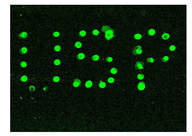
Identification of Bacterial Species Using 16S rRNA Microarray Technology

J.C. Pierce, K. Patel, R.Khan, R. Shah, J. Greenwood, A.S. Wenocur

Bioinformatics Program, University of the Sciences in Philadelphia

Poster # W-018



Abstract

The goal of this project is to develop DNA-based microarray technology for use in the undergraduate teaching laboratory. We have previously developed a PCR-based student exercise for the identification of bacterial species based on the DNA sequence fingerprint of the 5-prime region of the 16S ribosomal RNA (rRNA) gene. In an effort to continue to incorporate modern biotechnology into the classroom, we have integrated microarray technology into this project. Microarray technology allows for the quantitative analysis of large numbers of DNA and RNA molecules. Introducing microarray technology into the undergraduate classroom is a challenge due to the cost and the difficulty of analyzing large datasets. Our microarray exercise allows students to perform microarray experiments in a cost effective and goal-directed manner.

To develop the 16S rRNA identification microarrays, we generated an 800 bp DNA fragment by the PCR, which were spotted onto glass microarray slides. A total of nine different bacteria have been used (*Escherichia coli, Salmonella typhimurium, Pseudomonas aeruginosa, Vibrio fischeri (campbellii), Rhodosprillium rubum, Enterococcus faecium, Staphylococcus epidermidis (warneri), Bacillus subtilus, Mycobacterium avium (smegmatis)*). These bacteria were chosen to represent the types of organisms that are often encountered in the undergraduate microbiology-teaching environment. Unknown bacterial samples were labeled with cyamine-3 dCTP by direct incorporation during the PCR. Hybridization of the Cy3-labeled sample was performed and the slides scanned using a Bio-Rad ChipReader. Data was analyzed using the Excel spreadsheet program. A positive and negative hybridization control was developed based on the *Limulus* 18S rRNA gene.

Results of experiments have shown that we have been able to successfully construct 16S DNA microarray slides and purify labeled DNA fragments. Initial microarray analysis has shown that unknown bacterial species can be detected using the microarrays although problems with cross hybridization between species are evident. A web-based teaching resource is available (http://bioinformatics.usip.edu/undergraduate).

Introduction

Microarray Technology has become an important tool in biotechnology for the identification of large sets of nucleic acids (DNA and RNA) and proteins. The global analysis of cellular gene expression is one of the most promising developments of microarray technology. Major advancements in developmental biology, cellular adaptation, and molecular diagnostic predication have already been achieved with this technology. A second research area of promise for microarray technology is the identification of specific genotypes and species based on hybridization to arrays of predetermined sequences. As large databases of complete genomes and specific genes (ribosomal RNA genes) become available, scientists can theoretically design arrays that can identify any life form.

Introducing undergraduate students to microarray technology is a challenging educational task. The equipment and software to perform advanced microarray experiments are expensive and require significant expertise to use reliably. Commercial sources of spotted arrays are also expensive and allow for relatively little chance for students to "practice" learning this technology. Although the entry cost to microarray technology is high, there have been a number of efforts to give undergraduate students knowledge and experience in performing this technology. One successful method implemented by the Genome Consortium for Active Teaching (GCAT) obtains whole-genome microarray from academic sources and distributes those chips to academic participants. The host institution centralizes microarray scanning and data analysis is supported by custom image analysis software. We have built on these efforts by designing a microarray platform, which simplifies the construction, processing and data analysis of microarrays to give undergraduate students and educators an introductory experience in microarray technology. A central theme to our approach is pedagogical one; if students can perform relatively simple microarray experiments and data analysis on a system that they are already familiar with, then more complicated microarray experiments (eg. whole genome) will be easier to master.

The identification of unknown bacteria is a common and almost universally understood laboratory exercise in the undergraduate microbiology curriculum. Students are given an "unknown" bacterial culture and are require identifying the genus and species based on a number of physical and biochemical tests. We have used this exercise as a foundation for our microarray technology-teaching platform. A number of diverse bacteria are used to amplify an 800 bp fragment of the ribosomal 16S gene and these DNA's are spotted onto a glass microarray slide. An "unknown" bacterial culture is used as a template to label the 16S rRNA PCR fragment with Cy3-dCTP. The Cy3 labeled DNA is hybridized to the microarray and the data analyzed to determine the best match. We have developed a positive control sample that is labeled with Cy5 dCTP for inter-chip comparisons. Negative controls are integrated across the array for analysis of background non-specific hybridization. Our first generation 16S microarrays were successful in identifying the Mycobacterium species in our test set.

Methods

16S Microarray Construction

Grow Bacteria

↓

Isolate Bacterial Genomic DNA P

↓

PCR Amplification of 16S rRNA gene P

↓

Gel Electrophoresis P R

↓

DNA Sequencing P R

↓

Bioinformatic Sequence Analysis P R

↓

Microarray Construction P1 P2

16S Microarray Experiments

Grow Bacteria

↓

Cy3 Labeling by PCR P1 P2

↓

PCR Purification P

↓

Gel Electrophoresis P R

↓

Microarray Hybridization and Washing P1 P2

↓

Microarray Scanning

↓

Image Analysis

↓

Bioinformatic Microarray Analysis R

Detailed copies of Methods and Protocols can be accessed over the Internet via the following URL:

http://bioinformatics.usip.edu/undergraduate/

Results

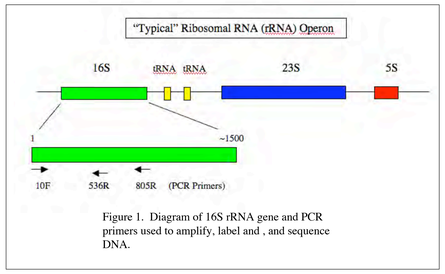


Figure 1. Diagram of 16S rRNA gene and PCR primers used to amplify, label and , and sequence DNA.

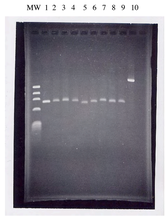


Figure 2. Gel electrophoresis of 800 bp 16S rRNA gene from 9 different bacteria. Lane 10 is the 18S rRNA gene from *Limulus polyphemus* (horseshoe crab).

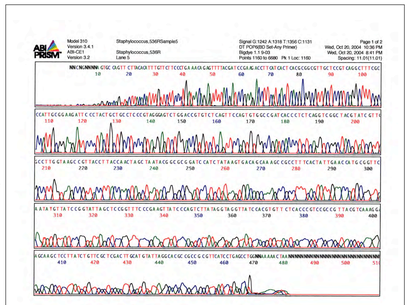


Figure 3. Electropherogram of 16S rRNA gene using the nested primer 536R.

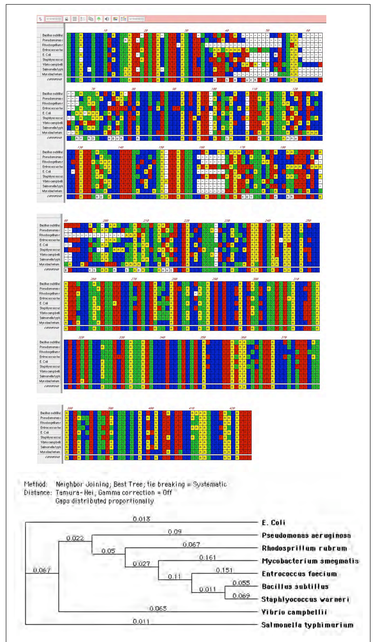


Figure 4. Multiple sequence alignment and phylogenetic reconstruction of the nine different bacteria used in this study.

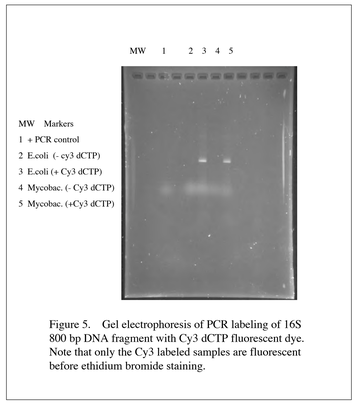


Figure 5. Gel electrophoresis of PCR labeling of 16S 800 bp DNA fragment with Cy3 dCTP fluorescent dye. Note that only the Cy3 labeled samples are fluorescent before ethidium bromide staining.

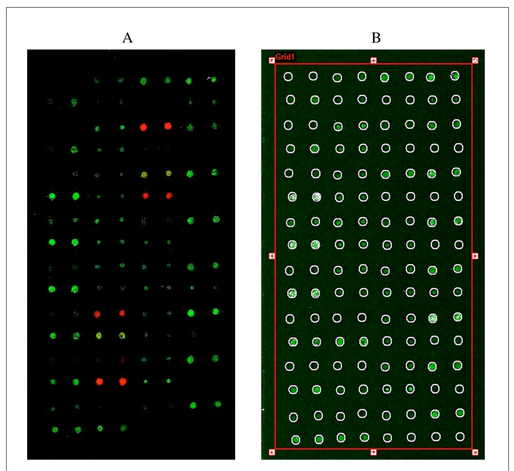


Figure 6. (A) Scan of 16S microarray slide that has been hybridized with a Cy3 dCTP-labeled *Mycobacterium* 800 bp DNA fragment. A Cy5 dCTP-labeled *Limulus* 18S DNA fragment has also been added to the hybridization reaction. (B) The same slide showing the Cy3 signal from the microarray scanner (BioRad Chipreader) and the grid pattern that was used to collect the hybridization signal.

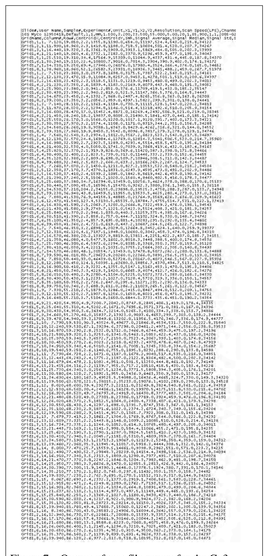


Figure 7. Output of text file report for the Cy3 fluorescent signal channel after image analysis of the 16S microarray slide. Each scan of the slide produces two such files, the Cy3 and Cy5 data reports.

Figure 8. Grid layout for the 16S rRNA microarray slides. Control spots are spotting buffer only. Positive controls are the *Limulus* 18S rRNA gene. Sample spots contain the 16S 800 bp fragment at a concentration of 50 ng/ul. All sample and positive control spots were also gridded at a 1:10 dilution.

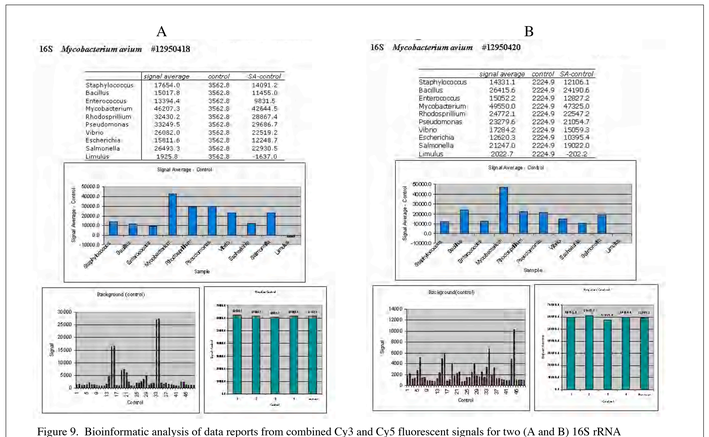


Figure 9. Bioinformatic analysis of data reports from combined Cy3 and Cy5 fluorescent signals for two (A and B) 16S rRNA microarray slides hybridized with a *Mycobacterium* labeled 800 bp fragment. Each sample is reported as the average signal from four (4) spots per slide. The combined negative controls are summed and the average value subtracted from each sample signal to generate the reported value. Total signal values are reported for all 48 negative controls. Positive controls (Cy5 signal) are reported for each spot and the average value generated for chip to chip comparisons.

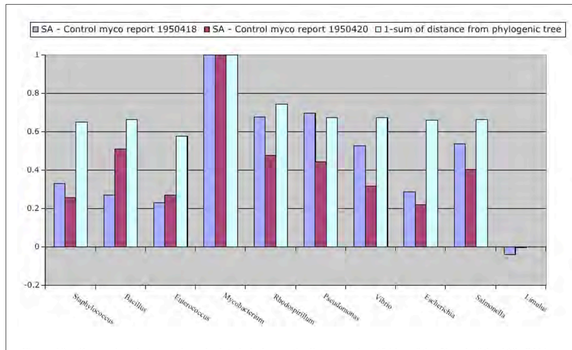


Figure 10. Comparison between the calculated phylogenetic distance using 424 bp of the 5' end of the 16S rRNA gene of the microorganisms used in this study and the average Cy3 fluorescent signal for the samples from the 16S microarray slides hybridized with *Mycobacterium* DNA. Sample signal values are reported as the ratio of each species signal to the highest signal. The phylogenetic distance is calculated as one minus the sum of the tree distances for each of the bacteria relative to the bacterial species with the highest Cy3 signal (*Mycobacterium*).

Conclusions

We have developed a microarray technology platform that uses the bacterial 16S ribosomal RNA (rRNA) gene as a target for species identification. The primary goal of this project was to build a robust system for the teaching microarray technology in the undergraduate curriculum. We have previously developed a microbiology laboratory exercise for the identification of "unknown" bacteria using a DNA sequencing approach. Students were given a culture of an "unknown" bacterium and used the polymerase chain reaction (PCR) to amplify an 800 base pair fragment from the 5' end of 16S rRNA gene. PCR primers were chosen based on the highly conserved sequences that are present at positions of 10 bp and 805 bp in the consensus *Escherichia coli* 16S rRNA genes. Use of these phylogenetically universal primers in the PCR allows for the amplification of an ~800 bp DNA fragment from numerous genera of bacteria including both Gram-positive and Gram-negative bacterial species. Using a nested reverse primer (536R), students are able to sequence the 5' portion of the 800 bp 16S rRNA DNA fragment. Analysis of the resulting sequence data using the BLAST program at the National Center for Biotechnology Information (NCBI) matches the sequence against the GenBank database for species identification. Interestingly, alignment results based on the BLAST matches showed that a number of bacteria from our collection and from commercial sources were keyed differently from their labeled genus/species identities. In this work we have adapted our bacterial identification studies to design a similar exercise that uses microarray technology.

In our first generation microarray slide design, we chose nine different bacterial strains that are often used in teaching in the microbiology undergraduate curriculum. The 800 bp PCR product was purified and the concentration normalized to 100 ng/ul for spotting on microarray glass slides. Using a manual 32-pin spotting apparatus (V & P Scientific) we designed a 128 spot grid pattern (16 columns x 8 rows) that contained DNA for each bacterial sample, a 1:10 dilution of each sample, a series of positive controls, and negative control spots. The 1:10 dilutions samples were used to determine the sensitivity of hybridization and were found to be too low to be useful. Positive control spots are based on the 18S rRNA gene from the horseshoe crab, *Limulus polyphemus*. A PCR product generated from *Limulus* DNA is labeled with Cy5-dCTP and used to calibrate the level of DNA hybridization, slide washing, and scanning sensitivity for each microarray slide used in our studies. This is an important molecular tool for inter-chip comparisons and for troubleshooting microarray experiments. The negative controls consist of buffer-only spots. Gridding errors, sample well contamination, slide washing and hybridization bias can be monitored using the dispersed negative controls. Early experiments showed a significant bias in the Cy3 hybridization signal across the microarray slide that were eventually resolved by gently mixing the slide on a rocking platform for 15 minutes after the hybridization mix was applied to the slide coverslip and then performing the overnight hybridization. Results from our first generation 16S rRNA microarray experiments showed that we could clearly identify the test bacteria when there was a significant phylogenetic distance between the test strain and the other samples spotted on the microarray. The *Mycobacterium* species, identified by BLAST analysis to be most similar to *Mycobacterium smegmatis*, was clearly identifiable by microarray analysis. Interestingly, as shown in Figure 10, when the calculated phylogenetic distance was plotted relative to the microarray fluorescent signals for each bacterial sample, there was a good correlation between the two replicate slides for the *Mycobacterium* sample and the other bacteria spotted on the array. This was true even though the Cy3 signal was less in the second replicate that the first *Mycobacterium* microarray slide. Issues with cross hybridization between species are evident in the data but this is to be expected due to the significant amount of sequence conservation in the ribosomal RNA genes. Other genomic loci may be more sensitive for identification by microarray hybridization. We have investigated the intervening spaces between rRNA genes as possible sites for future microarray experiments.

In this work we have develop a microarray technology platform for teaching in the undergraduate curriculum. The cost of constructing our 16S rRNA microarray slides are relatively inexpensive, especially when compared to commercial providers. The total cost for reagents are not excessive with the major expense being the Cy3 and Cy5 dCTP. DNA for microarray spotting is generated by the PCR negating the need for arrays of synthetic oligonucleotides. We used a manual printing apparatus to make the microarray slides. While robotic microarray slide printing would undoubtedly give better and more uniform results, having students actually gird their arrays is useful training and the cost of the manual gridder is much less than the investment and maintenance of robotic equipment. No advanced microarray bioinformatic software was used in the analysis of the microarray results. An Excel spreadsheet software program was used to normalize, analyze and view data. A major impediment is availability of a microarray scanner. Central scanning facilities are a workable concept as demonstrated by the GATC and availability of microarray scanners is hopefully expected to only increase in the future giving more students access to this fundamental biotechnology.

References

Reyes-Lopez, M.A., A. Mendez-Tenorio, R. Maldonado-Rodrigez, M.J. Doktycz, J.T. Fleming, and K.L. Beattie. 2003. Fingerprinting of prokaryotic 16S rRNA genes using oligodeoxyribonucleotide microarrays and virtual hybridization. Nucleic Acids Res. **31**:779-789.

Russo, E. 2003. The microarray family tree. The Scientist. **17 (16):**29-31.

Genome Consortium for Active Teaching (GATC) 2005. http://www.bio.davidson.edu/projects/GCAT/gcat.html

National Center for Biotechnology Information (NCBI) 2005. http://www.ncbi.nlm.nih.gov/

National Science Foundation WHERE DISCOVERIES BEGIN

This work was supported by the National Science Foundation NSF-CCLI grant # 0126814