

REVIEW ARTICLE

Microarrays in ecology and evolution: a preview

GREG GIBSON

Department of Genetics, North Carolina State University, Gardner Hall, Raleigh NC 27695–7614, USA

Abstract

Microarray technology provides a new tool with which molecular ecologists and evolutionary biologists can survey genome-wide patterns of gene expression within and among species. New analytical approaches based on analysis of variance will allow quantification of the contributions of among individual variation, genotype, sex, microenvironment, population structure, and geography to variation in gene expression. Applications of this methodology are reviewed in relation to studies of mechanisms of adaptation and divergence; delineation of developmental and physiological pathways and networks; characterization of quantitative genetic parameters at the level of transcription ('quantitative genomics'); molecular dissection of parasitism and symbiosis; and studies of the diversification of gene content. Establishment of microarray resources is neither prohibitively expensive nor technologically demanding, and a commitment to development of gene expression profiling methods for nonmodel organisms could have a tremendous impact on molecular and genetic research at the interface of organismal and population biology.

Keywords: cDNA microarray, gene expression profiling, parasitism, population structure, quantitative genomics, symbiosis, variation

Received 23 October 2001; accepted 23 October 2001

Introduction

Microarrays burst onto the scene of molecular biological research in the mid-1990s (Skena *et al.* 1995), and have quickly been established as an essential tool for gene expression profiling in relation to physiology and development. When used in conjunction with classical genetic approaches and the emerging power of bioinformatics, they are much more than a tool because they can force us to change our perspective on the process under study, and the way in which we approach experiments. For the past five years, the primary application of microarrays has been in the identification of sets of genes that respond in an extreme manner to some treatment, or that differentiate two or more tissues. Recently, the analysis and application of microarray technology has become more sophisticated, and it is increasingly clear that genome-wide surveys of transcription have much to contribute to ecology and evolutionary biology. In this review, after briefly describing the technology, I aim to preview some of the

areas in which we might expect rapid advances in the next few years.

There are actually several different ways in which nucleic acid probes can be arrayed at high density for interrogation of labelled mRNA samples. The two most common technologies, cDNA microarrays and oligonucleotide expression arrays, are contrasted in Fig. 1. Only the former is generally applicable to nonmodel organisms at the current time, as it requires only that a large library of cDNAs be available as a source of clones to be arrayed. Usually, microarray preparation is initiated by obtaining end-sequences for several thousand of the clones, and a unique set of these expressed sequence tags (ESTs), is selected for amplification. These products are robotically deposited at a density of around 20–30 clones per square millimetre on the end of a special glass microscope slide or filter, in batches of perhaps 100 slides. The cDNA microarray probe is then hybridized to radioactively or fluorescently labelled cDNA prepared by reverse transcription of mRNA isolated from the cells or tissues of interest. Competitive hybridization of two samples labelled with different dyes, commonly Cy3 and Cy5, allows an estimate of the ratio of transcript abundance in the two RNA samples being compared, for

Correspondence: Greg Gibson. Fax: (919) 515 3355; E-mail: ggibson@unity.ncsu.edu

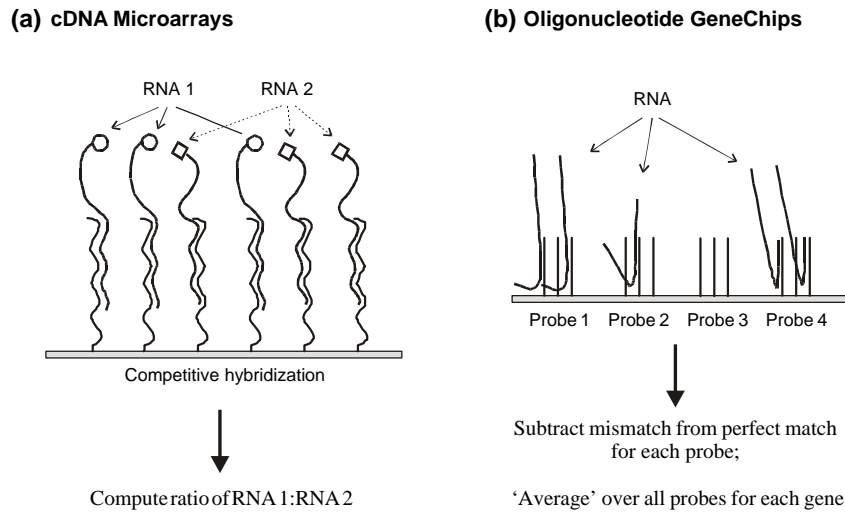


Fig. 1 Comparison of cDNA and oligonucleotide microarrays. (A) In cDNA microarray analysis, nanoliter amounts of concentrated 1–2 kb polymerase chain reaction product derived from cDNAs are deposited at high density in spots on a glass slide or filter. These are hybridized competitively to fluorescently labelled cDNA derived from two different RNA sources, and the ratio of the two signals at each spot reflects the relative levels of transcript abundance. (B) Affymetrix GeneChips® consist of up to 20 microsquares of 25 mer oligonucleotides per gene, including perfect and mismatch pairs (not shown) that will hybridize specifically or nonspecifically to a different part of the same transcript. Each square yields a different intensity measure, reflecting differences in GC content and folding of the RNA, so the values are massaged to produce a measure of gene expression that is contrasted with measures from other chips.

each spot (clone) on the microarray independently (Fig. 1A). The levels of fluorescence or radioactivity are not regarded as a reliable indicator of the absolute level of transcript, but as described below, it is possible to infer changes in gene expression from changes in the signal intensity of each clone relative to the sample mean.

The alternate oligonucleotide technology pioneered by Affymetrix GeneChips® differs in two important respects (Lockhart *et al.* 1996). First, the probes are a set of up to 20 short, 25 mer oligonucleotides that are specific for each gene or exon, along with the related set with single base mismatches incorporated at the middle position of each oligonucleotide. These are synthesized *in situ* on each silicon chip using genome sequence information to guide photolithographic deposition. Second, the arrays are hybridized to a single biotinylated amplified RNA sample (Fig. 1B), and the intensity measure for each gene is currently computed by an algorithm that massages the difference between the match and mismatch measurements and averages over each oligonucleotide. Rather than comparing ratios, inferences are drawn by contrasting differences in magnitude of these intensity scores. This technology is expensive, but has greater genome coverage than microarrays and may be more replicable and comparable across research groups, so is seeing wide application for model organisms such as yeast, flies, *Arabidopsis*, and mice.

To date, most microarray studies have focused on fold-change in transcript abundance as the measure of interest, often employing a common reference sample as the standard

against which experimental treatments are compared. That is, one experimental sample is competitively hybridized with a reference sample that consists of pooled RNAs from multiple treatments, and the fold-difference between two experimental samples is inferred by comparing the two ratio measurements. Ecologists and evolutionary biologists are much more used to dealing with quantitative data than are molecular biologists, so will often be looking to replace analytical terms expressing differences between means (such as fold-change relative to an arbitrary threshold, often twofold), with measures of significance and variance. Happily, there is no reason why standard statistical approaches common to quantitative genetic analysis cannot be applied to microarray data. In many respects microarrays resemble miniature agricultural plots (Kerr & Churchill 2001), and the data can be parsed with linear regression and mixed model analysis of variance. Replicate sample sizes of just five or six will generally be adequate to demonstrate that just a 1.5 fold-change in transcript level of a particular gene is statistically significant, while twice that number may be adequate to study differences as small as 1.2-fold. By contrasting expression relative to the sample mean, reference samples that provide no biological information can be avoided, so these quantitative microarray approaches work well with as few as twice as many replicates as the simplest duplicate experiments. At a cost of just a couple of hundred dollars per microarray, the extra effort is well justified, particularly if it opens up new avenues of research such as those discussed in the remainder of this review.

Mechanisms of adaptive and neutral evolution

One of the first applications of microarrays was in the study of the adaptive response of yeast grown under continuous culture conditions for hundreds of generations in a chemostat. Replicate clones allowed to evolve independently under glucose limitation all showed alterations in the expression of hundreds of genes, many of which are involved in central energy metabolism and metabolite transport (Ferea *et al.* 1999). These results indicated that selection favoured reduced glucose fermentation and improved glucose oxidation, much as occurs during the diauxic shift as glucose is depleted during growth in nonreplenished medium (De Risi *et al.* 1997). Approaches such as genetic fingerprinting (Smith *et al.* 1996) and rational hypothesis testing are now being used to pinpoint the actual mutations that have led to the adaptive response. Analyses of the profiles of fitness increase suggest that a handful of major effect substitutions may lead to significant changes in expression of large numbers of genes, while DNA-DNA hybridization on microarrays confirms that deletion and duplication of one or several genes can contribute to adaptive evolution (Hughes *et al.* 2000). Similar types of study are now being carried out with respect to adaptive evolution of yeast under different nutrient conditions, and of a variety of bacteria and fungi grown in chemostats as well as solid, heterogeneous environments.

Adaptation can also be studied in wild isolates by comparison of whole genome expression profiles. Cavalieri *et al.* (2000) isolated a homothallic strain of *Saccharomyces cerevisiae* from a vineyard in Tuscany, and observed segregation of several morphological and biochemical traits in tetrads after sporulation. As the complete genome sequence of this yeast is available, they were able to query the entire genome for transcriptional differences, and found that up to 6% of the genome showed more than a twofold difference in transcript abundance between the two haploid derivatives. Many of the genes are involved in amino acid uptake and biosynthesis, and it was again suggested that the differences in the transcriptomes are likely to trace to variation in a handful of regulatory loci that either act on hundreds of loci or initiate cascades of transcriptional response. This approach is being extended to higher eukaryotes, such as the fish *Fundulus*, for which 18% of a sample of 900 genes have been shown to differ significantly in expression level in heart muscle from a sample of 15 individuals from three populations along a latitudinal cline (M. Oleksiak, G. Churchill, and D. Crawford, manuscript submitted). Imposition of artificial selection might be envisaged as an aid to dissection of short-term adaptive responses of organisms in the wild, and under conditions of habitat disturbance.

Microarrays are also being used to examine the effects of neutral divergence as a result of mutation accumulation in

the nematode *Caenorhabditis elegans* and in *Daphnia* (M. Lynch, personal communication). Over 70 clonal lines derived from a single isogenic hermaphrodite worm have been propagated asexually for over 1000 generations with an effective population size of one (Vassilieva *et al.* 2000). Clear morphological and behavioural differences have evolved in the strains, with phenotypic expression also being influenced by environmental stressors such as temperature. Global gene expression profiling will provide a new perspective on the inherent variability and covariance of transcriptional variation, and of the power of mutation to induce correlated transcriptional responses at unlinked loci. Similarly, expression profiling of inbred lines of laboratory mice, including recombinant inbred lines, will establish baselines of transcriptional variation that will provide the error term for establishment of the significance of treatment effects in pharmacogenomic research (Pavlidis & Noble 2001; G. Churchill, personal communication).

Delineation of physiological and developmental pathways and networks

The fundamental applications of gene expression profiling for biomedical molecular biology apply equally well to ecological and evolutionary functional genomics. These include: (i) assembly of atlases of gene expression in diverse tissues at certain phases of development and under defined environmental conditions; (ii) drawing hierarchies and networks of gene activity based on temporal profiles of transcript abundance and changes in transcription that accompany mutation or other perturbation; (iii) prediction of gene function based on similarity of transcription profiles to those of known genes; and (iv) dissection of regulatory mechanisms, starting with surveys of transcription factor binding sites upstream of coexpressed genes. Bioinformatic strategies used in these pursuits include various forms of hierarchical clustering (Eisen *et al.* 1995), self-organizing maps (Tamayo *et al.* 1999), principal component analysis (Holter *et al.* 2000), and multiple regression (Bussemaker *et al.* 2001).

Two broad themes in evolutionary biology are likely to benefit from the application of these approaches to organismal biology. One is the utilization of comparative methods to tease apart the functional relevance of correlations between gene expression and the development of a phenotype. As resources are developed for more and more species, it will become possible to study the conservation of regulatory networks on the genome scale. While the recent history of evolution and development has seen great strides in our understanding of how individual components of regulatory pathways evolve (Carroll *et al.* 2001), the ability to survey in parallel the utilization of all classes of genes, rather than just focusing on those key genes that have been well characterized in model organisms, will

provide a fresh perspective on the evolution of gene networks. An early example of this is the adoption of macroarrays — hundreds of clones arrayed on filters and probed with radioactively labelled cDNA — to investigate the developmental bases of caste differences in honeybees (Evans & Wheeler 2000). Microarrays provide a general platform for molecular dissection of dimorphism in any organism.

The second broad theme will be the use of cDNA microarrays to identify candidate genes as modifiers of traits. It is now well recognized that quantitative trait locus (QTL) mapping technology is quite limited in its power to identify the actual genes that are responsible for quantitative variation, and that alternate types of evidence may be required to identify candidate genes that fall within the limits of a QTL interval. Quantitative gene expression profiling of lines that differ for a trait may prove useful in this regard (M. Wayne, personal communication). In fact, since transcript level is one step closer to the phenotype than is genotype, in more general terms the detection of 'transcriptotype'–phenotype associations may emerge as an efficient and productive approach to identification of QTL.

The genetic architecture of complex traits

More powerful mining of microarray data awaits the development and general acceptance of robust methods for the assessment of statistical significance of gene expression differences as well as the parsing of variance components due to different factors under consideration in an experiment. Three groups have recently reported slightly different statistical methods that use the ratio of treatment to experimental variance to assess significance of differences in transcript levels (Kerr, Martin & Churchill 2000; Thomas *et al.* 2001; Wolfinger *et al.* 2001). Figure 2 shows four ways in which transcriptional variance can be presented graphically. The simplest (Fig. 2A) is to plot averaged and normalized expression levels of one treatment against another, as is often done for the two dyes from a single array. Most points fall roughly along a diagonal, and if it is assumed that the variance in gene expression is constant across all genes, then those points that lie outside the limits of the dotted lines representing an arbitrary ratio cut-off can be tagged as responders.

A more rigorous approach is to measure the significance of an effect. Volcano plots (Fig. 2B) of the magnitude of an effect (commonly measured on a log₂ scale such that a value of 1 corresponds to a twofold difference) against the significance (presented as the negative logarithm of the *P*-value from an *F*-test so that increasing significance is at the top of the plot) provide an overall view of the fraction of the genome with large, small, and significant effects. These plots show clearly the consequences of selecting genes on the basis of fold-change rather than significance. By the

former criteria, genes lying in the bottom left and right dark-shaded sectors, which have large but insignificant effects (due to high variance) would be regarded as interesting but are probably false positives. At the same time, a large number of genes with small but highly replicable effects (in the light shaded sector) would be ignored, but these false negatives may well encode regulatory molecules that have a big effect even if the change in transcription is only 1.2-fold. Adjustment of the horizontal line up and down modifies the confidence in the statistical assessment of any particular gene. Whatever statistical threshold is chosen, this is preferable to using an arbitrary fold-change threshold.

Individual genes can be examined by plotting the replicate values from a series of experiments for each treatment on a normalized scale in which fluorescence intensity is measured relative to the sample mean. In the hypothetical example in Fig. 2(C), both sexes were measured under two experimental conditions, and it appears that, whereas males and females differ for treatment one, they have similar expression of the gene for treatment two. ANOVA would probably indicate a significant sex by treatment interaction effect for this gene.

Because it is not clear how to assess the appropriate level of significance for microarrays in which thousands of comparisons are performed, with unknown degrees of independence of the expression of individual genes, a fourth methodology is to plot the variance components of different treatment and/or treatment interaction effects on gene expression (Fig. 2D). Most genes fall in the bottom left hand corner, meaning that neither treatment contributes more than the experimental error to the observed variance, but many genes will fall along one of the axes, meaning that only one of the treatments may have had an effect, or between the axes, meaning that both may have had an effect. Such plots are useful for contrasting the magnitudes of the contributions of different fixed treatment effects, and show how statistical analysis can move well beyond the stark interpretation of differences in ratio of two treatments.

Quantitative genomics using microarray or oligonucleotide array analysis offers the ability to estimate fundamental parameters of gene expression variation, including the additivity, dominance, and heritability of transcription. In a side-by-side comparison of the effects of genotype, sex, and age on adult transcription in two inbred lines of *Drosophila melanogaster*, Jin *et al.* (2001) concluded that over one-half of the genome was affected by sex, up to one-quarter by genotype, and much less than 10% by the contrast of reproductive maturity at one week of age with senescence at six weeks. While it was not possible to estimate the among-individual contribution to variance because single flies do not provide sufficient mRNA for replicate arrays, the use of inbred lines clearly established that wild-type

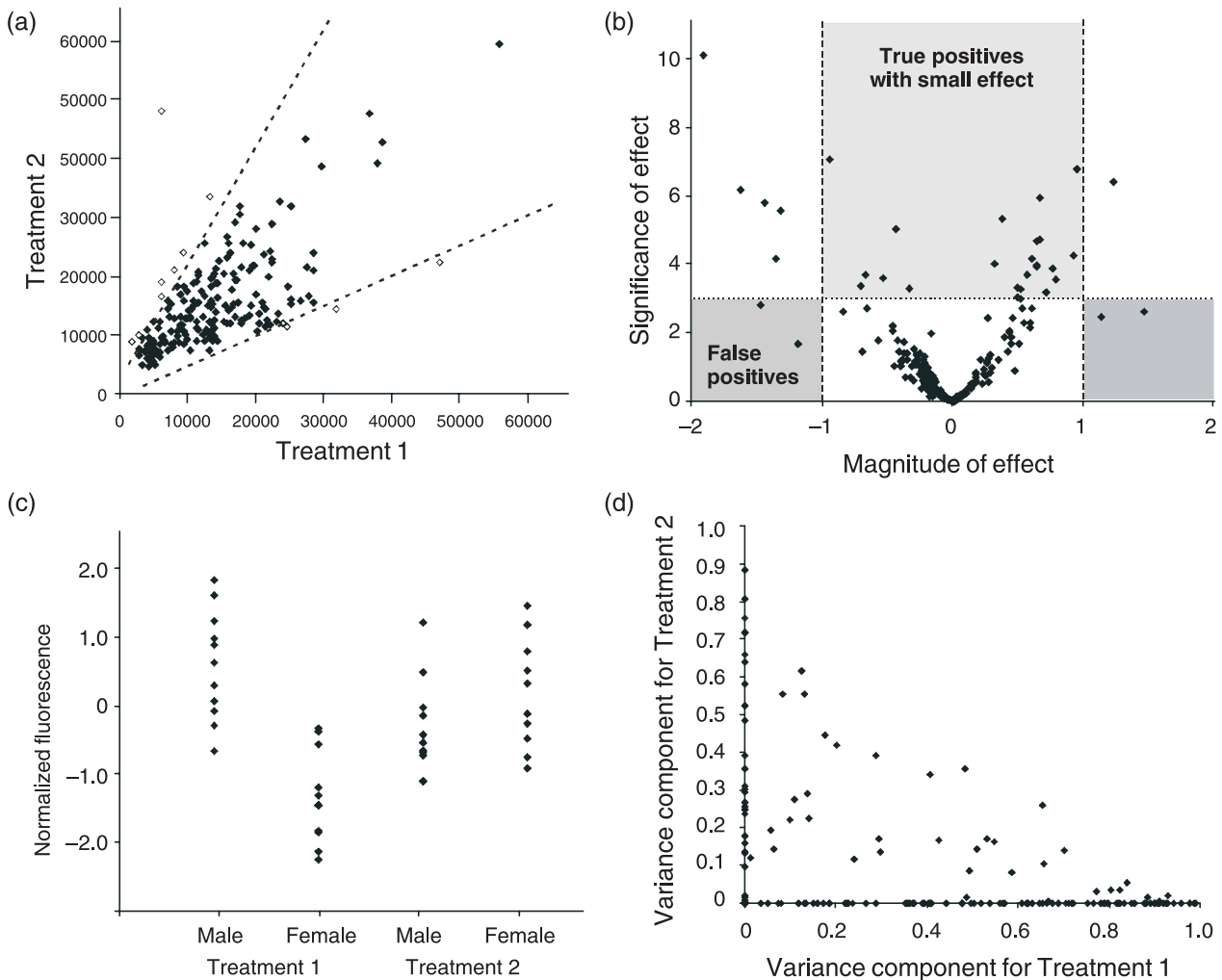


Fig. 2 Four modes of representing gene expression differences. (A) A plot of expression level of one treatment against another. This can be drawn for a single microarray, or for averaged signals over multiple replicates. White diamonds show genes with more than twofold differences, three of which are greater in Treatment 1, eight in Treatment 2. (B) A volcano plot of significance against magnitude. Significance is plotted as the negative logarithm of a P -value assessed by ANOVA, while magnitude is plotted as the difference between mean log base 2 transformed signal intensities. Vertical and horizontal dotted lines represent fold-change and significance thresholds, respectively, and define shaded hexants discussed in the text with respect to false positive and false negative interpretation of effects. (C) Individual gene effects can be plotted after normalization of overall array and dye effects. Each diamond represents one of 12 hypothetical independent measurement of the expression of the gene in each of four samples. (D) Variance components determined from a mixed model analysis of variance (Jin *et al.* 2001) indicate whether Treatment 1 or Treatment 2 has a larger effect on the expression of each gene. Similar plots can be drawn to characterize the magnitude of interaction effects, for example between genotype and sex, age or environment.

genotypes differ in their transcriptional profiles. Taken together with the *Fundulus* results mentioned above, for which the coefficients of experimental variation were similar (between 5 and 10% for each gene: unpublished data), it appears that the heritability of transcription can be substantial and may lie in the same numeric range as that of morphological traits. It will be important now to address questions such as whether all genes have the capacity to vary across a range of genotypes (or whether a subset of the genome is essentially invariant), to document levels of covariation in gene expression among lines, and to assess

the effects of population structure on gene expression in a wide variety of species.

The major limitation of this research strategy is cost because up to 10 arrays per treatment and at least 50 arrays per experiment are required. For this reason, it will make sense for research groups to pool resources and design comprehensive experiments that allow the results of individual groups funded by moderate research grants to be assessed in combination. The potential impact is great, as more is learned about the effect of sex on gene expression, patterns of covariance at the transcriptional level across

environments, and the extent of canalization of gene expression in mutant backgrounds. To the extent that a few regulatory polymorphisms of major effect can affect the expression of a few per cent of the genome, it is quite possible that transcriptional divergence between isolated subpopulations occurs more rapidly than molecular divergence at the DNA level. This scenario has fundamental implications for the appreciation of the importance of genetic drift as a mechanism for divergence in transcription profiles over short timescales (True & Haag 2001), as well as the levels of cryptic variation that are available at the transcriptional level.

Studies of symbiosis and parasitism

Turning to ecological genomics, it should be apparent that genome expression profiling has enormous potential to open up previously intractable systems to genetic analysis, notably in relation to parasitism and symbiosis. Classical molecular biology approaches have sketched the major events that occur during processes such as gall formation by *Agrobacterium tumefaciens* (Zupan & Zambriskie 1997) and nodulation by *Rhizobium* (Freiberg *et al.* 1997), and macroarray analyses have already been initiated to provide a global perspective on gene expression from symbiotic plasmids (Perret *et al.* 1999). As whole genome sequencing of a wide range of pathogenic and symbiotic bacteria elucidates the mechanisms of the evolution of genome content (Ochman & Moran 2001; Ogata *et al.* 2001), microarray analysis will provide a picture of the evolution of gene usage and of interactions between genomes and the environment. Prospects for better understanding of the genetic mechanisms of nematode parasitism will similarly benefit greatly from the ability to profile the expression of arrayed ESTs (Maizels, Blaxter & Scott 2001).

The general approach will be straightforward, building on existing resources. Genome sequencing projects are in place for many parasite species, often focusing on the collection of thousands of EST sequences, rather than complete genome sequence (see for example <http://www.nematodes.org> for a database of nematode genome resources). This is ideal for the development of microarrays as it facilitates the assembly of unigene sets by allowing the removal of the few dozen highly expressed transcripts that tend to constitute up to one quarter of the total mRNA in any organism. Although cDNA libraries can be normalized by molecular biological methods during their construction (Carnici *et al.* 2000), the presence of duplicate clones on an array is wasteful and under some circumstances may bias the statistical analyses. In most cases, it may be preferable to make separate microarrays for the host and parasite/symbiont genomes, where possible by separation of the two source DNA samples prior to library construction. By contrast, the mRNA source for hybridization to the

microarrays may be prepared directly from the tissue samples during the infection process, as cross-hybridization between the genomes should not be common, and purification of the microorganism is likely to alter gene expression. In animal systems such as the dog hookworm, where it is almost impossible to isolate the nematodes, *in vitro* hormonal treatment can be used to mimic aspects of the activation and infection process (Arasu 2001) as a prelude to microarray analysis (unpublished data).

In addition to quantifying the transcriptional responses of host and parasite or symbiont, it should be possible to identify candidate modifiers of infectivity, virulence, and pathogenicity by contrasting gene expression in different isolates under a variety of stimuli. Any gene that shows a significant difference in level of induction or repression between two different types of strain, or whose expression level regresses on an ecological variable of interest, will be flagged for follow-up studies. Because it is always possible (and advisable!) to go back to the plate of clones used to print the microarray and to resequence the clone in its entirety, preliminary microarray analyses can be performed even in the absence of high quality sequence information, which is still expensive and may be a practical impediment to initiation of this approach for many systems. As well as identifying candidate genes, population ecologists will be interested in documenting levels of transcriptional variation and covariation among isolates, with an eye to development of diagnostic arrays that may indicate the potential virulence of isolates in a particular environment or locality.

Genome-wide definition of population structure

Microarrays are beginning to be applied to DNA profiling as a means to characterize genetic differences among isolates and closely related species. Polymorphism for deletions and insertions can be detected as a reduction or elevation, respectively, of the ratio of fluorescence observed when two different genomic DNA samples are labelled and hybridized to a microarray consisting of genomic rather than EST clones. Whole genome comparisons of different strains of various microbes hint that polymorphism for gene content is not uncommon (Riley & Serres 2000), possibly reflecting adaptation to microhabitats, and similar data for yeast and higher eukaryotic genomes should soon be available. As a high throughput alternative to genome sequencing, microarray-based genome content profiling has the potential to reveal the existence of clines and other geographical patterns of structural genome evolution.

Another striking application of DNA content profiling has been the demonstration that aneuploidy accompanies natural selection in a wide range of cell types in culture, from *Escherichia coli* to mammalian cancer lines (Hughes *et al.* 2000). In fact, several instances of apparent co-induction of physically adjacent genes have later been shown to

result from local duplication of that portion of the genome. This is both a cautionary tale with regard to interpretation of expression profile changes, and a hopeful tale with respect to the potential for characterization of mechanisms of adaptation and oncogenesis. Similarly, microarrays might be used to profile heteroplasmy and rates of fixation of rearrangements in the mitochondrial DNA of cells in culture and even somatic differentiation of tissues during ageing (Coller *et al.* 2001).

Applications of microarrays to the characterization of single nucleotide polymorphism (SNPs) using the specificity of hybridization to short oligonucleotides have been reported (Fan *et al.* 2000; Pastinen *et al.* 2000). However, numerous other SNP genotyping methods that are more easy to perform and possibly more accurate exist, and it is unlikely that microarrays will emerge as a general method for profiling of population structure at the DNA level. They may, however, prove to be an invaluable tool for characterization of microbial diversity in soil or enteric samples. Because most organisms harbour a considerable fraction of fast-evolving genes that do not match sequences in any other species, generation of microarrays with these sequences as probes ought to be an efficient way to rapidly survey samples for species content.

Prospective

Gene expression profiling with cDNA microarrays in particular has numerous applications in ecology and evolution, from basic questions concerning the degree of covariation of gene expression, to applications in determination of strain diversity. There are perhaps two major obstacles to utilization of the technique, namely the perceived expense, and the availability of the technology. However, neither of these need constitute serious obstacles to progress. Robotics for handling large numbers of clones and spotting arrays can now be found on most campuses, and are not difficult to operate. While sequencing of several thousand ESTs remains costly, genome projects are already advanced for many species, and the sequencing step is not absolutely necessary in the first phase of development of a microarray resource as interesting clones will always need to be reconfirmed. Although technical difficulties must be overcome by each research group, it is feasible to establish a 5000 clone microarray resource within 12 months of commencing a project, and within the budget of a regular research grant. A typical basic experiment involving 50 or so pairwise hybridizations would then cost between five and 10 thousand US dollars in reagents, with a potentially high benefit to cost ratio. An investment of one million dollars on the part of the ecology and evolution community could see the development of microarray resources for up to 20 organisms of broad interest that would be available for a wide range of problems including many not even

broached in this review. Given the vast sums currently being spent on genome-scale sequencing, the returns would undoubtedly justify the investment.

References

- Arasu P (2001) *In vitro* reactivation of *Ancylostoma caninum* tissue-arrested third-stage larvae by transforming growth factor-beta. *Journal of Parasitology*, **187**, 733–738.
- Bussemaker HJ, Li H, Siggia E (2001) Regulatory element detection using correlation with expression. *Nature Genetics*, **27**, 167–174.
- Carninci P, Shibata Y, Hayatsu N *et al.* (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Research*, **10**, 1617–1630.
- Carroll SB, Grenier J, Weatherbee S (2001) *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science, Oxford, UK.
- Cavalieri D, Townsend J, Hartl D (2000) Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proceedings of the National Academy of Sciences of the USA*, **97**, 12369–12374.
- Coller H, Khrapko K, Bodyak N, Nekhaeva E, Herrero-Jimenez P, Thilly W (2001) High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection. *Nature Genetics*, **28**, 147–150.
- De Risi J, Iyer V, Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
- Eisen MB, Spellman P, Brown PO, Botstein D (1995) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the USA*, **95**, 14863–14868.
- Evans J, Wheeler D (2000) Expression profiles during honeybee caste determination. *Genome Biology*, **2**, Research1.1–1.6.
- Fan J *et al.* (2000) Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Research*, **10**, 853–860.
- Ferea T, Botstein D, Brown PO, Rosenzweig F (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proceedings of the National Academy of Sciences of the USA*, **96**, 9721–9726.
- Freiberg C, Fellay R, Bairoch A, Broughton WJ, Rosenthal A, Perret X (1997) Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature*, **387**, 394–401.
- Holter N, Mitra M, Maritan A, Cieplak M, Banavar J, Federoff N (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proceedings of the National Academy of Sciences of the USA*, **97**, 8409–8414.
- Hughes TR, Roberts CJ, Dai H *et al.* (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature Genetics*, **25**, 333–337.
- Jin W, Riley R, Wolfinger R, White KP, Passador-Gurgel G, Gibson G (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, in press.
- Kerr MK, Churchill G (2001) Statistical design and the analysis of gene expression microarray data. *Genetical Research*, **77**, 123–128.
- Kerr MK, Martin M, Churchill G (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, **7**, 819–837.

- Lockhart DJ, Dong H, Byrne MC *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*, **14**, 1675–1680.
- Maizels R, Blaxter M, Scott A (2001) Immunological genomics of *Brugia malayi*: filarial genes implicated in immune evasion and protective immunity. *Parasite Immunology*, **23**, 327–344.
- Ochman H, Moran N (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science*, **292**, 1096–1099.
- Ogata H, Audic C, Renesto-Audiffren P *et al.* (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science*, **293**, 2093–2098.
- Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvnen A-C (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Research*, **10**, 1031–1042.
- Pavlidis P, Noble W (2001) Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biology*, **2** (10), Research0042.1–0042.15.
- Perret X, Freiberg C, Rosenthal A, Broughton WJ, Fellay R (1999) High-resolution transcriptional analysis of the symbiotic plasmid of *Rhizobium sp.* NGR234. *Molecular Microbiology*, **32**, 415–425.
- Riley M, Serres M (2000) Interim report on genomics of *Escherichia coli*. *Annual Reviews of Microbiology*, **54**, 341–411.
- Schena M, Shalon D, Davis RW *et al.* (1995) Quantitative monitoring of gene expression patterns with a cDNA microarray. *Science*, **270**, 467–470.
- Smith V, Chou K, Lashkari D, Botstein D, Brown PO (1996) Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science*, **274**, 2069–2074.
- Tamayo P, Slonim D, Mesirov J *et al.* (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proceedings of the National Academy of Sciences of the USA*, **96**, 2907–2912.
- Thomas JG, Olson J, Tapscott S, Zhao L-P (2001) An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, **11**, 1227–1236.
- True J, Haag E (2001) Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development*, **3**, 109–119.
- Vassilieva LL, Hook A, Lynch M (2000) The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution*, **54**, 1234–1246.
- Wolfinger R, Gibson G, Wolfinger E *et al.* (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, in press.
- Zupan J, Zambryski P (1997) The *Agrobacterium tumefaciens* DNA transfer complex. *Critical Reviews in Plant Sciences*, **16**, 279–295.

The author is studying the molecular basis of quantitative variation for development and pharmacology, primarily in the fruitfly, *Drosophila melanogaster*. In addition to development of statistical methods for quantification of microarray data, much of his laboratory focuses on association studies between nucleotide polymorphism in candidate genes, and a variety of phenotypes.
