ELSEVIER

Journal of Microbiological Methods 47 (2001) 257–272

Journal of Microbiological Methods

www.elsevier.com/locate/jmicmeth

Review article

Applications of DNA microarrays in microbial systems

Rick W. Ye *, Tao Wang, Laura Bedzyk, Kevin M. Croker

E328 / 148B, DuPont Experimental Station, DuPont Central Research and Development, Route 141 and Henry Clay Road, Wilmington, DE 19880, USA

Received 1 June 2001; received in revised form 13 July 2001; accepted 16 July 2001

Abstract

DNA microarray technology allows a parallel analysis of RNA abundance and DNA homology for thousands of genes in a single experiment. Over the past few years, this powerful technology has been used to explore transcriptional profiles and genome differences for a variety of microorganisms, greatly facilitating our understanding of microbial metabolism. With the increasing availability of complete microbial genomes, DNA microarrays are becoming a common tool in many areas of microbial research, including microbial physiology, pathogenesis, epidemiology, ecology, phylogeny, pathway engineering and fermentation optimization.

Keywords: DNA microarrays; Microbial systems; Genome; bacteria

1. Introduction

The large-scale genome sequencing effort and the ability to immobilize thousands of DNA fragments on a surface, such as coated glass slide or membrane, have led to the development of DNA microarray technology. An entire microbial genome can be easily represented in a single array, making it feasible to perform genome-wide analysis (DeRisi et al., 1997). The two common applications of DNA microarray technology in microbiology are the exploration of genome-wide transcriptional profiles and the measurement of the similarities or differences in genetic contents among different microbes. Currently, DNA microarray technology is being used to study many bacterial species ranging from standard laboratory strains and pathogens to environmental isolates. They include *Escherichia coli* (Richmond et al., 1999; Tao et al., 1999), *Bacillus subtilis* (Fawcett et al., 2000; Ye et al., 2000; Yoshida et al., 2001), *Mycobacterium tuberculosis* (Wilson et al., 1999; Behr et al., 1999), *Helicobacter pylori* (Salama et al., 2000; Israel et al., 2001), *Streptococcus pneumoniae* (de Saizieu et al., 2000; Hakenbeck et al., 2001), *Neisseria meningitidis* (Eurogentec, Seraing, Belgium), *Pseudomonas aeruginosa* (Affymetrix, Santa Clara, CA), *Synechocystis* (Suzuki et al., 2001), *Caulobacter crescentus* (Laub et al., 2000), *Shewanella oneidensis* MR-1 (Murray et al., 2001) and *Methylomonas* sp. strain 16a (Ye et al., unpublished data). With an increasing number of microbial arrays being constructed, DNA microarrays are destined to become a standard laboratory tool, promising to

* Corresponding author. Tel.: +1-302-695-1750; fax: +1-302-695-1829.

E-mail address: rick.ye@usa.dupont.com (R.W. Ye).

0167-7012/01/$ - see front matter
PII: S0167-7012(01)00308-6

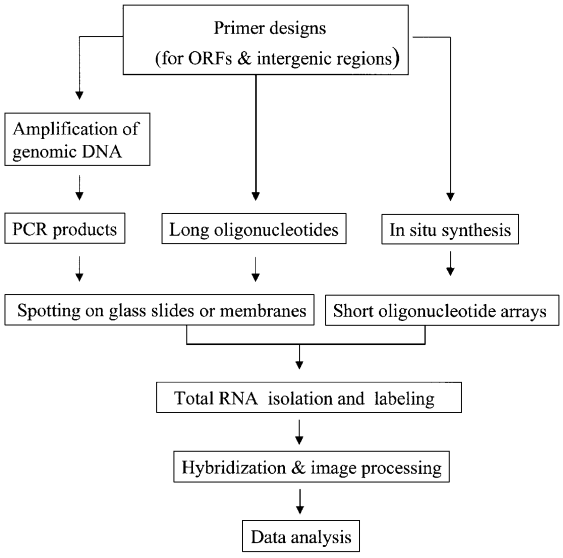

Fig. 1. Construction of DNA microarrays for microorganisms. Not shown here are the construction of oligonucleotide arrays by immobilization of oligonucleotides within a polyacrymide gel and the microelectronic technology developed by Nanogen (Table 1). DNA microarray can also be constructed with random clones for microbes if the complete genome sequence is not available.

greatly enhance and accelerate our investigation of genome-wide functions.

DNA microarrays are basically a miniaturized form of dot blot, but in a high-throughput format. There are two major types of DNA microarrays. One is the oligonucleotide-based array and the other is the PCR product-based array. A DNA microarray experiment consists of array fabrication, probe preparation, hybridization and data analysis (Fig. 1). An example of a DNA microarray experiment is shown in Fig. 2. Although the basic array technology is the same, there are fundamental differences in its application to prokaryotes and eukaryotes. For example, total RNA is usually labeled for a bacterial array experiment, while poly(A) RNA is often used for eukaryotic arrays. In general, bacterial mRNAs tend to have a much shorter half-life, making it difficult to isolate an intact mRNA population. More importantly, there are many specific applications that are unique to microbial systems. This review summarizes the construction and application of DNA microarrays for microorganisms with emphasis on bacteria.

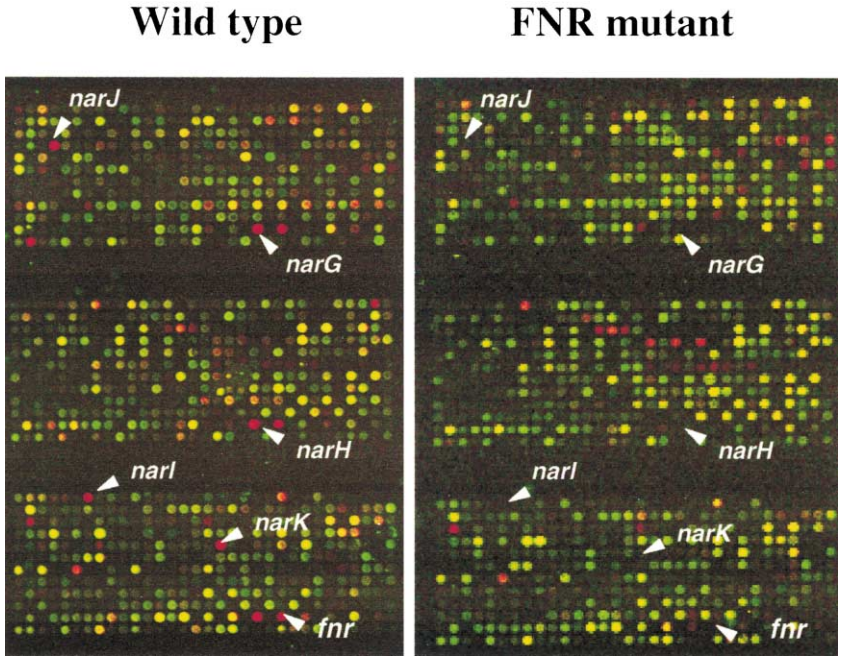

Fig. 2. An example of a DNA microarray experiment. This false-color image shows the induction of *nar* genes and their regulation by FNR protein under anaerobic conditions in *B. subtilis*. The *nar* genes are involved in anaerobic nitrate reduction. Genes induced under anaerobic conditions are shown in red.

2. Construction and use of PCR product-based DNA microarrays

2.1. Primer design

The first step of DNA microarray construction for microbes with known genome sequences is the design of primers to amplify specific regions of interest. In a bacterial genome, there are open reading frames (ORFs) and intergenic regions. The ORFs are the genes that encode specific proteins. The intergenic regions include promoters or regions encoding small RNA molecules which may have regulatory functions (Lease and Belfort, 2000). An ideal array should contain both ORFs and intergenic regions, although most of the current PCR product-based arrays only contain ORFs. Primer design for the whole genome can be carried out with a computer program such as PrimeArray (Raddatz et al., 2001) or Primer 3 (Rozen and Skaletsky, 1996, 1997 http://www-genome.wi.mit.edu/genome_software/other/primer3.html). PrimeArray is specifically designed to compute the oligonucleotide primer pairs for genome-scale gene amplification. The simplest way to design primers is to use the beginning and ending regions of a specific ORF. If an ORF is too long (>3.0 kb), primers can be designed to reduce the size of the PCR product. Since there are repeated regions in a genome, amplification of unique sequences is necessary to avoid cross-hybridization. Before proceeding to the making of all the primers for the whole genome, it is important to test a portion of the primer sets to ensure the primer quality and desirable amplification results.

2.2. PCR amplification

The goal of the whole genome PCR amplification is to achieve the highest success rate and yield in a high-throughput manner. Conditions for PCR amplification are initially optimized with a few 96-well plates. The optimized conditions vary with genomes since they could have different GC contents and secondary structures. After completion of the first round of PCR for all the plates, further optimization or redesign of primers for failed reactions is necessary.

To reduce chromosomal DNA contamination and increase yield, some arrays are constructed with second round PCR products (Ye et al., 2000; Wei et al., 2001). This practice may not be necessary as long as the amount of genomic DNA template remains low. In fact, the second round of amplification could increase the number of reactions that have multiple bands.

2.3. PCR product purification

To remove unincorporated nucleotides and primers, it is recommended to purify the PCR product. PCR purification can be done in either 96- or 384-plate format by ethanol precipitation (DeRisi et al., 1997) or by commercial purification systems (Millipore, Qiagen, Whatman). The 96-well multiscreen filter plates from Millipore have been found to give excellent DNA product recovery with no significant contamination at a relatively low cost (Hegde et al., 2000).

If the goal is to perform high-throughput analysis for gene discovery, purification of PCR products is not necessary. The binding efficiency of PCR products and the quality of the array may be slightly compromised, but there is a significant saving in time and money.

2.4. Spotting

The purified PCR products are spotted onto membranes or coated glass slides (Table 1). Currently,

Table 1
Examples of different types of DNA microarrays that can be used in microbial systems

Surface or matrix	DNA	Reference
Glass	Short oligonucleotides	Affymetrix (http://www.affymetrix.com/)
	Long oligonucleotides	Operon (http://www.operon.com/)
	PCR products	DeRisi et al., 1997, Eurogentec (http://www.eurogentec.com)
Membrane	PCR products	SigmaGenosys (http://www.genosys.com/), Eurogentec
Microelectronics	10–400-bp fragments	Nanogen (http://www.nanogen.com)
Polyacrymide	Oligonucleotides	Proudnikov et al., 1998

Panorama™ membrane macroarrays for several microbial genomes are available commercially (SigmaGenosys and Eurogentec; Table 2). DNA microarrays on coated glass slides are prepared by printing DNA products with high-speed robots. The arraying robots can be custom made (http://www.cmgm.stanford.edu/pbrown/mguide/index.html) or purchased from commercial sources (Table 2).

The common problems associated with glass slides are spot morphology, high background, and what appears to be batch variability. Although there are no perfect slides at the present time, aminosilane-coated slides (Corning, Telechem, Amersham Pharmacia Biotech) and poly-L-lysine-coated slides (DeRisi et al., 1997) are commonly used. The PCR products are resuspended in an appropriate solution before spotting. The two common solutions are high-salt buffer ($3 \times$ SSC) and 50% dimethyl sulfoxide (DMSO). One of the factors in determining which chemistry to use is the type of slides. For CMT-GAPS aminosi-

Table 2
Useful links to various web sites with information on microbial genomes and DNA microarray construction and applications

http://www.cmgm.stanford.edu/pbrown/	The Brown laboratory home page. It includes a complete guide to microarraying for the molecular biologist and other links.
http://www.jgi.doe.gov/tempweb/JGI_microbial/html	Joint Genome Institute. This site contains information on the current microbial genome-sequencing project.
http://www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl	Tigr's Comprehensive Microbial Resource (CMR) home page
christoph.dehio@unibas.ch	PrimeArray software for genome scale primer design (Raddatz et al., 2001)
http://www.genome.wi.mit.edu/genome_software/other/primer3.html	Software Primer3 for primer design
http://www.genosys.com/	SigmaGenosys. It provides membrane arrays and whole genome array sets. Other services include primer synthesis.
http://www.eurogentec.com	Eurogentec. It provides oligonucleotides and whole genome microbial arrays. It also performs custom array experiments.
http://www.microarrays.org/index.html	Protocols and other information
http://www.arrayit.com/index.htm	TeleChem International. It provides slides, software and other services.
http://www.corning.com/	Corning. It provides slides and commercial DNA arrays.
http://www.operon.com/	Operon. It provides 70-mer oligonucleotide array sets and primer synthesis service.
http://www.affymetrix.com/	It provides commercial high-density oligonucleotide arrays and services.
http://www.genemachines.com/	GeneMachines. It provides spotter and other robotics.
http://www.axon.com/	Axon Instruments. Scanner.
http://www.mdyn.com/	Molecular Dynamics. It provides spotter and scanners.
http://www.stat.berkeley.edu/users/terry/zarray/Html/index.html	Terry Speed's microarray data analysis group page
http://www.rana.lbl.gov/EisenResearch.htm	Michael Eisen's software Cluster and TreeView
http://www.lionbioscience.com/	Lion Bioscience. It provides arraySCOUT software and an integrated database platform based on SRS system.
http://www.sigenetics.com/cgi/SiG.cgi/index.smf	Silicon Genetics. It provides GeneSpring Suite for data analysis, presentation and storage.
http://www.genome-www4.stanford.edu/MicroArray/SMD/	Stanford Microarray Database
http://www.genomics.lbl.gov/~ecoreg/	The *E. coli* regulation consortium. A site for primary and secondary data on pathway regulation.
http://www.sghms.ac.uk/depts/medmicro/bugs/	Bacterial microarrays at St. George's Hospital Medical School
http://www.web.wi.mit.edu/young/location/	Methods for genome-wide protein–DNA binding studies

lane-coated glass microscope slides, DMSO has been found to offer many advantages (Hegde et al., 2000). It is hygroscopic and has a low vapor pressure, which allows DNA to be stored for long periods of time without significant evaporation.

After spotting, it is often necessary to check the quality of the slides and spotting results. This can be routinely performed with SYBR green staining (Battaglia et al., 2000) and by hybridizing the array with a Cy-labeled genomic DNA or random 9-mer sequence (http://www.pangloss.com/seidel/Protocols/9-merhyb.html). The SYBR green dye can be used to measure the amount of DNA in each spot and the spotting integrity, while the genomic DNA hybridization is a better indicator of the hybridization background and the quality of the slides.

2.5. RNA isolation

Most of the mRNAs made in bacteria do not have a poly(A) tail and are difficult to separate from the total RNA. An enrichment procedure has been described by Affymetrix (http://www.affymetrix.com). The enrichment is achieved by removing the ribosomal RNA. In addition, a method to isolate *E. coli* mRNA by polyadenylating it in crude cell extracts with *E. coli* poly(A) polymerase I and purifying it by oligo(dT) chromatography has been developed (Wendisch et al., 2001). The most important issue is, however, the mRNA stability. Most bacterial mRNA species only have a half-life in minutes, and the speed required to stabilize the mRNA population becomes crucial. There are numerous procedures to isolate total RNA from bacteria. The traditional procedure is the direct injection of bacterial samples into a hot phenol solution (Selinger et al., 2000). Alternatively, cells can be quickly frozen in liquid nitrogen and mechanically broken before isolation with acid phenol solution. A stop solution containing phenol and ethanol has been described for the isolation of total RNA from *E. coli* (http://www.microarrays.org/index.html). This stop solution could potentially be used for other bacteria. Recently, use of the RNA*later* solution (Ambion and Qiagen) is gaining popularity. The advantages of RNA*later* solution is its rapid stabilization of the mRNA population, allowing the samples to be stored for a long period of time under appropriate conditions prior to RNA isolation. It is especially useful for the collection of samples when immediate isolation of RNA is not possible. Permeability of RNA*later* has not been fully validated for different bacterial species, and its applicability needs to be further tested.

The method to lyse bacterial cells varies, depending on the type of organisms. After stabilizing the RNA population, a lysozyme or mechanical treatment such as a bead beater can be used. A variety of RNA isolation kits are available from different commercial sources (Qiagen, Ambion and others).

2.6. Total RNA labeling

The cDNA probes for array hybridization are synthesized from total RNA by reverse transcriptase. The nucleotides can be labeled with radioisotopes (such as P^{33}) or fluorescent markers. Either random primers or specific primers are used in the reaction. The amount of total bacterial RNA used varies with the organism, stage of growth, type of array and labeling method. Typically, 7–15 μg of total RNA in combination with 6 μg of random hexamers generally yields good labeling efficiency and reasonable signal intensity with Cy5 or Cy3 fluor for arrays on glass slides. The incorporation efficiencies of Cy5- and Cy3-labeled nucleotides are not equal. A two-step labeling procedure using aminoallyl-dUTP is gaining popularity due to the increased labeling efficiency and reduction in dye bias and cost. In this two-step procedure, primary aliphatic amino groups are first incorporated during cDNA synthesis. In the second step, the monofunctional *N*-hydroxylsuccinimide-activated fluorescent dye (Cy3, Cy5 or others) is coupled to cDNA by chemical reaction with the amino functional groups. Since the substrate for the reverse transcriptase is identical for all the samples and is less bulky, the two-step method could yield equivalent molarities of labeled probe with higher efficiency than the one-step labeling procedure. Detailed protocols for indirect labeling can be found on the web (http://www.microarrays.org/). Alternatively, a fluorescent labeling kit can be purchased from Stratagene (http://www.stratagene.com), Clontech (http://www.clontech.com) or Molecular Probes (ARES™ DNA Labeling Kits, http://www.molecularprobes.com). After labeling, it is necessary to remove the unincorporated dyes to

reduce background. This can be achieved by conventional DNA purification methods. To insure the probe quality, the labeling efficiency for Cy3 and Cy5 needs to be calculated. The calculation is based on the extinction coefficient with the following formula: $(O.D._{550} \times \text{dilution factor} \times \text{total volume})/0.15$ for Cy3, or $(O.D._{650} \times \text{dilution factor} \times \text{total volume})/0.25$ for Cy5. The total amount of incorporated dye obtained is in pmol.

It is worthwhile to point out that there are other fluorescent markers (for example, Alexa fluor from Molecular Probes) or labeling methods that are available or being developed. They may prove to be superior to the existing Cy3 or Cy5 technology.

2.7. Genomic DNA labeling

Genomic DNA probes can be used for normalization, slide quality control and comparative genomic studies. Genomic DNA can be labeled by nick translation or by random priming with the Klenow fragment of DNA polymerase (Richmond et al., 1999; Salama et al., 2000). Direct chemical labeling of nucleic acids is also a commonly used method. For example, the Universal Linkage System (ULS) is a technique for binding any marker group or label to DNA and RNA (Kreatech Diagnostics, Amsterdam, The Netherlands, http://www.kreatech.com). For the random priming method, the genomic DNA samples are first sheared by mechanical means such as nebulization or sonication. DNA fragments within 1–3 kb are collected and labeled by either the one- or two-step labeling procedure. A labeling reaction with 0.5–2 μg of genomic DNA often yields enough probes for a single hybridization experiment.

2.8. Hybridization and data acquisition

The amount of probe used for hybridization depends on the array format and labeling method. For arrays on glass slides, a reasonable signal to background ratio can be obtained with probes containing 100–200 pmol of incorporated fluorescent dye. In a typical hybridization reaction, equal amounts of Cy3- and Cy5-labeled probes based on the incorporated dye concentration are combined (Ye et al., 2000). To correct for the difference in labeling efficiency of Cy3 and Cy5, a dye swap procedure is used. In other words, the two samples are labeled with opposite dyes and the resulting probes are hybridized to two different slides.

The overall procedure for a PCR product-based DNA microarray hybridization is basically the same as for a Southern blot except for a few modifications. Before hybridization, most glass slides need to be treated to block or inactivate the non-specific binding sites. The procedure employed depends on the slide type and spotting chemistry. For aminosilane-coated slides, a prehybridization solution containing 1% BSA, 5 × SSC and 0.1% SDS has been found to be effective (Hegde et al., 2000). Usually, the hybridization solution containing the probe is placed onto the array and covered with a cover slip. The glass slide is then placed in a humidified chamber. The temperature of hybridization and washing conditions depend on the GC content of the organism. Bacteria with high GC content require a more stringent washing condition in order to minimize non-specific binding.

After hybridization, the signal intensities of all the spots on a glass slide are captured by scanners (GSI Lumonics, Molecular Dynamics, Genomic Solutions, Axon, and others). For membrane arrays hybridized with P^{33}-labeled probes, a phosphor imaging system (Molecular Dynamics) can be used. Processing of array images involves three steps: spot finding, quantification, and background estimation. These steps are performed with the software provided by the scanner vendors or by other sources (Table 2).

2.9. Data normalization

There are several systematic variables in a DNA microarray experiment that can affect the measurement of mRNA levels, making it difficult to make direct comparisons. Sources of the variations include the inherent errors from sample handling, slide to slide variation, difference in labeling or hybridization efficiency, and variations during image analysis. These differences are not due to the actual changes in gene expression levels. Normalization is a process of minimizing these variations, establishing a common base for comparison. Normalization can be done within the slide to adjust the dye incorporation efficiency, between the two slides for dye swap

experiments and across slides for replicates of the same experiment (Yang et al., 2001). After normalization, the ratio is calculated for each spot on the slide.

2.9.1. Global normalization

This method assumes that the overall mRNA abundance between two populations is similar and that only a small percentage of them differs, or that there is a symmetry in the expression levels of the up- and down-regulated genes. As a result, the overall signals between the test and the control can be statistically normalized. After global normalization, the ratios for the majority of the spots in the two samples should be in the range of 1 to 2. Since this method uses the overall signals generated by the array, global normalization is based on the detectable mRNA population.

2.9.2. Total RNA-based normalization

The mRNA population only makes up a small fraction of the total RNA (about 2–5% in exponential phase) in a bacterial cell. Most of the RNA species are ribosomal RNA (16S 23S and 5S RNA) and tRNAs. Ribosomal RNAs are relatively stable. The overall signal obtained from the array mainly represents the detectable mRNA population. Under many growth conditions such as late stationary phase or the presence of an inhibitor at high concentration, the mRNA population can change dramatically in bacterial cells. Total RNA-based normalization compares the difference in the mRNA abundance of two populations on the basis of the same amount of total RNA, and is more applicable in these situations. It is also applicable to partial arrays that contain only a portion of the microbial genome.

The total RNA-based normalization procedure requires the presence of internal controls. Before labeling, an equal amount of internal control RNA is added to the reaction mixtures that contain the same amount of total RNA. During hybridization, two probes with an equal amount of incorporated dyes are used for the same slide. The two sets of signal intensity data can be normalized with the internal controls before the ratio calculation. Normalization can also be performed after ratio calculation. Both total mRNA- and total RNA-based normalization methods often yield similar results if the compared mRNA populations do not show dramatic differences.

Regardless of the basis for normalization, the normalization factors are often determined by the total fluorescent intensity, mean, linear regression and other more sophisticated statistical methods (Hegde et al., 2000; Dudoit et al., 2000; Yang et al., 2001). Most manufacturers of scanners provide some basic tools for image analysis and basic array normalization. Some free or commercial software is also available (Tigr ArrayViewer from the Institute for Genomic Research, http://www.tigr.org/softlab; GeneSpring from Silicon Genetics, http://www.sigenetics.com/cgi/SiG.cgi/index.smf).

2.9.3. Internal controls

The internal control DNA fragments should have no homology to the DNA sequence of the genome being studied. Control DNA fragments can be selected from plants, yeast and mammalians. In designing internal controls for *B. subtilis*, we selected twelve 1.0-kb DNA fragments from different regions of lambda DNA. Each control is replicated in eight different positions on the array. This results in 96 control spots in the array. A T7 promoter is incorporated into one of the primers for the amplification of these controls so that the PCR products are used directly to make RNA via an in vitro transcription kit (Ambion, Austin, TX). An equal amount of internal control cocktail is spiked into the two labeling reactions that contain an equal amount of total RNA. The final control RNA concentrations range from 25 pg to 5 ng. The signal intensities from the controls cover the same range of signals from the samples. The slope obtained from the scatter plot of the internal controls is used for normalization (Fig. 3).

2.10. Data analysis

To determine if the observed difference between the two means is statistically significant, the *t*-test is often performed. A software that uses statistical methods based on the *t*-test to analyze array data has been described (Long et al., 2001).

To group genes with a similar expression pattern, cluster analysis is often employed. In an unsupervised mode, cluster analysis uses algorithms to arrange genes according to similarity in their expression pattern without applying predefined classes. In

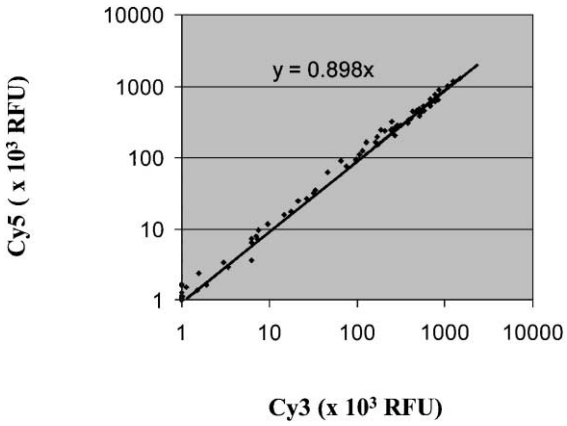

Fig. 3. A scatter plot of signal intensities of internal controls consisting of 1.0-kb fragments from different regions of lambda DNA. The signal intensities are presented in relative fluorescent unit (RFU). The control RNAs were spiked into the two labeling reactions containing the same amount of total RNA from *B. subtilis*. The final concentrations ranged from 25 pg to 5 ng.

the supervised mode, the task is to construct a set of classification rules which assigns predefined classes to given expression profiles (Brazma and Vilo, 2000). The most popular clustering method is hierarchical clustering (Eisen et al., 1998). A variety of types of cluster analysis and other types of processing of large microarray data sets can be performed with the software Cluster (http://www.rana.lbl.gov/ or through http://www.genome-www4.stanford.edu/MicroArray/SMD/). Currently, it includes hierarchical clustering, self-organizing maps (SOMs), *k*-means clustering and principal component analysis. The companion software Tree View allows graphical viewing of the results of clustering and other analyses from Cluster. Clustering and other types of data analyses can also be performed with commercial software such as GeneSpring (Silicon Genetics, http://www.sigenetics.com/cgi/SiG.cgi/index.smf) and arraySCOUT (Lion Bioscience, http://www.lionbioscience.com).

3. Oligonucleotide-based DNA microarrays

Instead of using PCR products, DNA microarrays can be constructed with short oligonucleotides. In the Affymetrix system, oligonucleotides are synthesized in situ on a derivatized glass surface using a combination of photolithography and combinatorial chemistry (Pease et al., 1994). The *E. coli* Genome Array system by Affymetrix uses a protocol for the enrichment and labeling of the non-polyadenylated mRNA of prokaryotes. The mRNA is directly labeled so that it represents the natural distribution of RNA species within the sample. No reverse transcription or amplification steps are involved. On the other hand, the enrichment procedure could also potentially alter the mRNA population. The comparison between direct labeling of RNA after enrichment and the use of reverse transcriptase for total RNA is not available in the literature.

Selinger et al. (2000) have reported the application of a 30-base pair resolution *E. coli* genome array for RNA expression analysis using the Affymetrix system. This array contains on average one 25-mer oligonucleotide probe per 30 base pairs over the entire genome, with one every six bases for the intergenic regions and every 60 bases for 4290 open reading frames (ORFs). Twofold concentration differences can be detected at levels as low as 0.2 messenger RNA (mRNA) copies per cell. The array also permits the investigation of intergenic regions of the genome.

A system using one optimized 70-mer probe per gene has been developed by Operon Technologies (http://www.operon.com). It is similar to a PCR-based microarray except that no amplification is required. The Array-Ready Oligo Set™ for the *M. tuberculosis* genome is currently available. It contains 4321 arrayable 70-mers including 26 controls. Another type of array has also been fabricated by immobilizing oligonucleotides in a polyacrylamide gel (Proudnikov et al., 1998).

Overall, the oligonucleotide-based DNA microarray has many advantages. (i) No amplification is required, and thus, there are no failed amplifications. It is difficult to obtain a high success rate of amplification for microorganisms that contain high GC content or complex DNA structure. Construction of oligonucleotide arrays can be a good option. (ii) There are fewer chances for contamination due to non-specific amplification and mishandling. (iii) There is a reduction in cross-hybridization and an increase in the differentiation of overlapping genes or highly homologous regions. (iv) It is easier to

normalize concentrations of oligonucleotides. (v) High-density oligonucleotide arrays enable high coverage of the genome, and thus, allow a precise mapping of the transcriptional regions and identification of alternative promoters. However, the cost of making long oligonucleotides is high at the present time. There are a limited number of whole microbial genome arrays that are available in the Affymetrix system, and the technology itself is proprietary.

4. DNA microarray database

DNA microarray experiments generate vast amounts of data. The goal of the array database is to allow researchers to retrieve, analyze and visualize the array data. It can also serve as a means to link array data to other information, such as DNA and protein sequences, protein expression profiles and cellular function. In addition, an array database will make it possible to compare gene expression profiles across microbial species. An example of an array database can be found at Stanford University (Sherlock et al., 2001). Stanford Microarray Database (SMD) will not act as a public repository for data, but will instead make all of its source code available to enable other institutions to set up their own databases using SMD's model. The recently established web site EcoReg (http://www.genomics.lbl.gov/~ecoreg/) is designed to be a repository of primary data (analogous to Genbank) for *E. coli* transcriptional control processes. It will be a bioinformatics database project to facilitate improved understanding and modeling of the transcriptional control of *E. coli* gene expression.

There are commercial software packages available to facilitate database construction (Table 2). For example, GeneSpring Suite™ from Silicon Genetics includes a web database and other tools for data sharing. The SCOUT platform developed by Lion Bioscience is based on a sequence retrieval system (SRS). It integrates the array data analysis package, arraySCOUT, to other analysis tools and databases. They include bioSCOUT for automated gene and genome analysis, pathSCOUT for metabolic pathway analysis, and πSCOUT for analysis of protein–protein interaction. This integration also allows a sharing of data on an enterprise-wide level.

5. Validation of DNA microarray data

There are three major sources of errors associated with the application of DNA microarrays: initial construction of the print-ready plates, array experiment, and data analysis. During the construction of the print-ready plates, there are handling issues such as plate transfer, which could lead to cross-contamination or other types of mishandling. During PCR amplification, the presence of multiple bands can lead to false results. Mistakes can also come from the generation of the final gene list, leading to a mismatch between the clone and spot position on the final array. As a result, it is advisable to check the hybridization results against well-characterized genes and internal controls when a new set of spotting plates is made. The common experimental errors can stem from uneven hybridization, inefficient labeling and problems during RNA preparations. During data analysis, it is difficult to have an accurate ratio calculation when the signal intensity is low. When the fold of induction is low, the result can be misinterpreted. All these issues lead to the need for validation of the array experiment. Supporting evidence can be obtained from enzymatic assays, reporter gene systems, and other direct RNA quantification methods such as quantitative real-time PCR (RT-PCR), nuclease protection assay, and primer extension. RT-PCR (Applied Biosystems, http://www.appliedbiosystems.com/) offers a high-throughput advantage and can be an excellent tool to supplement array analysis. As with any RT-PCR reactions, proper controls need to be implemented and care must be taken to avoid DNA contamination in RNA samples.

When the results of a DNA microarray experiment were compared with those obtained from a Northern blot, the sensitivity of the DNA microarray was found to be slightly less than that of a Northern blot analysis (Taniguchi et al., 2001). In most genes, the data obtained by the two methods were consistent. However, in 4 of 46 genes compared, the DNA microarray failed to detect the expression changes that were revealed by the Northern blot. The data demonstrated that DNA microarrays provide quanti-

tative data that are comparable to the Northern blot in general.

6. General applications of DNA microarrays in prokaryotic systems

6.1. Exploring genome-wide transcriptional activity

Cellular function starts with the expression of genes. A DNA microarray can measure the difference in transcriptional activity for every gene or region by comparing their mRNA levels when grown at different developmental stages or under different conditions. For example, 1529 genes were found to be differentially expressed when *E. coli* cells from stationary and log phases were compared (Selinger et al., 2000). Many previously unrecognized growth phase-regulated genes were identified. Global assays of gene expression and protein stability during the *C. crescentus* cell cycle revealed that a surprisingly large fraction of the genome and proteome was affected as cells grow and divide (Laub et al., 2000; Stephens, 2001). A single regulatory factor, the CtrA member of the two-component signal transduction family, was directly or indirectly involved in the control of 26% of the cell cycle-regulated genes. A DNA microarray was also used to explore the genome-wide changes in mRNA abundance in *B. subtilis* when grown under anaerobic conditions (Ye et al., 2000). The transcriptional activities of more than 100 genes were affected by the oxygen-limiting conditions. These include the genes that are involved in energy metabolism, iron uptake, antibiotic production, carbon metabolism and stress response.

Instead of comparing the mRNA levels of different samples, DNA microarrays can detect the presence of an mRNA transcript and can estimate its relative abundance to other mRNA species within the same sample. This type of study addresses the question of whether a gene or pathway is expressed in a given condition and at what level these genes are transcribed. It provides useful hints on the metabolic activity of the cell. Results obtained in *E. coli* have given a semi-quantitative view of each detectable mRNA species (Wei et al., 2001). Interestingly, many observations were consistent with the known *E. coli* physiology. One major issue of using genomic DNA as a control is the difference of the labeling efficiency between genomic DNA and mRNA. It can be difficult to label different mRNA species uniformly since many of them may have a stable secondary structure. Accurate quantification of mRNA in an array format is difficult. Therefore, much of the results obtained by this method are simply a reflection of the transcriptional activity, providing potentially useful information for hypothesis formulation.

Genome-wide transcriptional profiling generates patterns that are specific to a growth condition, developmental stage, or a drug treatment. These signature profiles are potentially a valuable diagnostic tool for a variety of applications such as drug discovery and fermentation process optimization. A software that can analyze, store, retrieve and compare signature profiles will be very useful for these applications.

6.2. Defining a regulon

A regulon is commonly referred to as the set of genes controlled by a regulator. Elucidation of a regulon is often carried out with RNA samples isolated from cells in which the regulatory region is mutated or expressed. For example, to reveal the NtrC regulon of *E. coli*, mRNA levels in a mutant strain that overexpresses NtrC-activated genes [*glnL*(Up)] were compared to those in a strain with an *ntrC* (*glnG*) null allele by using DNA microarrays (Zimmer et al., 2000). In all, approximately 2% of the *E. coli* genome appears to be under NtrC control. Other examples of using DNA microarrays to investigate regulons include the studies of sigma factors and two component regulatory systems in *B. subtilis* (Fawcett et al., 2000; Hecker and Engelmann, 2000; Ye et al., 2000) and competence in *S. pneumoniae* (de Saizieu et al., 2000; Rimini et al., 2000; Peterson et al., 2000).

One potential issue is the pleotrophic effects of mutation or over-expression of genes. This can lead to an induction or a reduction in the transcriptional activity for genes that are not directly associated with a specific regulator. It is especially difficult to sort out those genes that have low expression levels or those that are weakly regulated. Often, it is necessary to combine genomic and traditional methods to define a regulon.

6.3. Delineating operon structure

DNA microarrays can provide information on the coordinated expression of a set of genes located in the same region and with the same transcriptional orientation. If the same expression pattern was found, it could indicate the presence of an operon. If the expression patterns vary under different conditions, it would suggest the presence of multiple promoters for the operon, or whether there are simple multiple operons in the same region. Although PCR product-based DNA microarrays are not the best way to define an operon, this type of information is an added bonus from an array experiment. On the other hand, it is possible to use high-resolution arrays, such as the 30-base pair resolution *E. coli* genome array (Selinger et al., 2000), to obtain reasonable high-resolution information on the transcript starts and stops and operon structure.

6.4. Investigating unknown DNA regions

The DNA sequences of many microbial genomes are now available. Yet, a significant portion of the genome remains to be characterized. Defining the physiological functions of the unknown regions could be problematic, especially when there are no homologous sequences in the database or there are no obvious phenotypes after mutagenesis. DNA microarrays can facilitate the analysis of unknown regions in several ways. (i) Analysis of the regulon of an uncharacterized regulator may lead to the discovery of its physiological role. (ii) Mutations in an unknown pathway can lead to changes in the transcriptional level of associated known genes, providing hints on its potential function. DNA microarrays can easily detect these changes. (iii) Correlation between unknown and known genes in expression patterns when grown under the same conditions or at the same developmental stages could indicate a similar physiology role for these regions; or at least, define the conditions under which a particular unknown gene functions. A combination of DNA microarray with other genomic methods will certainly expedite our effort to characterize the unknown regions of microbial genomes.

6.5. DNA–protein interactions

Many proteins bind to DNA regions in the chromosome to regulate cellular functions or to maintain the genome. The investigation of genome-wide location and function of DNA-binding proteins with DNA microarray has been conducted in yeast (Ren et al., 2000; http://www.web.wi.mit.edu/young/location/). This method combines a modified chromatin immunoprecipitation procedure with a DNA microarray, which contains all the yeast intergenic sequences. Results of the experiment demonstrate that this approach can identify the global set of genes whose expression is controlled directly by transcriptional activators in vivo. DNA microarrays can also be used to study in vitro binding of proteins (Bulyk et al., 2001).

6.6. Comparative genomics and genotyping

Genomic hybridization of a whole genome array detects the presence or absence of similar DNA regions in other microorganisms, allowing genome-wide comparison of their genetic contents. It is an effective way to conduct a comparative genomic study in the absence of complete genome sequences. DNA microarrays have been used to investigate genome differences between *M. tuberculosis*, *M. bovis* and the various Bacille Calmette-Guérin (BCG) daughter strains (Behr et al., 1999) within the species of *H. pylori* and *M. tuberculosis* (Salama et al., 2000; Kato-Maeda et al., 2001) among different isolates of *S. pneumoniae* (Hakenbeck et al., 2001). These studies show that DNA microarrays can facilitate a better understanding of the genetic differences between closely related organisms, providing useful information for the identification of virulence factors, exploration of molecular phylogeny, improvement of diagnostics and development of vaccines.

DNA microarray technology is also an excellent way to identify changes in genetic content of the same strain after long-term adaptation or strain optimization. After adaptation for 2000 generations to a stressful high temperature of 41.5 °C, *E. coli* was examined on a genome-wide scale for duplication/deletion events by using DNA arrays (Riehle et al., 2001). A total of five duplication and deletion events were detected, providing additional evidence

for the idea that gene duplication plays an integral role in adaptation, specifically as a means for gene amplification.

7. Specific applications

7.1. Determination of virulence factors of microbial pathogens

Many genes associated with virulence are regulated by specific conditions. One way to determine the candidate virulence factors is to investigate the genome-wide gene expression profiles under relevant conditions, such as physiological changes during interaction with the host. A second approach would rely on comparative genomics. In a genome comparison study among *H. pylori* strains, a class of candidate virulence genes was identified by their coinheritance with a pathogenicity island (Salama et al., 2000). The whole genome microarray of *H. pylori* was also shown to be an effective method to identify differences in gene content between two *H. pylori* strains that induce distinct pathological outcomes (Israel et al., 2001). It is demonstrated that the ability of *H. pylori* to regulate epithelial cell responses related to inflammation depends on the presence of an intact *cag* pathogenicity island.

7.2. Host responses to pathogens or resident microflora

The host transcriptional profiles during the interaction of *B. pertussis* with a human bronchial epithelial cell line (BEAS-2B) were investigated using high-density DNA microarrays (Belcher et al., 2000). The early transcriptional response to this pathogen is dominated by the altered expression of cytokines, DNA-binding proteins, and NFκB-regulated genes. It was found that *B. pertussis* induces mucin gene transcription by BEAS-2B cells and then counters this defense by using mucin as a binding substrate. This result indicates the host defensive and pathogen counter-defensive strategies. A DNA microarray was also used to identify the host genes that were differentially expressed upon infection by *P. aeruginosa* to the A549 lung pneumocyte cell line (Ichikawa et al., 2000). Differential expression of genes involved in various cellular functions was found, and one of those genes encodes the transcription factor interferon regulatory factor 1. Both experiments with *B. pertussis* and *P. aeruginosa* demonstrate that host genomic transcriptional profiling, in combination with functional assays to evaluate subsequent biological events, provides insight into the complex interaction of host and human pathogens. A comparison of these two data sets and lessons from current microarray analyses of host–pathogen interactions were discussed in detail in a recent review (Diehn and Relman, 2001). In addition, monitoring of the cellular responses to *Listeria monocytogenes* and *Chlamydia pneumoniae* has been reported (Cohen et al., 2000; Coombes and Mahony, 2001).

In addition to the studies of host response to pathogens, DNA microarrays were used to investigate the global intestinal transcriptional responses to the residential colonization of *Bacteroides thetaiotaomicron*, a prominent component of the normal mouse and human intestinal microflora (Hooper et al., 2001). The results reveal that this commensal bacterium modulates the expression of genes involved in several important intestinal functions, including nutrient absorption, mucosal barrier fortification, xenobiotic metabolism, angiogenesis and postnatal intestinal maturation. These findings provide unique perspectives about the essential nature of the interactions between resident microorganisms and their hosts.

7.3. Gene expression profiles of drugs, inhibitors and toxic compounds

Inhibition of a particular cellular process may result in a regulatory feedback mechanism, leading to changes in gene expression patterns. Exploring the gene expression profiles with DNA microarrays may reveal information on the mode of action for drugs, inhibitors or toxic compounds. DNA microarray hybridization experiments have been conducted in *M. tuberculosis* to explore the changes in gene expression induced by the antituberculous drug isoniazid (INH) (Wilson et al., 1999). INH selectively interrupts the synthesis of mycolic acids, which are branched β-hydroxy fatty acids. Microarray experiments showed that isoniazid induced several genes

that encode proteins that are physiologically relevant to the drug's mode of action, including an operonic cluster of five genes encoding type II fatty acid synthase enzymes and *fbpC*, which encodes trehalose dimycolyl transferase. Insights gained from this approach may define new drug targets and suggest new methods for identifying compounds that inhibit those targets.

In addition to the alternation in gene expression patterns related to the drug's mode of action, drugs can induce changes in genes related to stress responses that are linked to the toxic consequences of the drug. The secondary effects of a drug may reveal information on the potential resistance mechanism, which may help design drugs that have less side effects but have high efficacy by reducing the bacterium's ability to neutralize the drug.

Each type of compound often generates a signature pattern of gene expression. A database populated with these signature profiles can serve as a guide to elucidate the potential mode of action as well as side effects of uncharacterized compounds.

7.4. Analyses of microbial evolution and epidemiology

DNA microarrays can be used to explore the variability in genetic content and in gene expression profiles within a natural population of the same or related species and between the ancestor and the descendents. As a result, it provides very rich information on the molecular basis of microbial diversity, evolution and epidemiology. Genomes within the species of *M. tuberculosis* have been compared with a high density oligonucleotide microarray to detect small-scale genomic deletions among 19 clinically and epidemiologically well-characterized isolates (Kato-Maeda et al., 2001). This study reveals that deletions are likely to contain ancestral genes whose functions are no longer essential for the organism's survival, whereas genes that are never deleted constitute the minimal mycobacterial genome. As the amount of genomic deletion increased, the likelihood that the bacteria will cause pulmonary cavitation decreased, suggesting that the accumulation of mutations tends to diminish their pathogenicity.

7.5. DNA microarray as a diagnostic tool

DNA microarrays could potentially be an assay method to address multiple questions for species identification in both clinical and environmental settings. It identifies their phylogenetic status based on unique 16S rRNA sequences and provides information related to the presence of antibiotic markers and pathogenicity regions. The use of a DNA probe array for species identification and detection of rifampin resistance in *M. tuberculosis* has been described (Troesch et al., 1999). Seventy mycobacterial isolates from 27 different species and 15 rifampin-resistant strains were tested. A total of 26 of the 27 species were correctly identified as well as all of the rifampin-resistant mutants.

For field applications, a portable system for microbial sample preparation and oligonucleotide microarray analysis has been reported (Bavykin et al., 2001). This portable system contained three components: (i) a universal silica mini-column for successive DNA and RNA isolation, fractionation, fragmentation, fluorescent labeling, and removal of excess free label and short oligonucleotides; (ii) microarrays of immobilized oligonucleotide probes for 16S rRNA identification; and (iii) a portable battery-powered device for imaging the hybridization of fluorescently labeled RNA fragments on the arrays. Beginning with whole cells, it takes approximately 25 min to obtain labeled DNA and RNA samples and an additional 25 min to hybridize and acquire the microarray image using a stationary image analysis system or the portable imager.

7.6. Pathway engineering and process optimization

Traditional approaches of biocatalysis optimization use random screening, mutagenesis and engineering improvement. While these methods are still very effective, a better understanding of the underlying physiology using genomic tools can accelerate these efforts. Information obtained by the DNA microarrays can help pathway engineering and process optimization in several ways. (i) Regulatory circuitry and coordination of gene expression among different pathways under different growth conditions can be measured by DNA microarray. (ii) The physiological state of the cells during fermentation can be assessed

by the genome-wide transcriptional patterns. (iii) DNA microarray can help identify genes involved in a production process if they are coregulated. (iv) The differences in genetic contents and expression profiles between wild-type and improved strains can be compared. (v) The actual array data can be incorporated into the mathematical models to describe a cellular process. Finally, general applications of DNA microarray technology to understand microbial physiology will continue to generate very large amounts of information that will ultimately benefit the pathway engineering and fermentation optimization effort.

Current research in using array information in pathway engineering and bioprocessing is at its early stage. Arrays containing genes involved central metabolism, key biosyntheses, some regulatory functions and stress response have been used to investigate the metabolic responses to protein overproduction and metabolic fluxes in *E. coli* (Oh and Liao, 2000a,b; Gill et al., 2001). Gene array analysis was also used as a tool to investigate the differences in the expression levels for 30 genes involved in xylose catabolism in the parent, strain B, and the engineered strain, KO11 (Tao et al., 2001). Increased expression of genes involved in xylose catabolism is proposed as the basis for the increase in growth rate and glycolytic flux in ethanologenic KO11.

8. Conclusion

With the availability of complete genome sequences of many microorganisms, the DNA microarray technology has become a very powerful tool to explore global gene expression profiles and to measure genome-wide differences in genetic contents. Since only the abundance of transcripts or presence and absence of DNA regions are measured with a DNA microarray, interpretation of the array data can be difficult in the absence of other supporting evidence. This is especially true when the physiological events are not well studied. In addition, it is not easy to sort out secondary effects caused by mutations, expression of certain genes, and different growth conditions. As a result, the greatest impact of this technology will not be realized until it is combined with other high-throughput genomic methods, biochemistry, genetics and physiology. Analysis of a systematically perturbed metabolic network in yeast clearly demonstrates the power of an integrated approach to build, test, and refine a model of a cellular metabolic pathway (Ideker et al., 2001).

Applications of DNA microarrays for gene expression profiling experiments between two samples appear to be relatively reliable. The array technology, however, cannot give a reasonable estimation of the actual amount of mRNA. The measurement of relative abundance of particular mRNA species within the same sample needs to be further tested and improved.

Currently, both oligonucleotide and PCR product-based arrays are used for the study of bacterial species. Whether one format will prevail in the future will largely depend on robustness, feasibility (cost and availability of technology) and purpose of the experiments. For example, short oligonucleotide arrays may not be suitable for comparative genomic studies for organisms that are not closely related. Additionally, an individual array could be made of both oligonucleotides and PCR products.

The use of DNA microarrays as a tool for phylogenetic studies and strain identification merits attention. For many organisms, the 16S rRNA approach often fails to truly reflect their genetic potential. This gap can be bridged by comparative genomic methods with whole genome arrays in the absence of genome sequences. It is not difficult to envision the future construction of a DNA array that will contain unique sequences of 16S rRNA, 23S rRNA, and many key functional genes for most of the representative bacterial species. This array could be useful in food, medical, environmental, and agricultural applications.

With the increasing applications of DNA microarrays and generation of enormous quantities of data, the construction of a database and the linking of relevant functional information will be the next important phase of technology development. Centralization of genomic data, including DNA sequences and array results, will be very beneficial to the research community. The construction of EcoReg, EcoCyc and EcoSal web sites for *E. coli* is an excellent starting point and could serve as a model for other prokaryotic microorganisms.

References

Battaglia, C., Salani, G., Consolandi, C., Bernardi, L.R., De Bellis, G., 2000. Analysis of DNA microarrays by non-destructive fluorescent staining using SYBR green II. BioTechniques 29, 78–81.

Bavykin, S.G., Akowski, J.P., Zakhariev, V.M., Barsky, V.E., Perov, A.N., Mirzabekov, A.D., 2001. Portable system for microbial sample preparation and oligonucleotide microarray analysis. Appl. Environ. Microbiol. 67, 922–928.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Schoolnik, G.K., Rane, S., Small, P.M., 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284, 1520–1523.

Belcher, C.E., Drenkow, J., Kehoe, B., Gingeras, T.R., McNamara, N., Lemjabbar, H., Basbaum, C., Relman, D.A., 2000. The transcriptional responses of respiratory epithelial cells to Bordetella pertussis reveal host defensive and pathogen counter-defensive strategies. Proc. Natl. Acad. Sci. U.S.A. 97, 13847–13852.

Brazma, A., Vilo, J., 2000. Gene expression data analysis. FEBS Lett. 480, 17–24.

Bulyk, M.L., Huang, X., Choo, Y., Church, G.M., 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc. Natl. Acad. Sci. U. S. A. 98, 7158–7163.

Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinot-Chazel, C., Galiegue, S., Hadibi, E.H., Casellas, P., 2000. Monitoring cellular responses to *Listeria monocytogenes* with oligonucleotide arrays. J. Biol. Chem. 275, 11181–11190.

Coombes, B.K., Mahony, J.B., 2001. cDNA array analysis of altered gene expression in human endothelial cells in response to *Chlamydia pneumoniae* infection. Infect. Immun. 69, 1420–1427.

DeRisi, J.L., Iyer, V., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., Keck, W., Amrein, K.E., Lange, R., 2000. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J. Bacteriol. 182, 4696–4703.

Diehn, M., Relman, D.A., 2001. Comparing functional genomic data sets: lessons from DNA microarray analyses of host–pathogen interactions. Curr. Opin. Microbiol. 4, 95–101.

Dudoit, S., Yang, Y.H., Callow, M.J., Speed, T.P., 2000. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. http://www.stat.berkeley.edu/users/terry/zarray/Html/papersindex.html.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 95, 14863–14868.

Fawcett, P., Eichenberger, P., Losick, R., Youngman, P., 2000. The transcriptional profile of early to middle sporulation in *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. 97, 8063–8068.

Gill, R.T., DeLisa, M.P., Valdes, J.J., Bentley, W.E., 2001. Genomic analysis of high-cell density recombinant *Escherichia coli* fermentation and "cell conditioning" for improved recombinant protein yield. Biotechnol. Bioeng. 72, 85–95.

Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W., de Saizieu, A., 2001. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. Infect. Immun. 69, 2477–2486.

Hecker, M., Engelmann, S., 2000. Proteomics, DNA arrays and the analysis of still unknown regulons and unknown proteins of *Bacillus subtilis* and pathogenic gram-positive bacteria. Int. J. Med. Microbiol. 290, 123–134.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N., Quackenbush, J., 2000. A concise guide to cDNA microarray analysis. BioTechniques 29, 548–562.

Hooper, L.V., Wong, M.H., Thelin, A., Hansson, L., Falk, P.G., Gordon, J.I., 2001. Molecular analysis of commensal host–microbial relationships in the intestine. Science 291, 881–884.

Ichikawa, J.K., Norris, A., Bangera, M.G., Geiss, G.K., van't Wout, A.B., Bumgarner, R.E., Lory, S., 2000. Interaction of *Pseudomonas aeruginosa* with epithelial cells: identification of differentially regulated genes by expression microarray analysis of human cDNAs. Proc. Natl. Acad. Sci. U. S. A. 97, 9659–9664.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., Hood, L., 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.

Israel, D.A., Salama, N., Arnold, C.N., Moss, S.F., Ando, T., Wirth, H.P., Tham, K.T., Camorlinga, M., Blaser, M.J., Falkow, S., Peek Jr., R.M., 2001. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. J. Clin. Invest. 107, 611–620.

Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N., Small, P.M., 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. 11, 547–554.

Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M., Shapiro, L., 2000. Global analysis of the genetic network controlling a bacterial cell cycle. Science 290, 2144–2148.

Lease, R.A., Belfort, M., 2000. A trans-acting RNA as a control switch in *Escherichia coli*: DsrA modulates function by forming alternative structures. Proc. Natl. Acad. Sci. U. S. A. 97, 9919–9924.

Long, A.D., Mangalam, H.J., Chan, B.Y., Tolleri, L., Hatfield, G.W., Baldi, P., 2001. Gene expression profiling in *Escherichia coli* K12: improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. J. Biol. Chem. 276, 19937–19944.

Murray, A.E., Lies, D., Li, G., Nealson, K., Zhou, J., Tiedje, J.M., 2001. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. Proc. Natl. Acad. Sci. U.S.A. 98, 9853–9858.

Oh, M.K., Liao, J.C., 2000a. DNA microarray detection of metabolic responses to protein overproduction in *Escherichia coli*. Metab. Eng. 2, 201–209.

Oh, M.K., Liao, J.C., 2000b. Gene expression profiling by DNA microarrays and metabolic fluxes in *Escherichia coli*. Biotechnol. Prog. 16, 278–286.

Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., Fodor, S.P., 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc. Natl. Acad. Sci. U. S. A. 91, 5022–5026.

Peterson, S., Cline, R.T., Tettelin, H., Sharov, V., Morrison, D.A., 2000. Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. J. Bacteriol. 182, 6192–6202.

Proudnikov, D., Timofeev, E., Mirzabekov, A., 1998. Immobilization of DNA in polyacrylamide gel for the manufacture of DNA and DNA-oligonucleotide microchips. Anal. Biochem. 259, 34–41.

Raddatz, G., Dehio, M., Meyer, T.F., Dehio, C., 2001. PrimeArray: genome-scale primer design for DNA microarray construction. Bioinformatics 17, 98–99.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., Young, R.A., 2000. Genome-wide location and function of DNA-binding proteins. Science 290, 2306–2309.

Richmond, C.S., Glasner, J.D., Mau, R., Jin, H., Blattner, F.R., 1999. Genome-wide expression profiling in *Escherichia coli* K-12. Nucleic Acids Res. 27, 3821–3835.

Riehle, M.M., Bennett, A.F., Long, A.D., 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 98, 525–530.

Rimini, R., Jansson, B., Feger, G., Roberts, T.C., de Francesco, M., Gozzi, A., Faggioni, F., Domenici, E., Wallace, D.M., Frandsen, N., Polissi, A., 2000. Global analysis of transcription kinetics during competence development in *Streptococcus pneumoniae* using high-density DNA arrays. Mol. Microbiol. 36, 1279–1292.

Salama, N., Guillemin, K., McDaniel, T.K., Sherlock, G., Tompkins, L., Falkow, S., 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. Proc. Natl. Acad. Sci. U. S. A. 97, 14668–14673.

Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., Church, G.M., 2000. RNA expression analysis using a 30-base pair resolution *Escherichia coli* genome array. Nat. Biotechnol. 18, 1262–1268.

Sherlock, G., Hernandez-Boussard, T., Kasarskis, A., Binkley, G., Matese, J.C., Dwight, S.S., Kaloper, M., Weng, S., Jin, H., Ball, C.A., Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., Cherry, J.M., 2001. The Stanford Microarray Database. Nucleic Acids Res. 29, 152–155.

Stephens, C., 2001. Bacterial cell cycle: seeing the big picture with microarrays. Curr. Biol. 11, R222–R225.

Suzuki, I., Kanesaki, Y., Mikami, K., Kanehisa, M., Murata, N., 2001. Cold-regulated genes under control of the cold sensor Hik33 in *Synechocystis*. Mol. Microbiol. 40, 235–244.

Taniguchi, M., Miura, K., Iwao, H., Yamanaka, S., 2001. Quantitative assessment of DNA microarrays—comparison with Northern blot analyses. Genomics 71, 34–39.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., Conway, T., 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. 181, 6425–6440.

Tao, H., Gonzalez, R., Martinez, A., Rodriguez, M., Ingram, L.O., Preston, J.F., Shanmugam, K.T., 2001. Engineering a homoethanol pathway in *Escherichia coli*: increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation. J. Bacteriol. 183, 2979–2988.

Troesch, A., Nguyen, H., Miyada, C.G., Desvarenne, S., Gingeras, T.R., Kaplan, P.M., Cros, P., Mabilat, C., 1999. *Mycobacterium* species identification and rifampin resistance testing with high-density DNA probe arrays. J. Clin. Microbiol. 37, 49–55.

Wei, Y., Lee, J.M., Richmond, C., Blattner, F.R., Rafalski, J.A., LaRossa, R.A., 2001. High-density microarray-mediated gene expression profiling of *Escherichia coli*. J. Bacteriol. 183, 545–556.

Wendisch, V.F., Zimmer, D.P., Khodursky, A., Peter, B., Cozzarelli, N., Kustu, S., 2001. Isolation of *Escherichia coli* mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. Anal. Biochem. 290, 205–213.

Wilson, M., DeRisi, J., Kristensen, H.H., Imboden, P., Rane, S., Brown, P.O., Schoolnik, G.K., 1999. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc. Natl. Acad. Sci. U. S. A. 96, 12833–12838.

Yang, Y.H., Dudoit, S., Luu, P., Speed, T.P., 2001. Normalization for cDNA Microarray Data. pp. 1–12(http://www.stat.berkeley.edu/users/terry/zarray/Html/papersindex.html).

Ye, R.W., Tao, W., Bedzyk, L., Young, T., Chen, M., Li, L., 2000. Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. J. Bacteriol. 182, 4458–4465.

Yoshida, K., Kobayashi, K., Miwa, Y., Kang, C.M., Matsunaga, M., Yamaguchi, H., Tojo, S., Yamamoto, M., Nishi, R., Ogasawara, N., Nakayama, T., Fujita, Y., 2001. Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in *Bacillus subtilis*. Nucleic Acids Res. 29, 683–692.

Zimmer, D.P., Soupene, E., Lee, H.L., Wendisch, V.F., Khodursky, A.B., Peter, B.J., Bender, R.A., Kustu, S., 2000. Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. Proc. Natl. Acad. Sci. U. S. A. 97, 14674–14679.