

Microarrays for microbiologists

S. Lucchini, A. Thompson and J. C. D. Hinton

Author for correspondence: J. C. D. Hinton. Tel: +44 1603 255352. Fax: +44 1603 255076. e-mail: jay.hinton@bbsrc.ac.uk

Molecular Microbiology, Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, UK

Keywords: DNA microarray, gene expression profiling, microbial genomotyping

Background

We are witnessing a remarkable change in the scale of molecular microbiological research and we are entering an era of ‘big science’. In the past decade we have moved from a time when entire research papers were based on the sequencing of a single gene or operon to a single paper describing the sequence of a whole genome. The completion of microbial genomes is continuing apace, with 37 genome sequences completed, and 142 in progress worldwide (http://www.tigr.org/tdb/mdb/mdbcomplete.html). The availability of this level of genetic information has spawned the terms ‘functional genomics’, ‘transcriptomics’ (Velculescu et al., 1997) and ‘proteomics’ (Wasinger et al., 1995), which describe the large-scale application of mass mutagenesis, gene expression profiling and global protein analysis. In this review we assess the role that gene expression profiling has begun to play in microbiology, discuss the potential for ‘genomotyping’ and consider future applications.

Assessment of transcription at the genomic scale has been achieved with DNA microarrays, which are glass slides containing an ordered mosaic of the entire genome as a collection of either oligonucleotides (oligonucleotide microarrays) or PCR products representing individual genes (commonly referred as cDNA microarrays).

The development of microarrays has been fuelled by the application of robotic technology to routine molecular biology, rather than by any fundamental breakthrough. The classical Southern and Northern blotting approaches for the detection of specific DNA and mRNA species (Southern, 1975; Alwine et al., 1977, 1979) provided the technological basis for microarray hybridization with fluorescently labelled cDNA. The idea of depositing multiple DNA spots representing different genes onto a solid surface is also nothing new, having been used by Blattner’s group to investigate Escherichia coli gene expression on membranes (macroarrays) as long ago as 1993 (Chuang et al., 1993). Commercially available macroarrays have continued to produce useful data, and should be considered before recourse to microarrays (Tao et al., 1999). The recent application of robotics to achieve high spotting densities of DNA on glass slides was innovative and facilitates the construction of microarrays containing up to 50000 genes on a single microscope slide (DeRisi et al., 1996; Shalon et al., 1996). This allows a single hybridization to be performed against multiple replicates of a single bacterial genome, or against copies of several unrelated genomes on a single glass slide. The development that has facilitated the reproducible comparison of gene expression between two samples, and hence between experiments, is dual fluorescent labelling (Schena et al., 1995). Simultaneous hybridization of two cDNA populations labelled with the fluorescent dyes Cy3 and Cy5 allows accurate assessment of relative levels of gene expression, which is unaffected by hybridization variability or the differences between individual microarrays which can complicate macroarray experiments.

Microarrays as a research tool

Microarrays allow us to produce a ‘gene expression profile’ or ‘signature’ for a particular organism under certain environmental conditions. Since the first report

---

**Fig. 1.** The dramatic increase in the number of publications involving DNA microarrays. ■, All microarray papers; □, microbial microarray papers only.
**Table 1. Applications of microarray technology to microbiological research**

<table>
<thead>
<tr>
<th>Application</th>
<th>Organism</th>
<th>Type of array*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to stress/environmental change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diauxic shift</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>DeRisi et al. (1997)</td>
</tr>
<tr>
<td>Growth in minimal/rich media</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Wodicka et al. (1997)</td>
</tr>
<tr>
<td>Heat shock, cold shock, growth with galactose/glucose</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Lashkari et al. (1997)</td>
</tr>
<tr>
<td>Growth in minimal/rich media</td>
<td><em>E. coli</em></td>
<td>cDNA membranes</td>
<td>Tao et al. (1999)</td>
</tr>
<tr>
<td>Growth in minimal/rich media</td>
<td><em>E. coli</em></td>
<td>cDNA glass slides</td>
<td>Wei et al. (2001)</td>
</tr>
<tr>
<td>IPTG, heat shock</td>
<td><em>E. coli</em></td>
<td>cDNA membranes and glass slides</td>
<td>Richmond et al. (1999); Wei et al. (2001)</td>
</tr>
<tr>
<td>Aerobic/anerobic growth</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Ye et al. (2000)</td>
</tr>
<tr>
<td>Aerobic/anerobic growth</td>
<td><em>B. subtilis</em></td>
<td>cDNA glass slides</td>
<td>Posas et al. (2000)</td>
</tr>
<tr>
<td>Osmotic stress</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Oh &amp; Liao (2000)</td>
</tr>
<tr>
<td>Growth in different carbon sources</td>
<td><em>E. coli</em></td>
<td>cDNA glass slides</td>
<td>Brocklehurst &amp; Morby (2000)</td>
</tr>
<tr>
<td>Metal-ion tolerance</td>
<td><em>E. coli</em></td>
<td>cDNA membranes</td>
<td>de Saizieu et al. (2000)</td>
</tr>
<tr>
<td>Quorum sensing</td>
<td><em>Streptococcus pneumonia</em></td>
<td>HDA</td>
<td></td>
</tr>
<tr>
<td>Response to changes that affect tryptophan</td>
<td><em>E. coli</em></td>
<td>cDNA glass slides</td>
<td>Khodursky et al. (2000a)</td>
</tr>
<tr>
<td>metabolism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell-cycle-associated gene expression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcriptional programme of sporulation in yeast</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Chu et al. (1998); Primig et al. (2000)</td>
</tr>
<tr>
<td>Identification of cell-cycle-regulated genes</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Spellman et al. (1998)</td>
</tr>
<tr>
<td>Analysis of mitotic cell cycle</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Cho et al. (1998)</td>
</tr>
<tr>
<td>Transcriptional changes during competence</td>
<td><em>Strep. pneumoniae</em></td>
<td>cDNA membranes</td>
<td>Rimini et al. (2000)</td>
</tr>
<tr>
<td>development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of cell-cycle-regulated genes</td>
<td><em>C. crescentus</em></td>
<td>cDNA glass slides</td>
<td>Laub et al. (2000)</td>
</tr>
<tr>
<td>Dissection of regulatory circuitry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effects of mutation in RNA polymerase II</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Holstege et al. (1998)</td>
</tr>
<tr>
<td>components, determination of mRNA stability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Promotor sequence analysis of coexpressed genes</td>
<td><em>Sacch. cerevisiae</em></td>
<td>‘<em>in silico</em> biology’</td>
<td>Wolfsberg et al. (1999); Zhang (1999)</td>
</tr>
<tr>
<td>Effects of nucleosome depletion on gene expression</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Wyrick et al. (1999)</td>
</tr>
<tr>
<td>Effects of gene dosage</td>
<td><em>Sacch. cerevisiae</em></td>
<td>HDA</td>
<td>Galitski et al. (1999); Giaever et al. (1999)</td>
</tr>
<tr>
<td>Effects of gene dosage</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Hughes et al. (2000b)</td>
</tr>
<tr>
<td>Snf/Swi protein complex mutants and nucleosome</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Sudarsanam et al. (2000)</td>
</tr>
<tr>
<td>structure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pdr1p/pdr3p</em> mutants (transcriptional activators</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>DeRisi et al. (2000)</td>
</tr>
<tr>
<td>involved in pleiotropic drug resistance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation of iron uptake</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Yun et al. (2000)</td>
</tr>
<tr>
<td><em>marA</em> mutants (transcriptional activator involved in drug resistance)</td>
<td><em>E. coli</em></td>
<td>cDNA membranes</td>
<td>Barbosa &amp; Levy (2000)</td>
</tr>
<tr>
<td>Characterization of the Fap1p zinc-sensitive</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Lyons et al. (2000)</td>
</tr>
<tr>
<td>regulon</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulation of glycolytic enzyme genes</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA membranes</td>
<td>Lopez &amp; Baker (2000)</td>
</tr>
<tr>
<td>Identification of copper regulon</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Gross et al. (2000)</td>
</tr>
<tr>
<td>Chromosomal position and gene expression</td>
<td><em>Sacch. cerevisiae</em></td>
<td><em>In silico</em></td>
<td>Cohen et al. (2000a)</td>
</tr>
<tr>
<td>Characterization of IHF regulon</td>
<td><em>E. coli</em></td>
<td>cDNA membranes</td>
<td>Arfin et al. (2000)</td>
</tr>
<tr>
<td>Function of histone deacetylase</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Bernstein et al. (2000)</td>
</tr>
<tr>
<td>Translational regulation</td>
<td><em>Sacch. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Kuhn et al. (2001)</td>
</tr>
</tbody>
</table>
of DNA microarray technology in 1995 (Schena et al., 1995; DeRisi et al., 1996; Lockhart et al., 1996) the potential of DNA microarrays has certainly captured the imagination of biologists worldwide. Naturally, we welcome the ability to monitor global gene expression in a single experiment rather than relying on the ‘one gene = one postdoc’ approach to eventually elucidate the function of all bacterial genes. However, there has also been a concern that this genome-wide approach might signal a move towards ‘non-hypothesis-driven’ or ‘data-driven’ science (Brent, 1999), a term that has been used rather pejoratively. It is clear that scientific inference

<table>
<thead>
<tr>
<th>Application</th>
<th>Organism</th>
<th>Type of array*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detection of point mutation polymorphism</td>
<td><em>M. tuberculosis</em></td>
<td>HDA</td>
<td>Gingeras et al. (1998)</td>
</tr>
<tr>
<td>Detection of gene deletions/insertions</td>
<td><em>E. coli</em></td>
<td>cDNA membranes</td>
<td>Riehle et al. (2001)</td>
</tr>
<tr>
<td>Genotyping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detection of point mutation polymorphism</td>
<td><em>M. tuberculosis</em></td>
<td>cDNA glass slides</td>
<td>Behr et al. (1999)</td>
</tr>
<tr>
<td>Strain comparison</td>
<td><em>Sacc. cerevisiae</em></td>
<td>HAD</td>
<td>Winzeler et al. (1999a)</td>
</tr>
<tr>
<td>Species identification and detection of point mutations in rpoB</td>
<td><em>M. tuberculosis</em></td>
<td>cDNA glass slides</td>
<td>Salama et al. (2000)</td>
</tr>
<tr>
<td>Drug target characterization/identification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kinase inhibitors</td>
<td><em>Sacc. cerevisiae</em></td>
<td>HDA</td>
<td>Gray et al. (1998)</td>
</tr>
<tr>
<td>Effects of immunosuppressant FK506</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Marton et al. (1998)</td>
</tr>
<tr>
<td>Isoniazid</td>
<td><em>M. tuberculosis</em></td>
<td>cDNA glass slides</td>
<td>Wilson et al. (1999)</td>
</tr>
<tr>
<td>Effect of amino acid biosynthesis inhibitor sulfometuron methyl</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Jia et al. (2000)</td>
</tr>
<tr>
<td>Antifungal agents</td>
<td><em>Sacc. cerevisiae</em></td>
<td>HDA</td>
<td>Bammert &amp; Fostel (2000)</td>
</tr>
<tr>
<td>Cellular response to bacterial infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infection of monocytes (THP-1) with <em>Listeria monocytogenes</em></td>
<td>Human</td>
<td>HDA, cDNA membranes, cDNA colony filters</td>
<td>Cohen et al. (2000b)</td>
</tr>
<tr>
<td>Infection of intestinal epithelial cells (HT-29 and T84) with <em>Salmonella enterica</em></td>
<td>Human</td>
<td>cDNA membranes</td>
<td>Eckmann et al. (2000)</td>
</tr>
<tr>
<td>Infection of macrophages cells (RAW 264.7) with <em>Sal. typhimurium</em></td>
<td>Mouse</td>
<td>cDNA membranes</td>
<td>Rosenberger et al. (2000)</td>
</tr>
<tr>
<td>Infection of epithelial cells (lung carcinoma A549 cell line) with <em>Pseudomonas aeruginosa</em></td>
<td>Human</td>
<td>cDNA glass slides</td>
<td>Ichikawa et al. (2000)</td>
</tr>
<tr>
<td>Infection of intestinal epithelial cells (HeLa 229) with <em>Chlamydia trachomatis</em></td>
<td>Human</td>
<td>cDNA membranes</td>
<td>Dessus-Babus et al. (2000)</td>
</tr>
<tr>
<td>Infection of bronchial epithelial cells (BEAS-2B) with <em>Bordetella pertussis</em></td>
<td>Human</td>
<td>HDA</td>
<td>Belcher et al. (2000)</td>
</tr>
<tr>
<td>In vivo effects of a commensal bacterium on the gene expression of the ileal epithelium</td>
<td>Mouse</td>
<td>HDA</td>
<td>Hooper et al. (2001)</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification of genetic markers</td>
<td><em>Sacc. cerevisiae</em></td>
<td>HDA</td>
<td>Winzeler et al. (1998)</td>
</tr>
<tr>
<td>Characterization of adaptive evolution</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Ferea et al. (1999)</td>
</tr>
<tr>
<td>RNA surveillance</td>
<td><em>Sacc. cerevisiae</em></td>
<td>HDA</td>
<td>Lelivelt &amp; Culbertson (1999)</td>
</tr>
<tr>
<td>Prediction of protein subcellular localization</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Diehn et al. (2000)</td>
</tr>
<tr>
<td>Analysis of topoisomerase function at replication forks</td>
<td><em>E. coli</em></td>
<td>cDNA glass slides</td>
<td>Khodursky et al. (2000b)</td>
</tr>
<tr>
<td>Role of RNA polymerase II subunit Rpb9 on transcription elongation</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Hemming et al. (2000)</td>
</tr>
<tr>
<td>Prediction of gene function by comparing expression profiles of deletion mutants</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Hughes et al. (2000a)</td>
</tr>
<tr>
<td>Genome-wide location of DNA-binding proteins</td>
<td><em>Sacc. cerevisiae</em></td>
<td>cDNA glass slides</td>
<td>Ren et al. (2000); Vishwanath et al. (2001)</td>
</tr>
</tbody>
</table>

*HDA, high-density oligonucleotide array.
uses a combination of deductive and inductive reasoning (Kell & King, 2000). Functional genomics allows us to make new and unexpected links between the function of unrelated and hitherto uncharacterized genes, and suggest hypotheses, which must subsequently be tested by more traditional methods of molecular genetics and biochemistry (Hughes et al., 2000a). An example lies with the increasing numbers of proteins which have unexpectedly been found to have dual cellular functions, such as enolase and aconitases: in addition to being a glycolytic enzyme enolase is also a vital constituent of the RNA degradosome (Py et al., 1996), and aconitases, besides their catalytic role in the TCA cycle, have been shown to act as a post-transcriptional regulator by binding mRNA (Tang & Guest, 1999). Genomic-scale research may be termed ‘non-hypothesis-driven’ science, but we suggest it should be viewed positively as it is likely to reveal the function of many genes which have been missed by more conventional approaches. The need for this is apparent when considering the genome sequence of *E. coli*, which still contains 1632 (38%) FUN genes (of unknown function; Hinton, 1997), which remain to be functionally characterized (Nelson et al., 2000). The role of FUN genes will not be discovered without the application of functional genomic technologies combined with creative experiments.

The explosive growth in the numbers of reviews discussing microarray technology has now been followed by many papers describing results obtained from gene expression microarray profiling (Fig. 1). Genomic and post-genomic approaches are likely to revolutionize our ability to understand how micro-organisms act, both in the laboratory and in the real world. But what effect has this new approach had on molecular microbiologists in general, and what difference is it likely to make in the future?

The answer depends on one’s field of research. Yeast researchers have already embraced microarrays as a useful and productive tool, as exemplified by 60% of the publications in Table 1. In other areas of microbiology, the application of microarray technology to bacteria has been slower. *E. coli*, *Mycobacterium tuberculosis* and other bacteria were the subject of just 37% of the microbial papers involving microarrays. The slow application of microarrays to academic bacterial research probably reflects the complex nature of this technology. Many pharmaceutical and biotechnology companies are successfully using bacterial microarrays to drive programmes of novel drug development, and this industrial experience suggests that technical problems will not be a barrier. We hope that work in our own and other laboratories worldwide will soon produce a raft of informative data concerning bacterial gene expression.

**Microbial gene expression profiling**

Microarrays have already been used to perform high-quality experiments, which have improved our understanding of microbial environmental responses and global gene expression (Fig. 2; Table 1) and the concept of the ‘global transcriptional response’ as a detailed molecular phenotype is now gaining acceptance (Hughes et al., 2000a). The first global transcriptional profile was obtained at the resolution of individual genes for *Saccharomyces cerevisiae*; Pat Brown’s group did this with gene-specific microarrays (DeRisi et al., 1997; Lashkari et al., 1997) and Affymetrix utilized oligonucleotide ‘gene chips’ (Wodicka et al., 1997). Eighteen months later, the first microarrays involving the whole genome of two prokaryotes were described: *M. tuberculosis* (Behr et al., 1999) and *E. coli* (Richmond et al., 1999; Tao et al., 1999). As genome sequencing projects are being completed and new applications of microarrays are being developed, the potential of microarray analysis are being rapidly applied to other micro-organisms (Table 1). It is not always appreciated
that a complete genome sequence is not essential for a microarray project. Interesting results can be obtained from microarrays assembled from partial genome data, or from an uncharacterized gene library, where spots of interest are sequenced once their expression profile has been determined (Pennis, 2000).

Yeast has been the micro-organism of choice for many research groups to analyse cell-cycle-associated gene expression and the effects of various environmental changes, such as osmotic shock, temperature shock, presence of DNA-damaging agents and growth in minimal or rich media (Table 1). The new level of analysis provided by whole-genome expression profiling has revealed the complexity of the cellular response to major changes in metabolism, exemplified by work on yeast diauxic shift (DeRisi et al., 1997). The expression levels of 1840 genes (30% of a total of 6116 genes tested) were found to be affected by the transition from anaerobic to aerobic growth. Similar complexity of gene expression in the transcriptional programme was reported for yeast going through the mitotic cell cycle (Cho et al., 1998; Spellman et al., 1998) or sporulation (Chu et al., 1998). Many of the responsive genes that were identified had previously been designated as FUN.

To understand the large amount of data created by microarrays, Mike Eisen developed a computer program to cluster genes according to their expression profiles (Eisen et al., 1998). Based on the important observation that functionally related genes show similar patterns of expression, identification of well-characterized genes that are co-expressed with FUN genes can give important clues towards function. Using this tool, Chu et al. (1998) defined seven sequential temporal classes of genes induced during yeast sporulation.

One of the most impressive examples of the use of microarrays for bacterial research has been provided by recent work on Caulobacter crescentus (Laub et al., 2000). The definition of the cell cycle of C. crescentus by microarray analysis revealed that 572 of 2966 genes (19.3%) were cell-cycle-dependent. Not only were a number of classes of cell-cycle-induced genes identified, but also the proportion dependent upon the global cell cycle regulator CtrA was recognized for the first time. This study led to recognition of the role of 11 novel sensor kinases and 5 new sigma factors. The identification of cascades of gene expression during the Caulobacter cell cycle is an important landmark for bacterial research. We look forward to similar studies describing gene expression cascades during sporulation of Bacillus subtilis, and during E. coli cell division.

Definition of entire regulons

Because cell cycle and environmental changes are multifactorial, involving important metabolic changes, it is difficult to unravel the role played by specific global gene regulators. The definition of important regulons by the use of appropriate regulatory mutants provides the framework for a better understanding of complex cellular responses. This approach has led to the global characterization of the IHF and MarA regulons of E. coli (Arfin et al., 2000; Barbosa & Levy, 2000), but does not distinguish between direct and indirect effects of regulatory mutations. More recently, DNA microarrays have been used to combine gene expression profile analysis with the localization of binding sites for DNA-binding proteins on the yeast genome. This approach involved the formaldehyde cross-linking of proteins to the DNA, followed by a modified chromatin precipitation procedure. After cell disruption, proteins were immunoprecipitated to enrich for DNA fragments containing DNA-binding sites. The enriched DNA was then amplified, labelled and hybridized to a microarray containing all yeast intergenic regions. This new approach permitted the identification of genes which were directly controlled by the regulatory proteins Gal4 and Ste12 (Ren et al., 2000).

Analysis of gene expression in vivo

The use of microarrays for the study of bacterial infection at the level of gene expression is in its infancy (Cummins & Relman, 2000; Hauf–fort & Hinton, 2000). We require improved technology for in vivo study of infection, both from the host’s and the pathogen’s point of view. Bacterial mRNA is less stable than eukaryotic mRNA, and is generally not polyadenylated, complicating mRNA purification (Sarkar, 1996). This does not cause problems for in vitro analyses, but does complicate the isolation of sufficient, good-quality bacterial mRNA from complex environments such as mammalian tissue. Advances have recently been made in applying the linear RNA amplification method (Eberwine et al., 1992), which has been used to extract sufficient eukaryotic mRNA for microarray analyses from small amounts of mammalian tissue (Lockhart et al., 1996; Luo et al., 1999; Wang et al., 2000). The successful application of this linear RNA amplification approach to polycistronic and non-polyadenylated bacterial mRNA remains to be demonstrated. An alternative way to decrease the amount of bacterial RNA required for the labelling of mRNA involves the new concept of genome-specific primers. Thirty-seven genome-directed primers (GDPs) were predicted to be sufficient to prime all genes in the M. tuberculosis genome, and were used to label mycobacterial RNA (Talaat et al., 2000). The cDNA probes generated by GDPs showed improved sensitivity and specificity when compared with probes obtained by random priming, and allowed a degree of selective amplification of mycobacterial RNA from a mix containing mammalian RNA.

Mammalian gene microarrays have recently been used to study host–pathogen interactions from the viewpoint of the host, by identifying gene expression patterns induced by the presence of a pathogen (Manger & Relman, 2000). Several in vitro studies have explored the effects of infection on the mRNA expression profile of human cells (Table 1). The effects of Listeria monocytogenes (Cohen et al., 2000b), Salmonella enterica (Eckmann et al., 2000) and Salmonella typhimurium
(Rosenberger et al., 2000) have recently been reviewed by Cummings & Relman (2000). Briefly, these studies reveal a specific host response, which is modulated by different host factors (Rosenberger et al., 2000). Other groups are using a more complex approach by comparison of the human cellular transcriptional signatures of pathogenic strains carrying well-defined mutations to get a more detailed view of the mechanisms underlying pathogen clearance (Manger & Relman, 2000).

**Genomotyping and microbial evolution**

Phylogenetic classification based on rRNA/rDNA and signature sequences is usually able to provide an accurate classification of micro-organisms above the species level (Gupta, 2000). However, it has become evident that lateral gene transfer is an important mechanism of evolution for prokaryotes, complicating phylogenetic analysis based on a small number of genes. Differences can be considerable between closely related bacterial pathogens: certain serovars of *Salmonella enterica* contain more than a megabase of sequence information than others (Ochman et al., 2000). Preliminary analysis suggests that more than 700 ORFs are present in *Salmonella typhi* and not in *Sal. typhimurium* (P. O’Gaora, personal communication). Acquisition of new DNA is clearly an important mechanism for the adaptation of bacteria to new ecological niches, as shown by the example of the acquisition of pathogenicity islands, which encode virulence factors and were probably acquired at different times (Groisman & Ochman, 1997). Whole-genome based methods are thus required to determine the repertoire of virulence genes found in bacterial pathogens. Strain comparison by hybridizing genomic DNA to microarrays (genomotyping) is a more realistic approach than the whole-genome sequencing of dozens of strains. Gene-specific microarrays have been used to compare the entire genome of a *Mycobacterium bovis* vaccine strain with the closely related *M. tuberculosis* H37Rv, resulting in a ‘gene-specific fingerprint’ at a resolution of approximately 2 kb (Behr et al., 1999). Alternatively, mutant alleles or single nucleotide polymorphisms (SNPs) can be detected with high-density oligonucleotide microarrays, which carry much shorter targets (typically about 25 nt on Affymetrix gene chips). A single nucleotide difference between target and probe can be sufficient to prevent hybridization, and has been used to identify 3000 polymorphisms between two strains of *Sacch. cerevisiae*. These polymorphisms were used as markers to map five uncharacterized loci to within 11–64 kb (Winzeler et al., 1998). Complete characterization of SNPs can be achieved for each base of a sequence of interest by using a set of four oligonucleotides, one oligonucleotide corresponding to the wild-type and the remaining oligonucleotides to the three possible mutations (Lemieux et al., 1998). Oligonucleotide-chip-based mutation analysis is limited by a lack of sensitivity for mutations in regions with high local A/T or G/C content or for small frameshift mutations (Favis et al., 2000). Nevertheless, this method promises to be extremely powerful as a diagnostic tool, as demonstrated by the identification of mutations in a 705 bp region of the *rpoB* gene, which caused rifampicin resistance in 44 clinical isolates of *M. tuberculosis* (Gingeras et al., 1998).

**Industrial applications of microarrays**

Gene expression profiling is being used to determine the effects of antibiotics, agrochemicals and pharmaceutical products on different organisms, and is being used in the search for new antimicrobials. Strategies for drug-target validation and the identification of secondary effects have been described previously (Rosamond & Allsop, 2000). Following determination of the ‘expression signature’ of a wide range of compounds, the prediction of the mode of action of a novel compound becomes possible, simply on the basis of analysis of the transcriptional changes made by the drug. Large biotech companies are already using this approach to obtain cost-effective information, which avoids large-scale mode-of-action studies.

**Microarray data analysis**

Microarrays provide huge amounts of data. This can be an advantage and a disadvantage; the potential of genome-scale information is incredible, but often the results of microarray analyses have been limited to the production of a catalogue of the induction or repression ratios of particular genes. This restricted approach fails to exploit the true value of microarray data. To obtain interesting and reliable hypotheses, and hence results, good mathematical and statistical tools are needed for the intelligent interrogation, or ‘mining’ of microarray data. Consensus must be reached on the level of differential gene expression that is truly significant, and similar approaches must be used by the worldwide community. This will be complicated by a recent analysis of global gene expression in *E. coli*, which concluded that there was ‘little correlation between the “fold difference” and the accuracy of differential gene expression levels’. Thus, the significance of differential gene expression measurements cannot be assessed simply by considering the magnitude of the difference between two experimental conditions (Arfin et al., 2000). For a single microarray experiment involving 5000 measurements, it is predicted that 250 false positives could arise from a Gaussian distribution of data points, emphasizing the importance of experimental replication and careful assessment of statistical significance (Arfin et al., 2000). To assess issues such as the number of assays required, the significance of changes in expression levels or within a cluster analysis will be essential (Zweiger, 1999). Fortunately, bioinformatic tools are being developed at great speed (Ochman et al., 2000). As well as extracting all we can from microarray data, we also need to be able to directly compare experiments within our research community. Since whole-genome transcriptional responses are very complex, all source of noise must be minimized, and good standardization procedures must be applied in
terms of experimental design, the description of results and the format for data storage (Aach et al., 2000). There is a strong argument for the use of universal controls for particular microarrays. For example, the proposed use of genomic DNA as a reference for all gene expression studies performed by members of the *Sal. typhimurium* microarray community should facilitate rapid exchange of meaningful data between laboratories.

A further step towards the prediction of gene function has been made by combining the high-throughput production of data provided by microarrays with a rigorous statistical analysis. Hughes et al. (2000a) have reported a large-scale approach that is intended to avoid problems of biological noise, and to build up a reference database or ‘compendium’ of patterns of gene expression profiles corresponding to 300 different mutations and chemical treatments in *Sacch. cerevisiae*. Two-dimensional hierarchical clustering of the obtained expression profiles identified several large groups of coregulated genes. Mutations in genes having similar known functions gave rise to similar profiles, which clustered together, giving an experimental basis for gene function prediction. This tactic has allowed small but coordinated differential gene expression levels to be observed across many different conditions, and to be related to gene function.

In the excitement of pursuing gene expression profiling for entire organisms, we must not lose sight of the fact that mRNA is only one intermediate between DNA and protein. Post-transcriptional and post-translational controls also play a major role in modulating protein expression. Transcriptional analysis may generate hypotheses, but more traditional molecular biological and proteomic approaches are still required to test these hypotheses.

The utility of microarrays now extends to the study of translational initiation. Kuhn et al. (2001) analysed translational regulation of specific mRNAs in yeast. Polysomal fractionation was used in conjunction with microarrays to study changes in translational initiation during diauxic shift. Although overall mRNA translation decreased, the authors identified one group of mRNA species (representing 610 out of 6275 genes examined) whose level of translational initiation was less affected by the change in carbon source. This group corresponded to the genes upregulated on diauxic shift, emphasizing the importance to the cell of mechanisms that ensure the translation of newly expressed genes.

### Reliability of microarray data

High-level mathematics has taught us that it is important to prevent the introduction of systematic bias when working with large numbers of variables. This concept holds true for microarray data analysis. Errors may be introduced at many points between the production of microarrays and the analysis of hybridization signals. Two critical steps that may strongly affect the results are the isolation of RNA and the generation of fluorescently labelled probes. RNA purification is inherently more difficult for bacterial than mammalian systems. The absence of polyadenylated mRNA means that cDNA labelling must be performed with total RNA, only approximately 3% of which is mRNA. Furthermore, bacterial mRNA is much less stable than eukaryotic mRNA. The half-life of mRNA molecules in *E. coli* can be as short as 30 s (Carpousis et al., 1999). Since differential mRNA instability is an important mechanism in the control of gene expression, great care must be taken to obtain quality RNA that has not been degraded. The rapid stabilization of RNA by addition of chaotropic agents such as guanidinium isothiocyanate is one important tool (Cox, 1968). Alternative commercial products such as RNALater (Ambion) perform a similar function. The initial stabilization of bacterial RNA is critical; otherwise one is in danger of studying cold-shock genes induced during centrifugation of bacteria at 4 °C prior to RNA extraction! The cDNA synthesis step is also critical because the cDNA probe must reflect a representative population of labelled mRNA species. It has recently been shown that the reverse transcription of *E. coli* RNA using a pool of primers specific to the 3’ regions of all mRNA molecules had a significant under-representation of about 30% of mRNAs when compared to the use of random hexamers (Arfin et al., 2000). Another problem that can be associated with labelling is that different fluorescent dyes do not have the same incorporation rate during labelling. This can be controlled by performing ‘dye swap’ experiments (Wei et al., 2001), but we recommend the ‘indirect’ labelling approach to avoid this problem (http://www.microarrays.org).

We have described some of the technical problems commonly encountered with microarrays. It is important to remember that the use of microarrays for gene expression profiling is a recent development, and some aspects of the approach are not completely understood. Therefore, important results obtained with microarrays must be confirmed with other techniques, such as real-time quantitative PCR or Northern blotting, until microarray-based methodologies are completely validated.

### To build or to buy, that is the question

Many laboratories and research centres are considering whether to invest in microarray technology or to obtain pre-printed microarrays from commercial sources. A number of microarrays (membrane-based arrays) and microarrays (glass slides) of interest to microbiologists are already available. Currently, the variety of arrayed microbial genomes on the market is restricted to relatively few organisms (Table 2), though on-going genome sequencing projects will rapidly expand this selection in the near future. However, current pricing is set at a high level (between £500 and £1000 per microarray), reflecting a lack of competition.

A common misconception is that gene expression profiling experiments will only require a small number
also essential. Simple basic errors, such as choice of the
approach, both at the technical and the 
microbiological levels. Appropriate experimental design is 
also essential. Simple basic errors, such as choice of the
wrong media or the stage of the growth phase used to 
compare mutant and wild-type strains, can dramatically 
affect gene expression! Changes in gene expression 
can be very transient. A 10 min exposure of yeast to 0·4 M 
NaCl resulted in the induction of 1300 genes, whereas 
only 172 induced genes were detected after 20 min 
(Posas et al., 2000). Learning to use microarrays often 
takes months rather than weeks, and requires the 
support of an experienced laboratory. Microarray 
experiments should not be entered into lightly!

The future

Genomotyping should revolutionize our ability to 
distinguish bacteria. The combination of other ‘chip 
technologies’ with genomotyping has already produced 
a prototype capable of separating E. coli from blood and 
performing subsequent microarray analysis (Cheng et al., 
1998). Whether or not this will prove applicable to 
the diagnostics market depends upon cost consider-
ations, and whether the technology can be made 
sufficiently robust to perform in the environment of a 
microbiology laboratory. We should remember that 
PCR promised to be a sensitive diagnostic tool, but has 
not led to many validated diagnostic approaches.

A significant area that needs to be investigated is the 
utility of microarrays for analysis of mixed bacterial 
communities. The application of gene expression 
profiling or genomotyping to obtain information about 
individual species within a natural community would 
prove invaluable for microbial ecology and for micro-
bial systematics alike. Assuming that appropriate 
hybridization stringencies are employed, and given a 
sufficient microbial diversity within the population of 
interest, there is no theoretical reason for this approach 
to fail.

Microarrays can also be used to gain clues to gene 
function through looking at knockout mutants, par-
particularly of predicted regulatory genes. This approach 
has already been successfully used by Winzeler et al. 
(1999b) to follow the growth of pools of 500 yeast 
knockout mutants under various environmental con-
ditions. Each mutant was tagged with a unique oligo-
nucleotide sequence (a ‘molecular barcode’) that was 
detected by hybridization to a custom-built microarray 
to determine growth conditions when certain mutants

## Table 2. Commercial ‘pre-printed’ microarrays (currently or soon to be available)

<table>
<thead>
<tr>
<th>Micro-organism</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacillus subtilis</td>
<td>Sigma-Genosys, Eurogentech</td>
</tr>
<tr>
<td>Candida albicans</td>
<td>Eurogentech, Incyte</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>Sigma-Genosys, TaKaRa Biomedical</td>
</tr>
<tr>
<td>Helicobacter pylori</td>
<td>Eurogentech</td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>Eurogentech</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>Corning, Eurogentech, Operon</td>
</tr>
<tr>
<td>Staphylococcus aureus</td>
<td>Incyte</td>
</tr>
<tr>
<td>Streptococcus pneumoniae</td>
<td>Eurogentech</td>
</tr>
</tbody>
</table>

of microarrays. Researchers must expect to use a 
significant number of microarrays to produce reliable 
data. Unfortunately, unlike radioactively labelled 
microarrays, fluorescently labelled microarrays cannot 
be reused. A typical microarray experiment involving 
four time points performed in triplicate will require 12 
microarrays for a single experiment, i.e. approximately 
£6000 at current prices plus the cost of labelling 
consumables. The majority of papers that have used 
microarray technology in the past year have described 
results from single experiments, without replicates or 
any indication of statistical significance. Such data 
require the ‘suspension of disbelief’ by the reader, and 
are unlikely to be acceptable in the future.

The scale of replication required to yield significant data 
could prove to be a significant barrier to the widespread 
application of microarray technology. And it is not yet 
possible for every medium-sized lab to design and print 
its own microarray slides. ‘In-house’ microarray tech-
nology is still expensive (especially at the level of 
consumables) and labour-intensive. The equipment for 
making and analysing microarrays is readily available, 
at a price. But fierce competition is pushing some 
companies to release machines before they are com-
pletely optimized. We would argue that it is worthwhile 
to rely on the robust homemade Stanford technology, 
which is responsible for the majority of microarray 
publications to date (Thompson et al., 2001). The 
significant investment now being made in genomic and 
post-genomic centres throughout Europe should allow 
researchers at all levels to pursue functional genomic 
approaches, either independently or through collabor-
ation, without needing to set-up ‘in-house’ facilities. 
Clearly, financial constraints can be overcome: Oh & 
Liao (2000) successfully used a small ‘subarray’ of 111 
E. coli genes involved in central metabolism to in-
vestigate metabolic flux.

Caveat emptor!

We must emphasize that the use of microarrays for 
molecular microbiology is still not straightforward. The 
generation of reliable data requires an extremely rig-
orous approach, both at the technical and the micro-
biological levels. Appropriate experimental design is 
also essential. Simple basic errors, such as choice of the

1410
were unable to grow. This methodology combined with a massive parallel analysis of mapped mutants (Ross-Macdonald et al., 1999; Spradling et al., 1999) offers a rapid route to determining the function of the FUN genes found in every microbial genome (Hinton, 1997).

The application of microbial gene expression profiling is only limited by our imagination! Bacteria have been used for decades as sensitive biosensors for mutagenicity (Maron & Ames, 1983), and this approach has recently been brought up to date. E. coli has been used to determine the effects of microwave radiation produced by mobile telephones. Macroarray analysis demonstrated that 13 genes were induced by a 2 h exposure to a commercial mobile telephone (A. Morby, personal communication). As we integrate the power of microarray analyses with our particular research interests, more creative applications are bound to arise.

We are moving from the period of genomics towards the post-genomic future and we are entering what is arguably the most exciting period in the history of microbiology. At last we have the potential to ask questions at a relevant scale, that of the whole genome and hence the whole organism. We are optimistic that swift progress will be made as we learn to implement microarray technology more effectively, and we look forward to the time when innovative ideas can be tested extremely quickly.

References


response to infection with the invasive enteric bacteria
Cluster analysis and display of genome-wide expression patterns.
Proc Natl Acad Sci USA 95, 14863–14868.


