs to be clustered around two peaks corresponding to 46% and 58% GC content. Two thirds of the ESTs from this library had a GC content of 58% and were thus predicted to be from the pathogen17. Many of these sequences showed strong identity to ESTs from non-interaction cDNA libraries. Additional cDNAs were confirmed to originate from the pathogen by Southern hybridization. Thus, in silico analyses can prove powerful in annotating sequence data and identifying candidate plant and pathogen genes to be selected for functional analyses, although it remains to be seen whether the clear differences in GC content between plant and Phytophthora cDNA sequences will also be observed in other pathosystems.

cDNA libraries enriched in differentially expressed genes

Methods to enrich libraries with differentially expressed cDNAs are available and can be combined with large-scale sequencing approaches. An example is suppression subtractive hybridization (SSH)18, a powerful technique that has become popular in recent years. SSH combines the selectivity of subtractive hybridization with the sensitivity of PCR. One of its main advantages is that it allows the detection of low-abundance differentially expressed transcripts, such as many of those likely to be involved in signalling and signal transduction, and might thus identify essential regulatory components in several biological processes.

As is the case for traditional subtractive hybridization methods, the cDNA population from which up-regulated (target) sequences are sought is termed the tester, and the cDNA population for comparison is termed the driver. Both tester and driver cDNA populations are digested with a restriction enzyme that recognizes a 4 bp site. The tester cDNA is divided into two equal portions and a different adaptor ligated to each. An excess of driver cDNA is hybridized to each tester sub-population, allowing equalization of high and low abundance cDNAs. Each hybridization mixture is then combined and further allowed to hybridize. After an extension reaction to fill in sticky ends, primers that anneal to each of the adaptors are used to PCR amplify only those molecules with both adaptors (enriched for sequences specific to the tester population), and these can be directly cloned for further study. Molecules with only one type of adaptor (i.e. with inverted terminal repeats) are not exponentially amplified because of the formation of stem-loop structures during the PCR annealing step, in a process called suppression PCR (Ref. 19).

The first application of SSH in the study of plant–microorganism interactions was for the isolation of potato genes that are up-regulated in the HR to P. infestans19. Tester cDNA was prepared from potato leaf material 24 h post-inoculation (hpi) with an isolate of P. infestans that elicits the HR. The driver cDNA was prepared from a compatible (susceptible) interaction with P. infestans, also at 24 hpi. Following SSH, cloned amplification products, arrayed in 384 well microtitre plates and spotted onto nylon membranes, were screened by hybridization to both tester and driver cDNAs. Clones hybridizing only to the former probe were sequenced and were shown to be specifically induced in the HR to P. infestans. PCR with primers designed to anneal to the cDNA sequences obtained by this SSH confirmed that they were derived solely from the plant. Among the HR-associated sequences were cDNAs that showed similarity to genes implicated in apoptosis in animals, providing evidence that programmed cell death in plants and animals might be related. SSH thus provides a ‘targeted EST’ strategy that allows those cDNAs that are up- or down-regulated in an interaction transcriptome (Fig. 1) to be rapidly identified, following subtraction of constitutively expressed cDNAs.
Future prospects: high throughput analyses of gene expression and interaction proteomics

Following the overflow of sequence data generated by the methods outlined above, from both plants and microorganisms, the next phase of study will require compiling an accurate set(s) of genes that are up- or down-regulated during the various plant–microorganism interactions. Such information should lend itself to high throughput, coordinated analyses of gene expression temporally across ranges of conditions by techniques such as microarrays or SAGE (serial analyses of gene expression). This, in turn, should provide information about the complexity of gene regulation, and might help to identify genes in common biochemical or signalling pathways.

Microarrays provide genome-scale information about patterns of gene expression and are currently being applied in several plant EST projects. Moreover, SSH has also recently been combined with cDNA microarray, allowing small-scale targeted EST projects to be analysed in the same way. In principle, either oligonucleotides or PCR amplified cDNAs are robotically printed onto a glass slide and hybridized simultaneously to two comparative cDNA populations, each labelled with a different fluorochrome, for example, cy-5 (red) for one and cy-3 (green) for the other. Red or green fluorescence of a hybridized clone indicates greater relative expression of that sequence in the cDNA population labelled with that fluorochrome, whereas yellow fluorescence indicates equal expression in both. Although available, relational databases for storing data from microarray experiments, and software for pattern-recognition and for extracting multidimensional information from microarray sets, is poorly developed. A list of currently available software packages and websites for analysing microarray data has been reviewed recently. Design of user-friendly tools for microarray data analyses nevertheless remains a major challenge.

The SAGE technique allows a rapid, quantitative and detailed analysis of thousands of transcripts simultaneously. In principle, short transcript-specific tags (9–11 bp) are generated from a cDNA population, and these are concatenated, cloned and sequenced. After sequencing thousands of clones, the frequency of specific tags is assessed, and this reflects the relative abundance of the respective mRNA in the cDNA population. Recently, the SAGE technique was applied to the study of gene expression in the model monocotyledonous plant, rice. In this report, 10 122 tags derived from 5921 genes expressed in rice seedlings were studied. Unexpectedly, the most highly expressed genes were a class of metallothionein genes, accounting for 2.7% of total gene expression.

A potential limitation of both microarrays and SAGE concerns whether they are sufficiently sensitive to detect transcripts derived from the microorganism during the interaction with a plant. As yet, this has not been investigated. However, to follow the expression of low abundance transcripts, and to quantify changes in expression levels accurately, it might be necessary to use quantitative RT-PCR or real-time RT-PCR based on TaqMan chemistry, which has been used to quantify low-level gene expression in mammals.

Studying the interaction transcriptome tells us about the genes that are transcribed during plant–microorganism interactions. It does not tell us if these transcripts are translated to proteins, or if constitutively expressed genes are differentially post-translationally modified. It is known, for example, that during the early stages of pathogen invasion, protein phosphorylation occurs in the plant, and it is possible that constitutively expressed plant proteins are either activated or inactivated by this process. Although not the subject of this review, clearly high throughput analyses of proteins generated during plant–microorganism interactions, or studies of the interaction proteome, will also be required for full characterization of the biochemical and physiological events that are occurring. Recently, two-dimensional polyacrylamide gel electrophoresis was used to identify ectomycorrhiza-regulated cell wall proteins in Pisolithus tinctorius, believed to be involved in cell–cell adhesion during its interaction with Eucalyptus globulus. Thus, future characterizations of plant–microorganism associations might use parallel analyses of both the interaction transcriptome and interaction proteome.

Conclusions

In the past six years, considerable effort has been directed towards the study of resistance in plants to pathogens, particularly focusing on the isolation and characterization of resistance genes in plants and avirulence genes in pathogens. Nevertheless, a dynamic exchange of molecular signals must occur between plants and microorganisms to result in resistance or susceptibility, infection or symbiosis. An important preliminary step in understanding the signalling events and mechanisms involved in plant–microorganism associations is a coordinated study of the transcriptional changes that take place in both the plant and the microorganism. We have therefore coined the term ‘interaction transcriptome’ to describe the sum of the transcripts from both host and microorganism that are expressed during their interaction. An understanding of the interaction transcriptome might lead to the identification of key biochemical or signalling pathways that can be used or manipulated to achieve durable plant resistance, or to promote beneficial symbiotic associations.

Several techniques have been developed for sampling the interaction transcriptome (Fig. 2). In addition to the full genome sequence of Arabidopsis, which should be completed this year, a vast wealth of information is emerging from EST programmes in several plant species, using cDNA libraries generated at various time-points in
a wide range of interactions. Nevertheless, characterization of ESTs from interaction cDNA libraries requires that both host and microorganism DNA sequences can be readily distinguished. EST and full genome sequencing projects are also on-going for many plant pathogens and symbionts. Bioinformatic analyses of ESTs from cDNA libraries derived solely from the host or microorganism will be needed to help to predict the origins of sequences generated from interaction cDNA libraries. Until such analyses are performed, and their results confirmed, researchers should be cautious about the origins of cDNAs (whether from host or microorganism) from interaction cDNA libraries presented in international databases.

An increasing choice of RT-PCR-based methods for isolating differentially expressed genes is contributing to our understanding of interaction transcriptomes. The sensitivity of these techniques facilitates the profiling of transcriptional changes in interactions involving small amounts of biological material. Moreover, they allow low-abundance transcripts, such as those involved in signalling and signal transduction, to be isolated. Such transcripts might escape the less sensitive blanket-sequencing approach of EST strategies. As sequence data accumulates, high throughput expression analyses of all the isolated genes using DNA microarrays or SAGE should provide direct comparison of common or distinct mechanisms, at the transcriptional level, between the different classes of pathogenic or symbiotic plant-microorganism interactions. However, a major bottle-neck in taking this growing body of knowledge about the expression of genes through to full understanding of mechanisms in plant-microorganism interactions, is the lack of high-throughput methods for analysing gene function, such as virus-induced gene silencing (VIGS) (Ref. 27), which might be applicable to a wide range of plant-microorganism interactions. This is an area that is likely to see intensive study over the coming years.

We are entering an era where signal perception and transduction in plant-microorganism interactions is of increasing importance and interest. This is personified by the recent efforts to characterize signalling events following resistance (R) gene activation26, and comparisons between the associated programmed cell death in the HR and apoptosis in animals2. High-throughput methods for studying differential gene transcription, or proteomics approaches for identifying the proteins that they encode26, will revolutionize our understanding of the molecular bases of plant-microorganism interactions.

References
10 Lapopin, L. et al. (1999) Comparative differential RNA display analysis of arbuscular mycorrhiza in Plant sativum wild type and a mutant defective in late stage development. Plant Mol. Biol. 41, 669–677
20 Birch, R.J. et al. (1999) Isolation of potato genes that are induced during an early stage of the hypersensitive response to Phytophthora infestans. Mol. Plant-Microbe Interact. 12, 356–361