

Bacterial genomics: the use of DNA microarrays and bacterial artificial chromosomes

Kristen D. Ball, J.T. Trevors*

Department of Environmental Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Received 20 November 2000; received in revised form 24 November 2001; accepted 26 November 2001

Abstract

Immense amounts of genetic information are contained within microbial genomes. As the number of completely sequenced microbial genomes is increasing, functional and comparative genomic techniques will be employed for sequence analysis and gene characterization. Sequence comparison and expression profiling by DNA microarrays can determine phylogenetic relationships and identify genes while bacterial artificial chromosomes (BACs) allow the study of entire biochemical pathways and permit the expression of bacterial genes in a foreign host. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Bacteria; Bacterial artificial chromosome (BAC); Genomics; Microarrays; Expression; Microorganism

1. Introduction

The term “genome” was introduced in 1920 to describe the sum of all genes on chromosomes. Since then, advances in DNA research launched the field of genomics. Genomics now encompasses the simultaneous structural and functional analysis of genes. The development of rapid and efficient DNA sequencing technology was a milestone in genomics. In 1995, *Haemophilus influenzae* became the first free-living microorganism to be sequenced (Fleischmann et al., 1995).

As sequence information on numerous organisms, both prokaryotic and eukaryotic, accumulates, the emphasis of study is shifting away from structural ge-

nomics to comparative and functional genomics, concerned with the analysis and application of whole-genome sequence information. Functional genomics attempts to assign a role to every gene within a genome and to examine the function and expression pattern of every gene in relation to all the other genes in the genome. Investigation of the relationships between genomes is the focus of comparative genomics. DNA microarrays facilitate the study of thousands of genes simultaneously by hybridization of mRNA to a high-density array of immobilized probe sequences. The arrays have been used primarily in biomedical research and clinical diagnostics (Osin et al., 1998; Schena et al., 1998).

A second useful tool in genomics research is the bacterial artificial chromosome (BAC). BACs are specialized cloning systems derived from the *Escherichia coli* F plasmid. Originally developed for chromosome mapping and analysis of the human genome, BACs can maintain DNA inserts exceeding 300 kilobase pairs

* Corresponding author. Tel.: +1-519-824-4120x3367; fax: +1-519-837-0442.

E-mail address: jtrevors@uoguelph.ca (J.T. Trevors).

(Shizuya et al., 1992), and allow expression of prokaryotic genes in a heterologous bacterial host (Rondon et al., 1999).

Microbial genomes, although simpler compared to eukaryotic organisms, are of great scientific, medical and economic importance. In bacterial genomes, there are genes encoding unique enzymes capable of catalyzing reactions under extreme conditions, as well as enzymatic pathways for degradation of chemical contaminants. Furthermore, the relative ease with which entire bacterial genomes can be sequenced assists in their functional analysis. Information obtained from gene function and expression in microbial cells can be used in the comparative analysis of more complex genomes. For example, a number of GTP-binding proteins, found in bacteria such as *E. coli*, *Neisseria gonorrhoeae* and *Bacillus subtilis*, have no known function yet, are encoded by a vital gene. Comparison of the sequence information of these microorganisms to the cumulating information in genomic databases has uncovered orthologous sequences in the genomes of diverse organisms, including eukaryotes. This information has been used to establish phylogenetic lineages and to identify genes that may be essential to life (Mittenhuber, 2001). In this manuscript, we examine the need for an integrative approach in the study of bacterial genomes. This will be accomplished by assessing the science and applications of bacterial genomics, and both the individual and combined roles of DNA microarrays and BAC technology. In addition, bioinformatics—the union of biology and computing—is needed to develop databases and data analysis programs to accelerate and enhance scientific research. Vast DNA sequence databases enable genome comparisons between completely different organisms. Likewise, functional genomic studies make significant contributions to the numerous databases that link gene structure to protein function. Although bioinformatics will not be emphasized in this exploration of bacterial genomics, it is imperative that computational analysis is necessary for the research and applications examined in this paper.

2. Bacterial genomics

Bacteria are ideal for genomic studies as the information in their small genomes (Trevors, 1996) is

relatively easy to access. One milestone in genomics occurred in 1995 with the sequencing of the first free-living organism, *H. influenza* (Fleischmann et al., 1995). A growing list of microbial genomes that have been sequenced can be found in the TIGR Microbial Database (<http://www.tigr.org/tdb/mdb/mdbcomplete.html>). For example, the pathogenic bacterium *N. meningitidis* was recently sequenced to identify potential vaccines against meningitis and septicemia (Tetelin et al., 2000; Pizza et al., 2000). Analysis of the sequence information uncovered 570 ORFs thought to encode surface proteins. Over 60% of these proteins were successfully expressed in *E. coli*, purified and tested as putative vaccine candidates in mice. Immunological analysis of the murine immune sera and comparative genomic analysis of pathogenic *N. meningitidis* strains showed that seven of the proteins were conserved among strains and stimulated effective antibody production. This genome-derived information is of immense value in the development of a potential vaccine against meningococcal disease. The industrial significance of the hydrolysis of complex biopolymers at high temperatures prompted genome sequencing of the thermophile, *Thermotoga maritima* (Nelson et al., 1999). Information derived from the completed genome sequence of the radiation-resistant *Deinococcus radiodurans* has been used to construct recombinant microorganisms with potential uses in radioactive waste management (Brim et al., 2000). The *Xylella fastidiosa* sequencing project has produced a wealth of information on a previously uncharacterized plant pathogen including the molecular basis for a number of metabolic functions and host–cell interactions (Simpson et al., 2000). Evidence for horizontal gene transfer between distantly related microbial species was uncovered by comparative genomic analysis of the thermophilic eubacterium *Thermotoga maritima* and Archaea (Nelson et al., 1999). This has led to a new understanding of bacterial evolution that includes not only diversification and adaptation through consecutive, single-gene mutations but transfer and loss of gene clusters (Ochman and Moran, 2001).

The National Centre for Biotechnology Information contains data on Clusters of Orthologous Groups (COGs) of proteins encoded in the genomes of diverse microorganisms. Each COG represents a functional pathway or system, such as ion transport or DNA re-

pair, conserved among at least three distant relatives. Comparative sequence analysis of the proteins encoded in fully sequenced genomes is used to discover COGs. Typically, 60–80% of the proteins from the genomes of entirely sequenced microorganisms belong to COGs (Tatusov et al., 2001). An exception is *Aeropyrum pernix* in which only 43% of proteins are in COGs. A COG database (<http://www.ncbi.nlm.nih.gov/COG/>) has been established to assist in phylogenetic and evolutionary classification, as well as contributing to the functional characterization of sequenced genomes (Galperin and Koonin, 1999).

3. Genomic technologies

Comparative and functional genomic data can provide fundamental scientific knowledge with applications in medicine, industry, agriculture and environmental biomonitoring (Beattie, 1997; Kotra et al., 2000). These approaches depend on bioinformatics and methods. The growing use of technologies, such as DNA microarrays and BACs, in the field of bacterial genomics, has immense potential with respect to beneficial applications.

4. DNA microarrays

DNA microarrays serve as platforms for genome-wide hybridization experiments for identifying DNA sequences, comparing different genomes, and monitoring gene expression (Lee and Lee, 2000; van Hal et al., 2000). DNA microarrays consist of thousands of unique DNA sequences each attached at a known location to a small, solid surface. Complementary, labelled mRNA or DNA can bind to the fixed sequences, to produce a pattern indicative of nucleic acid sequences that can be qualitatively and quantitatively analysed by a computer.

In this paper, the probe refers to the generally known DNA sequence that is used to identify the unknown 'target' sequence. This is the nomenclature accepted by Nature (Phimister, 1999). In a standard array, the probe consists of oligonucleotides or cloned genes that are attached at a known location to a solid support, such as glass or nylon. The targets are fluorescent or radioactive-labelled total mRNA, DNA

derived from mRNA by RT-PCR, or genomic DNA extracted from microorganisms in a given sample. The targets are hybridized to complementary probes of the array and are identified and quantified by the intensity of the resulting signal.

An oligonucleotide chip describes a specific type of array, originally produced by Affymetrix and marketed as GeneChip®. This technology involves the in situ synthesis of oligonucleotide on a glass array, using a process known as photolithography (Fodor et al., 1993). Using sophisticated, commercially available instruments, up to 300,000 distinct oligonucleotides, can be printed on a 1.28-cm² glass array (Lipshutz et al., 1999). Oligonucleotide chips have been used to provide rapid diagnosis of clinical samples and screening for mutations in cancer-related genes (Favis et al., 2000; Moch et al., 1999; Kononen et al., 1998).

The entire genome of the yeast *Saccharomyces cerevisiae* was mechanically microspotted on 110 glass microscope slides with an average density of 6400 different open reading frames per slide (DeRisi et al., 1997). This array uses cDNA, cloned DNA or DNA fragments obtained from other sources as the immobilized probes. Using a robotics system, the pre-synthesized DNA is bound to the array by chemical methods. DNA microarrays have been similarly produced using ink-jet technology analogous to that currently used in many printers (Castellino, 1997). A detailed description of how to make a DNA microarray can be found at the following website (<http://cmgm.stanford.edu/pbrown/>).

Gene expression in cells under different growth conditions or different environmental conditions can be monitored by hybridization of labelled mRNA to microarrays consisting of a large number of genes of a given organism. In a standard expression study, carbocyanine-labelled nucleotides Cy3-dCTP and Cy5-dCTP (Amersham Pharmacia Biotech, Piscataway, NJ, USA), which fluoresce red and green, respectively, are used to label mRNA extracted from two different samples. The fluorescent labels are incorporated into the target molecules by RT-PCR. Following hybridization, the intensity of red or green fluorescence at every position on the array is read by dual laser sources and converted into a signal. A computer displays the electronic symbol as a false-colour image in which a red or green spot corresponds to an ex-

pression of a gene in samples 1 or 2, respectively, and a yellow-orange spot indicates that a gene was expressed at similar levels in both samples. The intensity of a given spot is proportional to the level of gene expression. Richmond et al. (1999) used this method to produce a high-density array of all 4290 predicted *E. coli* open reading frames (ORFs) to compare the gene expression profiles of *E. coli* before and after heat shock treatment. They observed the expression of over 100 genes was affected by heat-shock. One third of the genes identified were previously uncharacterized, thus illustrating the importance of functional genomic analysis.

Although the probe sequence is known in a standard microarray expression study, useful information can be obtained from a shot-gun array of random, unsequenced DNA clones (Hayward et al., 2000). This approach allows accumulation of biological information on microorganisms without the need for complete genome sequencing. Using an array constructed from 3648 arbitrarily selected clones from a *Plasmodium falciparum* library, genes specific to different developmental stages of the malaria parasite were identified by expression profiling. As the proteins encoded by these genes are potential drug targets, this study provided information from which the development of anti-malaria drugs and vaccines may be possible (Hayward et al., 2000). These experiments reinforce the ability of genomics to bypass information gaps to provide significant information on organisms that have been traditionally difficult to study.

Arrays can be used to screen for genes involved in specific functions such as specific catabolic pathways, toxin production, nitrogen fixation, antibiotic production or cryoresistance (Tanghe et al., 2000). This can be accomplished by comparative array experiments, probing genomic libraries, or expression-profiling experiments using target mRNA extracted from a sample of mixed bacterial populations such as soil, sediment, food and clinical samples. Such studies underscore the potential importance of DNA microarrays in microbial ecology. DNA microarrays can be used for differential sequencing. Genome variation both within and between species, as well as the immense number of different organisms (many unculturable) on Earth, makes sequencing of all genomes virtually impossible. To overcome this problem, known probes from a completely sequenced microorganism can be hybridized to

an array of target DNA from a library of an unsequenced, but related microorganism. Only the clones represented by non-fluorescent spots on the array would have to be sequenced, as the rest of the DNA sequence would be inferred from homology.

DNA microarrays are allowing scientists to ask broader questions addressing the simultaneous interaction of all the genes in a cell under varying conditions. Aside from deducing functional characteristics and identification of a number of genetic elements within microbial genomes, comparative and functional genomic experiments using DNA microarrays have generated vast amounts of data, thereby accelerating the formation of new hypotheses to direct future research. There are numerous microbiological applications of such experiments. Oligonucleotide arrays have been used to detect pathogenic organisms in clinical samples and cDNA arrays have been used to monitor host response to bacterial infection (Manger and Relman, 2000). The transcriptional response of human microvascular endothelial cells (HMEC-1) to infection with *Chlamydia pneumoniae*, a respiratory pathogen also associated with chronic arterial inflammation, was observed through hybridization of cDNA from infected cells to a low-density human cDNA array (Coombes and Mahony, 2001). The expression profile revealed the upregulation of 20 genes involved in inflammation, thereby strengthening the association of *C. pneumoniae* with atherosclerosis. To understand alterations in host gene function due to infection by *Bordetella pertussis*, the causative agent of whooping cough in humans, human bronchial epithelial cells were infected with *B. pertussis* (Belcher et al., 2000). Labelled cRNA from both infected and uninfected cells was hybridized to an Affymetrix HU6800 GeneChip microarray. The host response to *B. pertussis* infection primarily involved the upregulation of a number of genes involved in inflammation, not previously associated with *B. pertussis* infection. These studies provide insight into how the host cell responds to infection and suggest possible mechanisms of pathogenesis, while eliminating the need for culturing pathogens outside of the host system. Consequently, the application of this approach has the potential to understand the alterations of the host gene functions due to the presence of poorly understood human diseases and a number of emerging infectious diseases.

5. Bacterial artificial chromosomes (BACs)

In the early 1990s, Shizuya et al. (1992) at the California Institute of Technology improved the human genome sequencing initiative with an improved large insert cloning vector for eukaryotic DNA sequences. The construction of large insert DNA libraries improved the physical mapping of the human genomes prior to the shot-gun sequencing approach adopted in 1994. These efforts led to the development of bacterial artificial chromosomes (BACs), cloning vectors capable of sustaining DNA inserts greater than 300 kb (Shizuya et al., 1992). BACs were an improvement over other commonly used cloning vectors, specifically yeast artificial chromosomes (YACs) and cosmids. BACs are more easily manipulated by conventional molecular biology techniques and the integrity of the recombinant DNA sequences is readily maintained (Shizuya et al., 1992). Today, BACs are the cloning vector of choice to generate genomic DNA libraries for genomic mapping and sequencing, complementation and hybridization experiments.

A BAC is a modified *E. coli* F-factor that regulates its own replication at a copy number of one or two per cell. BACs also have features typical of most cloning vectors, including a multiple-cloning site and rapid clone selection mechanisms such as antibiotic resistance or lacZ inactivation (Kim et al., 1996). A complete map of a standard BAC can be found at <http://www.brunel.ac.uk/depts/bl/project/genome/moltec/library/pbac.htm>. Transformation of BACs into recombination-defective (*recA*⁻) *E. coli* strains is critical to maintain the integrity of the cloned DNA. However, modification of BACs by specific recombination in recombination-competent strains is useful in minimizing redundancy in libraries (Hill et al., 2000). The construction of a BAC library is described at http://www.tree.caltech.edu/protocols/BAC_lib_construction.html.

Traditional applications of BACs have focused mostly on eukaryotic genomes. These applications include DNA fingerprinting, fluorescent in situ hybridization (FISH), the development of BAC-end sequencing databases and of sub-libraries containing smaller inserts for sequencing and mapping (Kim et al., 1996). BACs have assisted prokaryotic, as well as eukaryotic, whole-genome sequencing efforts by reducing the number of clones required to cover the

genome. A smaller number of clones reduces the time and effort involved in library screening and mapping. For example, an average bacterial genome, such as that of *E. coli*, is approximately 4.5 Mbp (Blattner et al., 1997). Ideally, only 30 clones would be required for a minimal overlap library using a conservative average insert size of 150 kb. A minimal overlap library covering 97% of the 4.4 Mbp *Mycobacterium tuberculosis* genome was possible with only 68 BAC clones (Brosch et al., 1998). The large insert size of BACs allows entire biochemical pathways and other gene clusters to be cloned into a single vector, thereby permitting the study of several genes at once. Furthermore, BAC clones allow the study of certain prokaryotes, such as many archaea, whose genomes were previously unclonable due to their instability in high copy number plasmids (Kim et al., 1996). BACs have also provided a suitable system for the characterization of microorganisms that cannot be isolated or sustained in pure cultures (Rondon et al., 2000). For example, DNA extracted directly from a planktonic marine sample was used to create a BAC library (Beja et al., 2000). Subsequent analysis of BAC clones revealed that a significant amount of DNA from unculturable yet abundant marine microorganisms. The applications of BACs are numerous. First, there is the obvious application of whole-genome sequencing and mapping. BAC libraries can be used to construct DNA microarrays for the functional and comparative genomics experiments described in the previous section. This demonstrates the need for technological integration in genomic studies.

BAC libraries created from DNA directly extracted from a mixed-population sample can be used in the study of metagenomics. The term metagenome describes the sum of the diverse genomes within a mixed population, such as the genetic material extracted from an environmental sample such as soil, water or sediment (Rondon et al., 2000). BAC libraries of DNA extracted directly from environmental samples allow the phylogenetic and genomic analysis of the entire microbial populations. This is useful when applied to environments such as soil in which the majority of microorganisms are viable but non-culturable (VBNC). Rondon et al. (2000) showed that BAC clones directly isolated from agricultural soil expressed various biological functions in *E. coli*, including bactericidal activity. These libraries have the potential to revolu-

tionize phylogenetics, promote the discovery of novel genes and to assist in the hunt for new and useful natural products, including antibiotics.

There is accumulating evidence that recombinant prokaryotic genes within BACs can be expressed in heterologous bacterial hosts. BAC-encoded genes from the Gram-positive bacterium *B. cereus* have been expressed in the Gram-negative *E. coli* (Rondon et al., 1999). This genomic technique allowed rapid identification of functional loci in *B. cereus*, a microorganism not easily studied by traditional molecular techniques. Sosio et al. (2000) created BACs harbouring actinomycete genomic clusters that could replicate autonomously in *E. coli* host or stably integrate into a *Streptomyces* genome by site-specific recombination. The *E. coli*–*Streptomyces* artificial chromosomes (ESAC) were made using the previously described approach or by homologous replication of overlapping clones in *E. coli*. ESACs will permit the analysis of previously genetically uncharacterizable actinomycetes. Optimization of expression of integrated genes can potentially lead to the synthesis of production hosts for antibiotic and other industrially and medically important actinomycete metabolites (Sosio et al., 2000).

6. Discovery of novel antibiotics

One of the most promising areas of genomic research is the search for novel antibiotics. There is an increasing need to develop antibiotics against multi-drug-resistant pathogenic bacteria. Currently, there are limited known vital cell activities, each of which is targeted by a single class of antibiotics. Furthermore, occurrences of resistant microbial strains are typically documented within a few years of prescribed use of a given antibiotic. Using an integrative approach that includes the implementation of both BACs and DNA microarrays, the potential exists to find novel antibiotic targets and to generate new antimicrobial agents.

Antibiotics must be selective for pathogens while not adversely affecting humans. An antibiotic target may consist of a conserved enzyme of a pathway that is essential to microbial survival under conditions of infection. Possible targets include cell division, secretion and various metabolic pathways (Moir et al., 1999). Comparative genomic approaches using both

sequence databases and DNA microarrays can be used to find such targets. To accomplish this, BAC libraries can be constructed for numerous pathogenic microorganisms. Once sequenced, microarrays of all ORFs of a given microorganism can be fabricated, to find targets through comparative screening of numerous pathogens. Comparative genomics approaches and expression profiling may also serve to identify vital or virulence factors unique to a specific species or group of microorganisms.

While comparisons between microorganisms may reveal apparently well-suited targets for antibiotics, screening of eukaryotic genomes against those of bacteria is crucial for antibiotic selectivity without toxicity (Loferer et al., 2000). Based on the findings of the comparative studies, a microarray of all potential targets can be constructed and used to screen for homology in eukaryotic organisms to avoid toxicity. However, sequence homology does not necessarily correspond to functionality. This is demonstrated by the fact that many antibiotics currently in use act on targets whose nucleotide sequences are distinctly similar to eukaryotic sequences. Thus, expression experiments using the antibiotic target array may be in order to determine the function of each array component.

From the information obtained through the experiments above, potential protein targets can be purified and characterized. Assays can be developed to screen for the inhibitors of the target protein and to monitor their potency, both in vitro and in vivo. The details of the experimental procedure for developing drugs and screening them against such pathways are beyond the scope of this paper but can be found in a recent review by Kotra et al. (2000).

Many microorganisms, such as actinomycetes, naturally produce antibiotic compounds (Demain and Elander, 1999). Through comparative and functional genomic and metagenomic studies using both microarrays and bioinformatics, microorganisms can be screened for natural antibiotic production. Gene clusters encoding antibiotic synthesis could be cloned into a BAC vector and expressed in an appropriate host. As previously mentioned, *E. coli*–*Streptomyces* artificial chromosomes (ESAC) have been developed in which genes from non-culturable actinomycete strains have been integrated into a foreign genome and expressed in a culturable host (Sosio et al., 2000). Classical genetic, molecular and biochemical tech-

niques can be employed to optimize expression of the recombinant genes. This can potentially lead to the large-scale production of natural antibiotics.

The availability of entirely sequenced microbial genomes has led to the multidisciplinary use of bacteria in genomic applications. In a preliminary study geared to the development of toxicogenomic approach to cost-effective assessment of contaminated soils, *E. coli* was used as an indicator of toxicity (Fredrickson et al., 2001). Hybridization of mRNA from uranium nitrate-exposed cells to an array of all predicted *E. coli* genes showed the upregulation of more than 100 genes, many of which are associated with stress response. Probes for the upregulated genes are being incorporated into complex arrays called genosensors that will be used to measure toxicity of contaminated soils in terms of the metagenome. The integration of bacterial genomics and novel technologies into various scientific fields may lead to unprecedented research initiatives and the building industrial and environmental applications.

7. Summary

The availability of complete genomic sequences for a growing number of microorganisms has made genomic analysis, particularly, suitable to the study of bacteria. Functional and comparative genomic characterization of microbial genomic sequencing data using bioinformatics, DNA microarrays and BACs, either alone or collaboratively, has revolutionized our understanding of microbial genomes organization and molecular mechanisms underlying cell-functions. The scientific, medical, industrial and environmental applications of bacterial genomics and their future potentials are exhaustive.

Through large-scale identification of genome sequence similarity and variability, DNA microarrays can help solve previously unanswerable questions concerning evolutionary concepts such as speciation and horizontal gene transfer. Hybridization of DNA from various fluorescent *Pseudomonas* strains to an array of genome fragments from reference Pseudomonads confirmed the degree of relatedness between well-characterized strains (Cho and Tiedje, 2001). This rapid technique can be used to determine phylogenetic lineages and identify isolates at the species to strain level. A

modified approach used an array of *Shewanella oneidensis* MR-1 ORFs to determine the degree of relatedness between other unsequenced species of *Shewanella* (Murray et al., 2001). In this study, it was discovered that genes in operons had conserved levels of hybridization over the operon across species. This knowledge may be applied to identify horizontally transfer genes between distantly related bacterial species.

BACs serve as a reservoir of microbial genomic information as genomic libraries. In addition to being used to clone large random DNA, BACs can be developed to contain sequences encoded multiple genes and regulatory elements in a heterologous host. Subsequently, such gene clusters can be easily manipulated by traditional microbiological methods for further study in vitro and in vivo (John et al., 2001; McGregor and Schleiss, 2001). BACs have also been developed to regulate expression of cloned genes in bacterial and mammalian cells (Hong et al., 2001). Bacteria may someday play a role as production vessels for BACs modified as gene therapy vectors and vaccines against viruses (Borst and Messerle, 2000; Suter et al., 1999).

An integrative approach to bacterial genomics that combines the use of BAC and DNA microarray technologies may significantly contribute to the promotion of both pure and applied scientific research including the development of novel antibiotics and the determination of the minimal genome. The minimal genome set is the theoretical smallest number of genetic elements capable of sustaining a free-living microorganism. The smallest known genome is that of the *Mycoplasma genitalium*. The complete 0.58 Mb genome sequence of this parasitic bacterium has already been determined (Himmelreich et al., 1996). Through sequence analysis and comparative studies between completely sequenced microorganisms and transposon mutagenesis, approximately 300 genes have been predicted to be essential to life (Hutchison et al., 1999). Interestingly, the function of one third of these genes remains undetermined. DNA microarrays constructed from BAC libraries of completely sequenced microorganisms can assist in both genome comparisons and analysis of gene function via expression profiling of diverse bacteria species (Mushegian, 1999). The potential exists for the integration of genomic analytical tools, such as bioinformatics, DNA microarrays and BACs, to exploit the wealth of diverse information contained in bacterial genomes.

Acknowledgements

This work was supported by a NSERC (Canada) operating grant to J.T.T.

References

- Beattie, K.L., 1997. Analytical microsystems: emerging technologies for environmental biomonitoring. In: Sayler, G.S., Sanseverino, J., Davis, K. (Eds.), *Biotechnology in the Sustainable Environment*. Plenum, New York, pp. 249–260.
- Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C., Jovanovich, S.B., Feldman, R.A., DeLong, E.F., 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2, 516–529.
- Belcher, C.E., Drenkow, J., Kehoe, B., Gingeras, T.R., McNamara, N., Lemjabbar, H., Basbaum, C., Relman, D.A., 2000. From the cover: 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.
- Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.
- Borst, E., Messerle, M., 2000. Development of a cytomegalovirus vector for somatic gene therapy. *Bone Marrow Transplant.* 25 (Suppl. 2), S80–S82.
- Brim, H., McFarlan, S.C., Fredrickson, J.K., Minton, K.W., Zhai, M., Wackett, L.P., Daly, M.J., 2000. Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments. *Nat. Biotechnol.* 18, 85–90.
- Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B.G., Cole, S.T., 1998. Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. *Infect. Immun.* 66, 2221–2229.
- Castellino, A.M., 1997. When the chips are down. *Genome Res.* 7, 943–946.
- Cho, J.C., Tiedje, J.M., 2001. Bacterial species determination from DNA–DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* 67, 3677–3682.
- Coombs, 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.
- Demain, A.L., Elander, R.P., 1999. The beta-lactam antibiotics: past, present, and future. *Antonie van Leeuwenhoek* 75, 5–19.
- DeRisi, J.L., Iyer, V.R., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.
- Favis, R., Day, J.P., Gerry, N.P., Phelan, C., Narod, S., Barany, F., 2000. Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. *Nat. Biotechnol.* 18, 561–564.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512.
- Fodor, S.P., Rava, R.P., Huang, X.C., Pease, A.C., Holmes, C.P., Adams, C.L., 1993. Multiplexed biochemical assays with biological chips. *Nature* 364, 555–556.
- Fredrickson, H.L., Perkins, E.J., Bridges, T.S., Tonucci, R.J., Fleming, J.K., Nagel, A., Diedrich, K., Mendez-Tenorio, A., Doktycz, M.J., Beattie, K.L., 2001. Towards environmental toxicogenomics—development of a flow-through, high-density DNA hybridization array and its application to ecotoxicity assessment. *Sci. Total Environ.* 274, 137–149.
- Galperin, M.Y., Koonin, E.V., 1999. Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* 106, 159–170.
- Hayward, R.E., DeRisi, J.L., Alfadhli, S., Kaslow, D.C., Brown, P.O., Rathod, P.K., 2000. Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Mol. Microbiol.* 35, 6–14.
- Hill, F., Benes, V., Thomasova, D., Stewart, A.F., Kafatos, F.C., Ansoorge, W., 2000. BAC trimming: minimizing clone overlaps. *Genomics* 64, 111–113.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., Herrmann, R., 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24, 4420–4449.
- Hong, Y.K., Kim, D.H., Beletskii, A., Lee, C., Memili, E., Strauss, W.M., 2001. Development of two bacterial artificial chromosome shuttle vectors for a recombination-based cloning and regulated expression of large genes in mammalian cells. *Anal. Biochem.* 291, 142–148.
- Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C., 1999. Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* 286, 2165–2169.
- John, R.M., Aparicio, S.A., Ainscough, J.F., Arney, K.L., Khosla, S., Hawker, K., Hilton, K.J., Barton, S.C., Surani, M.A., 2001. Imprinted expression of neuronatin from modified BAC transgenes reveals regulation by distinct and distant enhancers. *Dev. Biol.* 236, 387–399.
- Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., Shizuya, H., 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34, 213–218.
- Kononen, J., Bubendorf, L., Kallioniemi, A., Barlund, M., Schraml, P., Leighton, S., Torhorst, J., Mihatsch, M.J., Sauter, G., Kallioniemi, O.P., 1998. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* 4, 844–847.
- Kotra, L.P., Vakulenko, S., Mobashery, S., 2000. From genes to sequences to antibiotics: prospects for future developments from microbial genomics. *Microbes Infect.* 2, 651–658.

- Lee, P.S., Lee, K.H., 2000. Genomic analysis. *Curr. Opin. Biotechnol.* 11, 171–175.
- Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21, 20–24.
- Loferer, H., Jacobi, A., Posch, A., Gauss, C., Meier-Ewert, S., Seizinger, B., 2000. Integrated bacterial genomics for the discovery of novel antimicrobials. *Drug Discovery Today* 5, 107–114.
- Manger, I.D., Relman, D.A., 2000. How the host ‘sees’ pathogens: global gene expression responses to infection. *Curr. Opin. Immunol.* 12, 215–218.
- McGregor, A., Schleiss, M.R., 2001. Molecular cloning of the guinea pig cytomegalovirus (GPCMV) genome as an infectious bacterial artificial chromosome (BAC) in *Escherichia coli*. *Mol. Genet. Metab.* 72, 15–26.
- Mittenhuber, G., 2001. Comparative genomics of prokaryotic GTP-binding proteins (the Era, Ogb, EngA, ThdF (TrmE), YchF and YihA families) and their relationship to eukaryotic GTP-binding proteins (the DRG, ARF, RAB, RAN, RAS and RHO families). *J. Mol. Microbiol. Biotechnol.* 3, 21–35.
- Moch, H., Schraml, P., Bubendorf, L., Mirlacher, M., Kononen, J., Gasser, T., Mihatsch, M.J., Kallioniemi, O.P., Sauter, G., 1999. High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.* 154, 981–986.
- Moir, D.T., Shaw, K.J., Hare, R.S., Vovis, G.F., 1999. Genomics and antimicrobial drug discovery. *Antimicrob. Agents Chemother.* 43, 439–446.
- Murray, A.E., Lies, D., Li, G., Neelson, 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.
- Mushegian, A., 1999. The minimal genome concept. *Curr. Opin. Genet. Dev.* 9, 709–714.
- Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., Fraser, C.M., 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.
- Ochman, H., Moran, N.A., 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* 292, 1096–1099.
- Osin, P., Shipley, J., Lu, Y.J., Crook, T., Gusterson, B.A., 1998. Experimental pathology and breast cancer genetics: new technologies. *Recent Results Cancer Res.* 152, 35–48.
- Phimister, B., 1999. Going Global. *Nat. Genet.* 21, 1.
- Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capocchi, B., Galeotti, C.L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broecker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D.W., Jeffries, A.C., Saunders, N.J., Granoff, D.M., Venter, J.C., Moxon, E.R., Grandi, G., Rappuoli, R., 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816–1820.
- 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.
- Rondon, M.R., Raffel, S.J., Goodman, R.M., Handelsman, J., 1999. Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6451–6455.
- Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R., Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne, M.S., Clardy, J., Handelsman, J., Goodman, R.M., 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66, 2541–2547.
- Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E., Davis, R.W., 1998. Microarrays: biotechnology’s discovery platform for functional genomics. *Trends Biotechnol.* 16, 301–306.
- Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., Simon, M., 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U. S. A.* 89, 8794–8797.
- Simpson, A.J., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M., Araya, J.E., Baia, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, J.M., Briones, M.R., Bueno, M.R., Camargo, A.A., Camargo, L.E., Carraro, D.M., Carrer, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El Dorry, H., Facincani, A.P., Ferreira, A.J., Ferreira, V.C., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furlan, L.R., Garnier, M., Goldman, G.H., Goldman, M.H., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C., Lemos, E.G., Lemos, M.V., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M., Madeira, H.M., Marino, C.L., Marques, M.V., Martins, E.A., Martins, E.M., Matsukuma, A.Y., Menck, C.F., Miracca, E.C., Miyaki, C.Y., Montero-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L., Netto, L.E., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmieri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M.R.A., de Jr., R.V., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C., da Silva, A.M., da Silva, F.R., da Jr., S.W., da Silveira, J.F., Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsuhako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Meidanis, J., Setubal, J.C., 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* consortium of the organiza-

- tion for nucleotide sequencing and analysis. *Nature* 406, 151–157.
- Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A.M., Donadio, S., 2000. Artificial chromosomes for antibiotic-producing actinomycetes. *Nat. Biotechnol.* 18, 343–345.
- Suter, M., Lew, A.M., Grob, P., Adema, G.J., Ackermann, M., Shortman, K., Fraefel, C., 1999. BAC-VAC, a novel generation of (DNA) vaccines: a bacterial artificial chromosome (BAC) containing a replication-competent, packaging-defective virus genome induces protective immunity against herpes simplex virus 1. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12697–12702.
- Tanghe, A., Teunissen, A., Van Dijck, P., Thevelein, J.M., 2000. Identification of genes responsible for improved cryoresistance in fermenting yeast cells. *Int. J. Food Microbiol.* 55, 259–262.
- Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V., 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28.
- Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson J.D., Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Citti, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., Venter, J.C., 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287, 1809–1815.
- Trevors, J.T., 1996. Genome size in bacteria. *Antonie van Leeuwenhoek* 69, 293–303.
- van Hal, N.L., Vorst, O., van Houwelingen, A.M., Kok, E.J., Peijnenburg, A., Aharoni, A., van Tunen, A.J., Keijer, J., 2000. The application of DNA microarrays in gene expression analysis. *J. Biotechnol.* 78, 271–280.