npg

SHORT REVIEW

The next generation of microarray research: applications in evolutionary and ecological genomics

S-H Shiu[1] and JO Borevitz[2]

[1]Department of Plant Biology, Michigan State University, East Lansing, MI, USA and [2]Department of Ecology and Evolution, Division of Biological Sciences, University of Chicago, Chicago, IL, USA

Microarray technology is one of the key developments in recent years that has propelled biological research into the post-genomic era. With the ability to assay thousands to millions of features at the same time, microarray technology has fundamentally changed how biological questions are addressed, from examining one or a few genes to a collection of genes or the whole genome. This technology has much to offer in the study of genome evolution. After a brief introduction on the technology itself, we then focus on the use of microarrays to examine genome dynamics, to uncover novel functional elements in genomes, to unravel the evolution of regulatory networks, to identify genes important for behavioral and phenotypic plasticity, and to determine microbial community diversity in environmental samples. Although there are still practical issues in using microarrays, they will be alleviated by rapid advances in array technology and analysis methods, the availability of many genome sequences of closely related species and flexibility in array design. It is anticipated that the application of microarray technology will continue to better our understanding of evolution and ecology through the examination of individuals, populations, closely related species or whole microbial communities.

Introduction

Microarrays can be broadly defined as tools for massively paralleled ligand binding assays where features (e.g. oligonucleotides) are placed on a solid support (e.g. a glass slide) at high density for recognizing a complex mixture of target molecules (Ekins and Chu, 1999). For biological applications, the features on the arrays can be DNA, RNA, proteins, polysaccharides, lipids, small organic compounds or even whole cells (Hoheisel, 2006). Therefore, microarray technology in principle allows the estimation of target abundance and the detection of biological interactions at the molecular or cellular level. Among the possible feature types, DNA microarrays are the most popular and well developed. In many ways, DNA microarray technology has contributed to fundamental changes in how molecular biologists look at genes and is one of the methodological advances that propelled us into the post-genomics era, that is, from structural to functional analysis of genomes.

Compared to one of the first DNA microarray applications using a spotted array to examine the expression of 64 *Arabidopsis* genes (Schena *et al.*, 1995), microarray technology has come a long way in terms of the number of features available on an array and the range of potential applications. The most well known use of DNA microarrays is for profiling messenger RNA levels; however, DNA microarrays have also been used to detect DNA–protein (e.g. transcription factor-binding site and transcription factor) interactions, alternatively spliced variants, the epigenetic status of the genome (such as methylation patterns), DNA copy number changes and sequence polymorphisms (Kapranov *et al.*, 2003; Borevitz and Ecker, 2004; Stoughton, 2005; Hoheisel, 2006). In addition to the ability to examine a large number of genes in parallel, the success of microarrays can largely be attributed to the versatility and flexibility of array designs. Currently many microarray platforms are available, and custom array designs are feasible and relatively cost efficient. As microarrays are designed to represent a subset of, or the whole genome, it is straightforward to envision their use for understanding evolution at the molecular level and genome evolution. In addition, the advances in using microarrays for mapping and quantitative trait analyses can contribute greatly to ecological genetic studies.

The potential uses for microarray applications in ecology and evolution have been summarized in two insightful reviews by Gibson (2002) and Ranz and Machado (2006). We will not cover the use of microarrays for dissecting the genetic components of adaptive traits. Advances in this area can be attributed to the use of microarrays for genotyping and mapping studies that

Correspondence: Professor JO Borevitz, Department of Ecology and Evolution, Division of Biological Sciences, University of Chicago, Cummings 915E, Chicago, IL 60637, USA.
E-mail: borevitz@uchicago.edu

have been reviewed elsewhere (Borevitz and Nordborg, 2003; Borevitz and Ecker, 2004). In this review, we will first provide an overview of current microarray technology. We will then focus on recent studies applying this technology to the field of evolutionary and ecological genomics, its limitations and future directions.

Microarray technology

The ancestors to DNA microarrays are dot blots and spotted arrays on nylon membranes developed in the late 1980s (Jordan et al., 2001; Southern, 2001). Over the past decade, microarray technology has undergone rapid development in both the number of features on an array and statistical methods for analyzing the array results. The diverse platforms, technical aspects of microarray manufacturing and applications are detailed in several book chapters and reviews (Baldi and Hatfield, 2002; Heller, 2002; Kapranov et al., 2003; Stekel, 2003; Stoughton, 2005; Hoheisel, 2006). These microarray platforms can be classified into two broad categories depending on whether they are manufactured using presynthesized DNA sequences or in situ synthesis methods.

Among platforms using presynthesized sequences, most are fabricated using either pen tip or inkjet deposition. Microarrays based on pen tip deposition (spotted arrays) on glass slides were used nearly exclusively in earlier microarray applications. Typically a glass slide can accommodate ~10 000−40 000 features (~10 features per mm²). Spotted arrays, however, have several quality control issues including non-uniform feature sizes, irregular feature shapes or donuts and variation in quantity deposited across slides and batch runs. These quality issues can be overcome to some extent by the second method, inkjet deposition. Inkjet deposition works the same way as the commonly found inkjet printer. Since there is no physical contact between the spray machinery and the substrate, such as glass slides, the fabrication quality of sprayed arrays is significantly higher than that of spotted arrays. In addition, the feature density of sprayed arrays, such as those from Agilent technology, can be as high as 185 000 features per glass slide, nearly 20-fold higher than spotted arrays. The cost for array fabrication with both methods is low. However, platforms based on presynthesized DNA are not necessarily cheaper than in situ synthesized arrays (detailed below) considering the cost of DNA presynthesis, particularly when presynthesis requires cloning and amplification.

The second type of array platform is fabricated by direct DNA synthesis on the arrays. Several array manufacturers synthesize features via repeated rounds of protection and deprotection of the oligo 5′ reactive end on a solid surface (Figure 1; Fodor et al., 1991; Pease et al., 1994). By combining solid phase chemistry and photolithography, high-density oligo arrays with ~400 000 to >6 250 000 features can be manufactured. The most widely used oligo arrays generated with this approach are manufactured by Affymetrix, where the synthesis of DNA oligos is achieved by successive rounds of deprotection from UV light with many photolithographic masks (Fodor et al., 1991; Pease et al., 1994; Chee et al., 1996). These masks contain openings in predefined locations that allow the synthesis chemistry to take place. A second method is based on digitally controlled
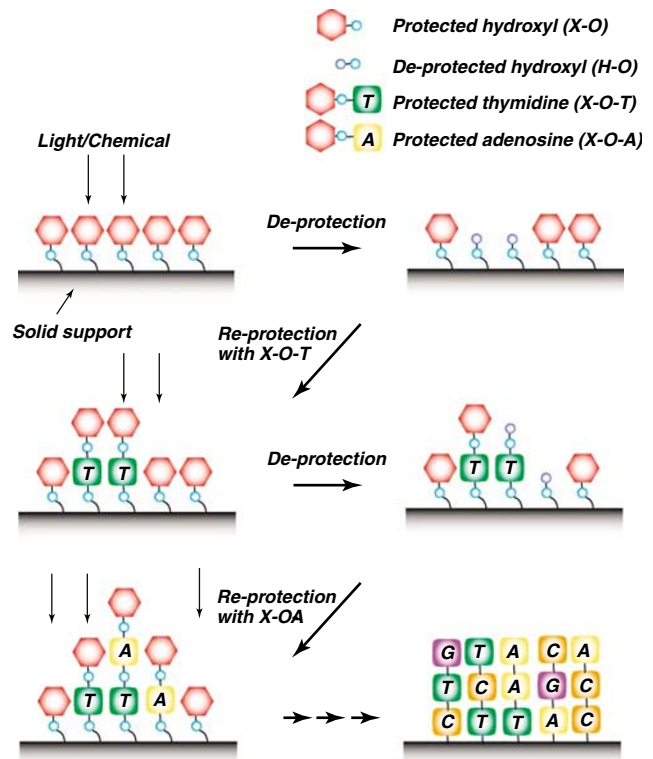

Figure 1 DNA microarray fabrication with in situ synthesis. A solid support with protected hybroxyl (X–O) can be deprotected by either ultraviolet light or chemical. The deprotected hydroxyl is reprotected with the 5′ protected derivative of the nucleotide of interests (e.g. X–O–T or X–O–A). This protection and deprotection process is repeated until desired oligonucleotides are generated.

micromirrors instead of photolithographic masks (Singh-Gasson et al., 1999). One of the drawbacks of mask-based in situ synthesis is the requirement to fabricate a number of masks. This can be costly, particularly if the oligos are long and the total number of features per array is high. This along with compounding errors during synthesis has limited the lengths of features to 25 bp. Therefore, if a relatively small number of arrays are needed, high-density arrays can be manufactured with substantially lower costs with mask-less rather than mask-based methods. The most popular platform for this type of oligo array is provided by NimbleGen Systems, which currently offers flexible length oligos ranging from 24 to 85 bases, instead of the fixed 25 bp oligo used in Affymetrix arrays. Considering feature density, however, Affymetrix arrays currently have a much greater density (>6 250 000 per array) than that of Nimblegen arrays (~390 000). In addition to photolithography-based methods, chemical-based deprotection can also be used for in situ synthesis. This method is used by sprayed array manufacturers such as Agilent.

As the array density can be very high and the array design can be flexible, there is essentially no limitation as to what kinds of features can be included as long as sequence information is available. In addition to cDNA or EST sequences, array features can target predicted genes, intergenic regions, introns, bacterial artificial chromosomes (BACs) or even whole genomes (Borevitz

and Ecker, 2004). This flexibility greatly increases the number of questions that can be addressed through microarray technology. How these microarrays can be used to address questions in evolution and ecology is the focus of the following sections.

Changes in genome content

Genomes constantly experience contraction, expansion and other changes due to deletion, insertion, translocation and inversion events. Ploidy level changes are also frequently encountered, particularly in plants. A better understanding of genome dynamics will help answer not only how the rate of changes can be estimated but also what the nature of selection on genome architecture is. In cancer research, genome content changes have long been recognized as both a diagnostic tool and a mechanism explaining the abnormal growth patterns of cancer cells. These changes can be identified with comparative genomic hybridization (CGH) where DNA sequence copy number differences throughout the entire genome are monitored by hybridizing differentially labeled test DNA and normal reference DNA to normal chromosome spreads (Kallioniemi et al., 1992). As DNA microarrays became available, CGH was conducted using various array platforms containing BAC, cDNA, EST or oligonucleotide features (Pinkel et al., 1998; Carvalho et al., 2004; Ishkanian et al., 2004).

CGH can be widely applied to many questions in genome evolution concerning the patterns and rate of changes. For example, through genome sequence comparisons, it is clear that bacterial genomes are very dynamic with large numbers of gains and losses of genetic material (Doolittle, 1999; Jain et al., 1999). To determine the extent of genome content changes, the genomic DNA of five Escherichia coli strains diverged 25–40 million years ago were hybridized to the spotted nylon membrane array with 4290 open reading frames derived from one sequenced strain (Ochman and Jones, 2000). They found that repeated events of gene acquisition and the concomitant loss of sequences have created divergent lineages of E. coli strains each possessing a unique sets of genes. Using a BAC array representing 12% of the human genome, 63 putative DNA copy number variations were identified between human and the great apes (Locke et al., 2003). In addition, it was found that most copy number variations differentiating the great ape and human genomes occur within genic regions and are localized to intrachromosomal segmental duplicated areas. These examples illustrate the use of DNA microarrays to detect changes in genome content of related species. CGH can be used to examine genome content changes in individuals of a population or populations of the same species. For example, two CGH studies show that there are substantial copy number polymorphisms in the human genome (Iafrate et al., 2004; Sebat et al., 2004; Conrad et al., 2006; Hinds et al., 2006; McCarroll et al., 2006). CGH of DNA from wild-type and knockout yeast strains showed that many of the knockout strains are aneuploids due to intense negative selection pressure in eliminating individuals that cannot complement the loss of function (Hughes et al., 2000).

Array CGH have also been applied to detect variation within or between plant species. By hybridizing genomic DNA from Arabidopsis thaliana accessions to a partial genome RNA expression microarray, ~4000 single-feature polymorphisms (SFPs) were identified in a comparison between the reference strain Columbia (Col) and the strain Landsberg erecta (Ler) (Borevitz et al., 2003). Of these SFPs, ~111 involved potential gene losses. Whole genome arrays have identified >300 genic indels when comparing Arabidopsis ecotypes (Wolyn et al., 2004; Werner et al., 2005a, b). Gene loss is expected to be common since most duplicate genes are expected to be lost relatively quickly (Lynch and Conery, 2000). On the other hand, most genes have at least one relative in the same genome, and some gene families seem to have expanded in a lineage-specific fashion (Shiu et al., 2004, 2005, 2006).

CGH can be conducted both within species and between species in various lineages to discover gene gain and loss events and correlate the gain/loss patterns to parameters such as gene function, expression level, involvement in protein complexes and timing of gene duplication. This type of studies can also generate empirical estimates of gene loss rates. CGH is not restricted to genes since arrays with non-genic sequences or even the whole genomes are either commercially available for several model organisms or can be fabricated on demand. Questions regarding the variation in intergenic sequences including cis-element or other functional elements can be addressed.

Uncovering novel functional elements in the genome

One of the major findings of large-scale cDNA sequencing projects is that many cDNAs were not predicted by gene annotation programs (Yamada et al., 2003; Ota et al., 2004). In addition to evidence of novel genes from cDNA data, expression data from studies using array with features covering the whole genome (tiling array) reveal that significant levels of expression can be detected in intergenic regions in human (Kapranov et al., 2002; Rinn et al., 2003; Bertone et al., 2004), fly (Stolc et al., 2004), A. thaliana (Yamada et al., 2003; Stolc et al., 2005b), and rice (Stolc et al., 2005a; Li et al., 2006). These transcriptional activities raise many questions. The first is whether these positive features are truly transcribed. This has been shown in many cases via PCR-based verification (Kapranov et al., 2002). The second is whether the production of these RNAs is regulated or not. As many of these transcribed intergenic regions are differentially expressed in different human cell lines (Cheng et al., 2005), it is likely that the expression of some of these regions is regulated. Does the expression of these regions have any functional relevance and what is the mechanism of action? Are these transcribed sequences coding or noncoding? Microarray expression analysis across species will reveal if intergenic expression is conserved and pave the way for molecular evolutionary analyses. Functional studies can be performed through analysis of knockouts, as was performed in the functional genomic analyses of small open reading frames in yeast (Kastenmayer et al., 2006).

In addition to genes, the genome contains other functional elements and motifs such as cis-regulatory sequences and matrix attachment sites that serve as the

targets for regulatory or structural protein binding. The mapping of these functional non-genic regions has been undertaken in several studies (van Steensel, 2005). One of the techniques involves combining traditional chromatin immunoprecipitation (ChIP) with DNA microarray hybridization (chip) (Buck and Lieb, 2004; Hanlon and Lieb, 2004; Sikder and Kodadek, 2005). First, the cells expressing the DNA-binding protein of interest or its tagged version is permeated with crosslinking agents to crosslink the protein and its DNA targets. After lysing the cells, chromatin is sheared into ~500 bp fragments, and fragments that are associated with the protein are enriched by immunoprecipitation. The DNA fragments are released from the protein by reversing the cross-linking reaction, ligated to linkers, amplified with linker primers, labeled with fluorescent dyes and hybridized to the microarray. The ChIP-chip approach was first applied to identifying transcription factor targets in yeast (Ren et al., 2000; Iyer et al., 2001). With the availability of microarrays covering the whole genome, the number of ChIP-chip applications is expected to rise quickly. In the genome evolution context, an exciting use of ChIP-chip is in comparison of binding sites of orthologous transcription factors or transcription factor family members. This information will provide insight into the dynamics of *cis*-regulatory element evolution and how the functions of transcription factor duplicates diverge. Heterologous expression of transcription factors from another species followed by ChIP-chip can determine how binding site preferences have diverged genome-wide *in vivo*.

Evolution of regulatory networks

In order to explain how species such as humans and chimps with highly similar, and even identical, genes can differ so substantially in their anatomy, physiology, behavior and ecology, it was suggested that evolutionary differences are more often based on changes in the mechanisms controlling the expression rather than on amino-acid changes (King and Wilson, 1975). Although this conjecture is mainly supported by evidence from metazoan development, it underlies the importance of transcriptional regulation in phenotypic evolution. Several studies show that substantial differences exist in gene expression between related species (Enard et al., 2002; Meiklejohn et al., 2003; Ranz et al., 2003; Rifkin et al., 2003; Nuzhdin et al., 2004; Rustici et al., 2004; Gilad et al., 2006). For example, Rifkin et al. (2003) examined variation in genome-wide gene expression among *Drosophila simulans*, *D. yakuba* and four strains of *D. melanogaster* during the start of metamorphosis. In addition to finding that gene expression differs significantly between species and between strains, they found that the expression patterns of transcription factor genes are relatively more stable than their downstream targets. Cross-species comparisons of expression patterns can also reveal the selection regime on gene expression. By examining expression levels of genes among humans and three other primates, Gilad et al. (2006) found evidence for stabilizing selection in the expression of certain genes. Interestingly, they also found that the expression of a number of human genes experienced lineage-specific selection, judging from significantly elevated or reduced expression in the human lineage compared to the other primate lineages.

In addition to studies focused on differences between species, several recent studies have contributed significantly to our understanding on expression variation at the population level. Assessment of expression variation between populations is the first step in understanding if it is evolutionarily relevant and affects fitness in a reasonable fashion. For example, widespread expression variation has been uncovered in genes involving in amino-acid metabolism, sulfur assimilation and processing and protein degradation among natural isolates of budding yeast (Townsend et al., 2003). Taking one step further, similar experiments on budding yeast was conducted in the presence of copper sulfate, an anti-microbial agent used in vineyard to investigate the relationships between genetic variation in gene expression and phenotypic variation (Fay et al., 2004). Interestingly, among 633 genes with significant expression differences between yeast strains, only 44 were associated with the presence of copper sulfate. The expression variation that can be ascribed to copper sulfate provides insights into the molecular basis of naturally occurring traits and its selection.

Several other aspects of regulatory networks can be analyzed with microarrays. ChIP-chip can be used to identify DNA target regions of not only transcription factors but also other proteins that interact with genomes, such as matrix proteins and proteins responsible for chromatin remodeling. RNA immunoprecipitation-chip (RIP-chip) can also be used to determine genome-wide RNA protein binding interactions (Gerber et al., 2004; Schmitz-Linneweber et al., 2005). Another important mechanism of transcriptional regulation is through modulation of the epigenetic states of promoters by cytosine methylation. Several microarray-based approaches have been developed to determine the location of methyl cytosines in the genome (van Steensel, 2005; Schumacher et al., 2006) and recently the whole genome of Arabidopsis at 35 bp resolution (Zhang et al., 2006).

The integration of methylation, chromatin precipitation, gene expression and other functional genomics data contributes to a genome-wide description of regulatory networks. In the near future, comparable data sets generated in related species will lead to a better understanding of the nature of regulatory changes, the rate of changes in evolutionary time and ultimately their role in causing phenotypic differences between organisms and adaptation.

Phenotypic plasticity vs gene expression

Behavioral and phenotypic plasticity is a crucial determinant of organism's fitness as they result in appropriate responses to environmental variation such as seasonal changes and reproductive opportunities. As complex phenotypes are the consequences of the interaction between genes and environment, an essential step for understanding complex traits and their evolution in general is to identify the proximal mechanism of behavioral and phenotypic plasticity (Gibson, 2002). Several studies have been carried out attempting to identify expression changes in this context.

The first example is a comparison of expression profiles between nursing and foraging bees using cDNA

microarrays (Whitfield *et al.*, 2003). As the transition to foraging in the honey bee involves environmentally modulated behavioral changes that are associated with changes in brain structure and neurochemistry, the expression profile differences between brains of nursing and foraging bees were compared. Of the approximately 5000 genes tested, 39% showed significant change in transcript abundance between these two types of bees. Although the nurse-to-forager transition is age-related, they were able to compare nurse and precocious forager bees of the same age and found that most gene expression differences between these two types of bees are due to behavior instead of age. In addition, brain messenger RNA profiles of individual bees were found to be a good predictor whether it is a nurse or forager.

Another interesting study regarding plasticity and gene expression is a comparison of the expression profiles between sneaker and anadromous male salmon (Aubin-Horth *et al.*, 2005a, b). Typical anadromous male salmon undergo marine migration before homing to freshwater to spawn. Interestingly, some males mature earlier at greatly reduced size without leaving fresh water and adopt the alternative sneaker tactic. By comparing the expression profiles of brains from age-matched sneaker and immature males, they found 15% of ~3000 genes tested showed significant differences, and individuals with the same reproductive tactic had similar expression profiles (Aubin-Horth *et al.*, 2005a). They further disentangle the effect of rearing environments by comparing immature and sneaker males under wild or hatchery-like environments; expression of 225 genes were affected by the interaction between reproductive tactic and rearing environment (Aubin-Horth *et al.*, 2005b).

These studies exemplify the use of microarrays to identify gene expression changes that can be attributed to behavioral or phenotypic differences between genetically similar individuals. Although major questions remain unanswered, such as how these changes are triggered by the environment and if the behavioral and phenotypic differences are due to some or all the expression changes, microarrays can be more broadly used to dissect expression differences between individuals or populations in different environments.

Metagenomics and microbial community structure

Metagenomics is the culture-independent cloning, sequencing and analysis of microbial DNA extracted directly from an environmental sample (Riesenfeld *et al.*, 2004; Schloss and Handelsman, 2005). Several metagenomic studies have turned up many microbes unknown to traditional microbiology. For example, using a whole genome shotgun approach, more than 1800 microbial species, including 148 previously unknown bacterial 'species' were found in seawater samples from the Sargasso Sea (Venter *et al.*, 2004). In addition, more than 1.2 million genes were predicted in these samples, and nearly 70 000 of them are potential novel genes. More environmental sequencing efforts are underway, and sequences specific to each microbial species will increase dramatically in the near future.

As species or population-specific sequences become available, a microarray with these specific sequence features can be fabricated to quickly sample the diversity and abundance of microbial communities (Cook and Sayler, 2003; Eyers *et al.*, 2004). For example, a small subunit (SSU) ribosomal RNA (rRNA) gene array with probes representing 8903 non-redundant sequences has been fabricated to analyze the microbial diversity in natural outdoor aerosols (Desantis *et al.*, 2005). In addition to showing that the SSU array is capable of distinguishing various species present in the sample, they also found the concentrations of rRNA from five bacterial species correlate well with their hybridization intensities on the array. Similar SSU arrays have been used to detect and quantify individual bacterial species in a more complicated mixture representing >100 species (Palmer *et al.*, 2006). Although the number of species analyzed is likely far lower than the complexity encountered in a typical environmental sample, these studies illustrate the future use of these 'barcode' arrays to monitor microbial communities to monitor the interactions between microbial species and changes in community structures under different environments. The outcome of these studies will provide insights for understanding not only the dynamics of microbial communities but also the effects of environmental parameters on microbial adaptation.

Arrays have also been designed for examining host–pathogen interactions. These include soybean and two pathogenic *Phytophthora* species (Moy *et al.*, 2004) and human and malaria pathogen (Winzeler, 2006). In addition to examining the interaction between host and pathogen genes through their expression patterns, the variation in the responses of individuals to pathogens can be revealed by the high-density microarrays covering both host and pathogen genomes. Furthermore, such variation can be further examined under various ecologically relevant conditions to capture the dynamics between hosts and pathogens in a more realistic setting.

Experimental design and array data analysis

A typical microarray experiment involves comparisons of sample from various sources in an attempt to disentangle the contribution of a single factor and interactions between factors. Therefore, a successful microarray experiment requires consideration of the biological question in mind, treatments and comparisons that minimize the unwanted variation or noise, and a sound experimental design maximizing the information return while minimizing the cost (Leung and Cavalieri, 2003). Poorly designed array experiment may produce data that are not suitable for answering the biological questions or hypotheses, incur unnecessary cost or have insufficient replications. These issues are discussed in details in two excellent reviews by Churchill (2002) and Yang and Speed (2002).

A microarray typically consists of thousand to million features. While it provides a wealth of information leading to many interesting discoveries, analyzing the vast amount of data generated can be a daunting task. Although a number of statistical methods and software have been developed for microarray data processing and analysis, how the data sets are best analyzed remains a complicated question. Among the software available,

Bioconductor, an open source and open development software project for the analysis of genomic data primarily based on the R programming language, contains a number of program packages for microarray data analyses and is arguably the most comprehensive resource for such applications (Gentleman *et al.*, 2004). Although the analysis tools have come a long way compared to just 5 years ago, there is substantial disagreement in, for example, which normalization method to use, which statistics should be used for reporting significant differences, and (for Affymetrix arrays) whether mismatch probe information is useful or not. The appropriate choice of analysis tools depends on the experimental design, the array technology used, the number of replications and the questions at hand.

Limitations of microarrays

One major limitation of microarrays is crosshybridization between features and related sequences in addition to the targets intended. For spotted arrays with long features such as cDNAs or ESTs, the significance of crosshybridization has been evaluated using microarrays of several large gene families (Evertsz *et al.*, 2001; Xu *et al.*, 2001). It was found that the crosshybridization signal drops off quickly as sequence identities decrease. For example, for genes that are 80% identical, the signal level is ~ 5 to 10% of the perfect match. Therefore, the impact of crosshybridization may not be as serious if the features of interests do not have close relatives. However, for metagenomic applications, crosshybridization is likely a more significant issue due to the relatedness of array features, such as SSUs. As for oligonucleotide microarrays, although they have greater specificity than cDNA microarrays, crosshybridization remains a concern (Kane *et al.*, 2000; Dai *et al.*, 2002). For oligo arrays, one way to alleviate the crosshybridization problem is by excluding the features that have a high potential to crosshybridize. Another possibility is to optimize the hybridization conditions. However, both of these approaches would require knowledge of the target sequences or genomes and will not be useful for most cross-species comparisons.

A related problem in cross-species comparisons using microarrays is caused by sequence divergence between species. For expression studies, the hybridization intensity differences between species are the contribution of both expression level differences and differential hybridization (Gilad *et al.*, 2005). One solution is to first hybridize the genomic DNA of the species of interest to the microarray to detect SFPs (Borevitz *et al.*, 2003). These SFPs are likely probes with sequence divergence between these species and can be excluded from further analysis. In addition to the specificity issue, there is also the problem of sensitivity where the target sequences may be of low abundance in the samples or only limited material is available for sample preparation. This issue may be overcome by augmenting the number of signals per transcript or by amplifying all sequences within the sample (Nygaard and Hovig, 2006). However, the amplification step can be biased and sequences may be differentially amplified or there may be a nonlinear relationship between the original relative abundance of sequences and the amplified sample.

Another major limitation for applying microarray technology to any discipline is the availability of sequences. The research community has somewhat bypassed this problem by analyzing DNA or RNA from species related to the organisms the arrays are designed against. Although this approach has been very fruitful, uncertainty in the genome contents of related species is an important concern (Gilad *et al.*, 2005). In addition, studies using arrays of a related species provide an incomplete picture since the arrays do not have elements specific to the related species. The good news is that new sequencing technologies with greatly increased throughput, efficiency and lower cost are being developed (Shendure *et al.*, 2004). For example, using fiber-optic slides with picoliter-sized reaction wells, 1.6 million sequencing reactions can take place simultaneously, generating ~ 25 million bases in 4 h (Margulies *et al.*, 2005). Another recently published high-throughput sequencing method involves 'polymerase colony' (polony) where millions of sequencing-by-ligation reactions are performed using common laboratory equipment (Shendure *et al.*, 2005). Although there are some limitations in these technologies, such as relatively much shorter read length and assembly issues, their rapid development suggests that truly high-efficiency and low-cost sequencing will be within the reach of individual laboratories in the next few years.

Conclusion

The availability of genome sequences has contributed greatly to our understanding of structural aspects of genomes. Functional analyses of components of genomes are dominated by large-scale gene expression analyses, particularly those using microarrays. We have discussed how microarrays can be applied to study some of the questions in genome evolution and ecology and what the current limitations are. The limited sequence availability may not be as much of a concern since many genome sequencing projects are underway and the sequencing technology has advanced very rapidly in both throughput and cost-efficiency. In addition, microarray technology is still improving with higher feature density, lower fabrication cost and, most importantly, more flexible array design. This, along with the ongoing effort in developing open source software for microarray analyses are reasons to be optimistic that microarray results will be robust and analyses will be relatively straightforward. In this review we have sampled the uses of microarray for comparisons that provide better understanding of ecological and evolutionary processes and phenomenon. Arrays for non-model organisms are either available now or will be available soon. Their creative use in the study of evolution and ecology is something to watch for in the near future.

Acknowledgements

References

Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA (2005a). Alternative life histories shape brain gene expression profiles in males of the same population. *Proc Biol Sci* **272**: 1655–1662.

Aubin-Horth N, Letcher BH, Hofmann HA (2005b). Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon. *J Hered* **96**: 261–278.

Baldi P, Hatfield GW (2002). *DNA Microarrays and Gene Expression: From Experiments to Data Analysis and Modeling*. Cambridge University Press: Cambridge. p 230.

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**: 2242–2246.

Borevitz JO, Ecker JR (2004). Plant genomics: the third wave. *Annu Rev Genom Hum Genet* **5**: 443–477.

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D et al. (2003). Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res* **13**: 513–523.

Borevitz JO, Nordborg M (2003). The impact of genomics on the study of natural variation in Arabidopsis. *Plant Physiol* **132**: 718–725.

Buck MJ, Lieb JD (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**: 349–360.

Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B (2004). High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* **57**: 644–646.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D et al. (1996). Accessing genetic information with high-density DNA arrays. *Science* **274**: 610–614.

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**: 1149–1154.

Churchill GA (2002). Fundamentals of experimental design for cDNA microarrays. *Nat Genet* **32** (Suppl): 490–495.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006). A high-resolution survey of deletion polymorphism in the human genome. *Nat Genet* **38**: 75–81.

Cook KL, Sayler GS (2003). Environmental application of array technology: promise, problems and practicalities. *Curr Opin Biotechnol* **14**: 311–318.

Dai H, Meyer M, Stepaniants S, Ziman M, Stoughton R (2002). Use of hybridization kinetics for differentiating specific from non-specific binding to oligonucleotide microarrays. *Nucleic Acids Res* **30**: e86.

Desantis TZ, Stone CE, Murray SR, Moberg JP, Andersen GL (2005). Rapid quantification and taxonomic classification of environmental DNA from both prokaryotic and eukaryotic origins using a microarray. *FEMS Microbiol Lett* **245**: 271–278.

Doolittle WF (1999). Lateral genomics. *Trends Cell Biol* **9**: M5–M8.

Ekins R, Chu FW (1999). Microarrays: their origins and applications. *Trends Biotechnol* **17**: 217–218.

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P et al. (2002). Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340–343.

Evertsz EM, Au-Young J, Ruvolo MV, Lim AC, Reynolds MA (2001). Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* **31**: 1182, 1184, 1186 passim.

Eyers L, George I, Schuler L, Stenuit B, Agathos SN, El Fantroussi S (2004). Environmental genomics: exploring the unmined richness of microbes to degrade xenobiotics. *Appl Microbiol Biotechnol* **66**: 123–130.

Fay JC, McCullough HL, Sniegowski PD, Eisen MB (2004). Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae. *Genome Biol* **5**: R26.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**: 767–773.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80.

Gerber AP, Herschlag D, Brown PO (2004). Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast. *PLoS Biol* **2**: E79.

Gibson G (2002). Microarrays in ecology and evolution: a preview. *Mol Ecol* **11**: 17–24.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006). Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* **440**: 242–245.

Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP (2005). Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res* **15**: 674–680.

Hanlon SE, Lieb JD (2004). Progress and challenges in profiling the dynamics of chromatin and transcription factor binding with DNA microarrays. *Curr Opin Genet Dev* **14**: 697–705.

Heller MJ (2002). DNA microarray technology: devices, systems, and applications. *Annu Rev Biomed Eng* **4**: 129–153.

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006). Common deletions and SNPs are in linkage disequilibrium in the human genome. *Nat Genet* **38**: 82–85.

Hoheisel JD (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* **7**: 200–210.

Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D et al. (2000). Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* **25**: 333–337.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y et al. (2004). Detection of large-scale variation in the human genome. *Nat Genet* **36**: 949–951.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP et al. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* **36**: 299–303.

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533–538.

Jain R, Rivera MC, Lake JA (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci USA* **96**: 3801–3806.

Jordan B (2001). DNA arrays for expression measurement: a historical perspective. In: Jordan B (ed) *DNA Microarrays: Gene Expression Applications*. Springer-Verlag: Berlin. pp 3–12.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F et al. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**: 818–821.

Kane MD, Jatkoe TA, Stumpf CR, Lu J, Thomas JD, Madore SJ (2000). Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* **28**: 4552–4557.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP et al. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* **296**: 916–919.

Kapranov P, Sementchenko VI, Gingeras TR (2003). Beyond expression profiling: next generation uses of high density oligonucleotide arrays. *Brief Funct Genom Proteom* **2**: 47–56.

Kastenmayer JP, Ni L, Chu A, Kitchen LE, Au WC, Yang H et al. (2006). Functional genomics of genes with small open reading frames (sORFs) in S. cerevisiae. *Genome Res* **16**: 365–373.

King MC, Wilson AC (1975). Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.

Leung YF, Cavalieri D (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet* **19**: 649–659.

Li L, Wang X, Stolc V, Li X, Zhang D, Su N et al. (2006). Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* **38**: 124–129.

Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D *et al.* (2003). Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* **13**: 347–357.

Lynch M, Conery JS (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376–380.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC *et al.* (2006). Common deletion polymorphisms in the human genome. *Nat Genet* **38**: 86–92.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL (2003). Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci USA* **100**: 9894–9899.

Moy P, Qutob D, Chapman BP, Atkinson I, Gijzen M (2004). Patterns of gene expression upon infection of soybean plants by *Phytophthora sojae*. *Mol Plant Microbe Interact* **17**: 1051–1062.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004). Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol Biol Evol* **21**: 1308–1317.

Nygaard V, Hovig E (2006). Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Res* **34**: 996–1014.

Ochman H, Jones IB (2000). Evolutionary dynamics of full genome content in *Escherichia coli*. *EMBO J* **19**: 6637–6643.

Ota T, Suzuki Y, Nishikawa T, Otsuki T, Sugiyama T, Irie R *et al.* (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nat Genet* **36**: 40–45.

Palmer C, Bik EM, Eisen MB, Eckburg PB, Sana TR, Wolber PK *et al.* (2006). Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res* **34**: e5.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc Natl Acad Sci USA* **91**: 5022–5026.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D *et al.* (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207–211.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**: 1742–1745.

Ranz JM, Machado CA (2006). Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol Evol* **21**: 29–37.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I *et al.* (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.

Riesenfeld CS, Schloss PD, Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38**: 525–552.

Rifkin SA, Kim J, White KP (2003). Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet* **33**: 138–144.

Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S *et al.* (2003). The transcriptional activity of human chromosome 22. *Genes Dev* **17**: 529–540.

Rustici G, Mata J, Kivinen K, Lio P, Penkett CJ, Burns G *et al.* (2004). Periodic gene expression program of the fission yeast cell cycle. *Nat Genet* **36**: 809–817.

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

Schloss PD, Handelsman J (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol* **6**: 229.

Schmitz-Linneweber C, Williams-Carrier R, Barkan A (2005). RNA immunoprecipitation and microarray analysis show a chloroplast pentatricopeptide repeat protein to be associated with the 5′ region of mRNAs whose translation it activates. *Plant Cell* **17**: 2791–2804.

Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P *et al.* (2006). Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res* **34**: 528–542.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P *et al.* (2004). Large-scale copy number polymorphism in the human genome. *Science* **305**: 525–528.

Shendure J, Mitra RD, Varma C, Church GM (2004). Advanced sequencing technologies: methods and goals. *Nat Rev Genet* **5**: 335–344.

Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM *et al.* (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**: 1728–1732.

Shiu SH, Byrnes JK, Pan R, Zhang P, Li WH (2006). Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc Natl Acad Sci USA* **103**: 2232–2236.

Shiu S-H, Karlowski WM, Pan R, Tzeng YH, Mayer KF, Li WH (2004). Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. *Plant Cell* **16**: 1220–1234.

Shiu SH, Shih MC, Li WH (2005). Transcription factor families have much higher expansion rates in plants than in animals. *Plant Physiol* **139**: 18–26.

Sikder D, Kodadek T (2005). Genomic studies of transcription factor-DNA interactions. *Curr Opin Chem Biol* **9**: 38–45.

Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR *et al.* (1999). Maskless fabrication of light-directed oligonucleotide microarrays using a digital micro-mirror array. *Nat Biotechnol* **17**: 974–978.

Southern (2001). DNA microarrays: history and overview. *Meth Mol Biol* **170**: 1–15.

Stekel D (2003). *Microarray Bioinformatics*. Cambridge University Press: Cambridge. p 280.

Stolc V, Gauhar Z, Mason C, Halasz G, van Batenburg MF, Rifkin SA *et al.* (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**: 655–660.

Stolc V, Li L, Wang X, Li X, Su N, Tongprasit W *et al.* (2005a). A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray. *Plant Mol Biol* **59**: 137–149.

Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC *et al.* (2005b). Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays. *Proc Natl Acad Sci USA* **102**: 4453–4458.

Stoughton RB (2005). Applications of DNA microarrays in biology. *Annu Rev Biochem* **74**: 53–82.

Townsend JP, Cavalieri D, Hartl DL (2003). Population genetic variation in genome-wide gene expression. *Mol Biol Evol* **20**: 955–963.

van Steensel B (2005). Mapping of genetic and epigenetic regulatory networks using microarrays. *Nat Genet* **37** (Suppl): S18–S24.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Werner JD, Borevitz JO, Uhlenhaut NH, Ecker JR, Chory J, Weigel D (2005a). FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170**: 1197–1207.

Werner JD, Borevitz JO, Warthmann N, Trainer GT, Ecker JR, Chory J *et al.* (2005b). Quantitative trait locus mapping and DNA array hybridization identify an FLM deletion as a cause for natural flowering-time variation. *Proc Natl Acad Sci USA* **102**: 2460–2465.

Whitfield CW, Cziko AM, Robinson GE (2003). Gene expression profiles in the brain predict behavior in individual honey bees. *Science* **302**: 296–299.

Winzeler EA (2006). Applied systems biology and malaria. *Nat Rev Microbiol* **4**: 145–151.

Wolyn DJ, Borevitz JO, Loudet O, Schwartz C, Maloof J, Ecker JR *et al.* (2004). Light-response quantitative trait loci identified with composite interval and extreme array mapping in *Arabidopsis thaliana*. *Genetics* **167**: 907–917.

Xu W, Bak S, Decker A, Paquette SM, Feyereisen R, Galbraith DW (2001). Microarray-based analysis of gene expression in very large gene families: the cytochrome *P*450 gene superfamily of *Arabidopsis thaliana*. *Gene* **272**: 61–74.

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ *et al.* (2003). Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* **302**: 842–846.

Yang YH, Speed T (2002). Design issues for cDNA microarray experiments. *Nat Rev Genet* **3**: 579–588.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H *et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**: 1189–1201.