Application of the Polymerase Chain Reaction (PCR) and Bioinformatics in Bacterial Identification

Harriet S. Wilson
Professor of Microbiology, Biological Sciences Department, Sierra College

Abstract

Students in microbiology classes are frequently unable to identify bacterial isolates using traditional laboratory methods including observation of colonies, microscopic examination of stained materials and analysis of phenotypic characteristics. In this study various methods for extracting DNA from live bacterial cultures were explored, and the Polymerase Chain Reaction was used to amplify a portion of 16S ribosomal DNA approximately 1500bp in length from 25 different types of bacteria commonly encountered in a non-clinical laboratory setting. Different methods and materials for concentrating and purifying DNA samples were investigated, and segments of DNA between 500 and 800bp from 14 different types of bacteria were sequenced. Nucleotide sequences were compared to data recorded in the public database of NCBI using the BLAST algorithm and sequence homologies were determined. Of the 14 samples sequenced, five showed 100% homology with previously identified bacterial strains and one was found to be 100% homologous with an Antarctic bacterium of unknown identity. Six of the remaining isolates were 99% homologous with previously identified strains, one was 98% homologous and one 97% homologous. Samples showing low homology were suspected of being from mixed cultures. Application of the PCR, nucleotide sequencing and bioinformatics was found to be effective addition to traditional methods of bacterial identification in undergraduate microbiology laboratories.

Introduction

Students working on independent projects associated with the major's course in microbiology at Sierra College often choose to identify organisms (primarily bacteria) from the environment. Many of these organisms are isolated from air samples, while others are collected from water, soil, plant materials or other sources. Identification techniques typically involve observation of colonies on solid media, microscopic examination of stained materials, antimicrobial sensitivity testing and analysis of data obtained from a variety of enzymatic tests. The Bergey's Manuals of Systematic and Determinative Bacteriology serve as primary sources of reference information, and additional resources are utilized when available. Although many types of bacteria are readily identified using these methods and materials, others are not, and students are frequently unable to collect enough data to assign putative names to their isolates. Additional methods for bacterial identification are then needed.

The objective of this project was to develop reliable and relatively inexpensive methods for applying DNA technology and bioinformatics to the identification of unknown bacteria. Initially this involved determining reliable methods for obtaining DNA samples from live bacterial

cultures, amplifying portions of 16S ribosomal DNA from these using the PCR and obtaining sequence data that could be used for bacterial identification. Although successful amplification of DNA samples with the polymerase chain reaction presented the first major challenge, it was not the only one. Finding suitable methods for maintaining DNA samples, visualizing results using gel electrophoresis, determining DNA concentrations and accumulating DNA in quantities sufficient for sequencing, added to the complexity of the project. Finally, utilization of computer technology to receive, analyze, compare and present the data obtained provided a learning experience for everyone involved. Since another goal of this project was to make information available for the assistance of others, this document is broken into segments intended to improve access.

I. The Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR), a powerful diagnostic tool developed by Kary Mullis and his coworkers at Cetus Corporation (1986), provides an interesting addition to traditionally used bacterial identification methods. The polymerase chain reaction allows specific segments of DNA to be amplified (replicated over and over again) in vitro, by exploiting specific features of DNA structure and its replication process as follows. A cellular DNA molecule is composed of two nucleotide strands that are complimentary to one another and antiparallel (up-side-down relative to each-other). The complimentary base pairs of a DNA double helix are held together by relatively weak hydrogen bonds, and can be induced to separate in vitro by the application of heat. During DNA replication, DNA-dependent DNA polymerase enzymes use the nucleotide sequences of existing DNA strands as templates or patterns for building new, complimentary strands. DNA polymerase builds DNA by catalyzing the formation of phosphodiester bonds binding nucleotides to the free 3' ends of existing nucleotide chains (i.e., builds from 5' to 3' and requires a primer). A single-stranded oligonucleotide sequence that is complimentary to a region of DNA can hybridize with it (anneal to it) and serve as the primer sequence (providing the 3' end) required for DNA polymerase. The energy necessary for DNA synthesis (bond formation) in vitro can be provided by nucleoside triphosphates (dNTPs or rNTPs). The polymerase chain reaction can be readily initiated in vitro by mixing template DNA, oligonucleotide primers, DNA polymerase and dNTPs with reaction buffer in a thin-walled tube, and then modulating the temperature to stimulate alternate cycles that denature the DNA, allow the primers to anneal, and then promote chain elongation. By repeating this process over and over again, it is possible to amplify a single strand of DNA more than a million times within a few hours.

In order to apply the PCR to the identification of bacteria isolated from random environmental samples, it is useful to amplify regions of DNA that are common to all organisms and that contain highly conserved regions, i.e., nucleotide sequences (base sequences) that have remained stable over time. Regions of DNA (genes) that code for ribosomal-RNA (r-RNA) exhibit these features. All known cellular organisms contain ribosomes, structures essential for protein synthesis, and consequently all of these organisms contain r-RNA. A bacterial ribosome (70S) contains one molecule each of three types of RNA identified as 23S, 16S and 5S (S=Svedberg unit – sedimentation coefficient). Since the function of ribosomes within cells has apparently remained constant over time, the nucleotide sequences of ribosomal-RNA molecules tend to be highly conserved. Certain regions of the16S r-RNA have remained extremely highly conserved during evolution, so provide ideal material for identification and classification. Sequence

homology studies involving 16S r-RNA or the DNA coding for it (16S ribosomal DNA) have been used extensively to determine evolutionary relationships between organisms, particularly bacteria and archaea. For this study, regions of 16S ribosomal-DNA approximately 1500 base pairs in length were amplified and portions of these were sequenced in an attempt to determine bacterial identity.

Materials and Methods

Bacterial strains – The bacteria used in this study included both Gram-positive and Gram-negative forms obtained from cultures available in the laboratory. Some of these, purchased from Becton Dickinson and Company as in vitro diagnostic discs, included *Escherichia coli* (ATCC#25922), *Citrobacter freundii* (ATCC#8090), *Pseudomonas aeruginosa* (ATCC#27853), *Salmonella choleraesuis* (ATCC#14028), *Klebsiella pneumoniae* (ATCC# 13883), *Serratia marcescens* (ATCC# 8100), *Proteus vulgaris* (ATCC#13315), *Enterobacter aerogenes* (ATCC#13048), *Enterobacter cloacae* (ATCC#23355) and *Acinetobacter baumannii* (ATCC#19606). Others were collected from air, soil or water, tentatively identified using traditional laboratory methods and assigned putitive names. These included *Janthinobacterium lividum*, *Azotobacter nigricans*, *Pseudomonas fluorescens*, *Proteus mirabilis*, *Klebsiella oxytoca*, *Escherichia coli*, *Xenorhabdus poinarii*, and *Micrococcus roseus*. Additional cultures being investigated by students during the spring semester of 2004 included isolates from the genera *Staphylococcus*, *Bacillus*, *Vibrio*, *Rhodococcus, Nocardia,* and *Porphyrobacter* (as indicated by nucleotide sequence analysis). None of the cultures investigated were clinical isolates. Media used for culture maintenance were selected as indicated by the growth requirements of the organisms. Commercial media types included tryptic soy broth, tryptic soy agar, nutrient agar, Mueller-Hinton agar, and MacConkey agar (Difco-BBL). Defined media prepared in the laboratory included a nitrogen-free mannitol medium formulated for the isolation of *Azotobacter* species.

Genomic DNA Extraction – Various methods were used for extracting chromosomal DNA from live cultures. Some Gram-negative organisms grown for 18-24 hours (over night) in broth cultures were placed in 1.5 mL microfuge tubes (1 mL each) and boiled for 10 minutes. These were then centrifuged for 5 minutes to pellet the cellular material, the supernatant was poured off, and the solids were resuspended in 25μL of Tris buffer (10mM, pH 8.5) by vortex mixing. Some Gram-negative cultures were grown on solid media (over night or longer as required to obtain colonies) and then transferred into Tris buffer before boiling for 10 minutes. A single, medium-sized colony (cell mass 2mm in diameter) placed in 500μl of liquid, or a single small colony (cell mass 1mm in diameter) placed in 100μl of liquid provided a sample of suitable density. This was vortex-mixed to obtain a uniform suspension and then boiled for 10 minutes. Some genomic DNA was extracted from Gram-negative cultures with mini-spin extraction columns (Qiagen – DNeasy Tissue Kit) and Genomic DNA from some putative Gram-negative and all Gram-positive cultures was extracted with bead-beater kits (MoBio – Ultra Clean Microbial DNA Extraction Kit).

Enzymatic amplification – All DNA amplification involved the primer set Bacteria 8 forward (5'AGAGTTTGATCCTGGCTCAG3') and Universal 1492 reverse (5'ACGGCTACCTTGTTACGAC3') flanking a region of 16S ribosomal DNA approximately

1500 base pairs in length. Primers were purchased from Qiagen and were suspended in Tris buffer (10mM, pH 8.5). Each reaction mixture contained 25μL Taq PCR Master Mix (Qiagen – Taq PCR Master Mix Kit), 5μL primer mix at a final concentration of 5μM each (5 pica moles per μL each), 5μL template DNA and 15μL sterile, distilled water (pH 8.0). Amplification cycles (35) were proceeded by a 4-minute denaturation at 94ºC (loading samples into a hot block) and included 45 seconds at 55ºC (to anneal), 2 minutes at 72ºC (to extend), and 30 seconds at 94ºC (to denature). During the final cycle, the extension time was lengthened to 20 minutes and the denaturing step was omitted. All amplifications were run on a Gene Cycler (BioRad). DNA samples were electrophoresed in 1% agarose gels, stained with ethidium bromide, and quantified under UV illumination by comparison to bacteriophage lambda DNA cut with restriction endonuclease (either *Hin*dIII or *Pst*I). Repeat amplifications of PCR products were performed by diluting 1μL amplicon samples in 999μl Tris buffer (10mM, pH 8.5) and using this dilution as template. Some PCR product was purified with QIAquick Gel Extraction kits (Qiagen) prior to repeat amplification.

Additional Equipment – All DNA, enzyme and water samples were kept on ice prior to and during PCR protocols. Shaved-ice (Hawaiice) supported the thin-walled, 2ml PCR tubes. All liquid samples were transferred with Eppendorf Reference digital pipettes (.5-10μL and 10-100μL volumes) using standard tips. Reaction mixtures were set up by placing Taq master mix in the tube bottoms, and then adding specified volumes of additional liquids to the inside surfaces of tubes as separate droplets. After all components of the reaction mixtures had been added, tubes were placed into a Microfuge E (Beckman) for a momentary spin, and then transferred to the preheated block of the thermal cycler. Electrophoresis was conducted in homemade Plexiglas gel boxes or an E-C Minicell Electrophoretic Gel Systems (E-C). Electrophoresis results were made visible and photographic images were captured with a Foto/Phoresis I Transluminator and Polaroid camera setup (Fotodyne).

Results

A region of 16S ribosomal DNA approximately 1500 base pairs in length was amplified from 25 different strains of bacteria representing 19 genera (as listed) during the spring and summer of 2004. DNA extraction methods involving the boiling of cellular suspensions in Tris buffer were successful for all ATCC, Gram-negative and for most putative Gram-negative cultures, but did not yield results with any Gram-positive cultures. Some cultures identified by students as Gram-negative, and grown on media selective for Gram-negative forms (e.g., MacConkey agar), did not generate PCR product when boiled, but did when their genomic DNA was extracted using a kit. The DNeasy Tissue Kit (Qiagen) worked well for extracting genomic DNA from Gram-negative cultures, but required additional reagents (not supplied) for Gram-positive cultures. The Ultra Clean Microbial DNA Extraction Kit (MoBio) worked well for all Gram-positive and some putative Gram-negative cultures that would not yield results with other methods.

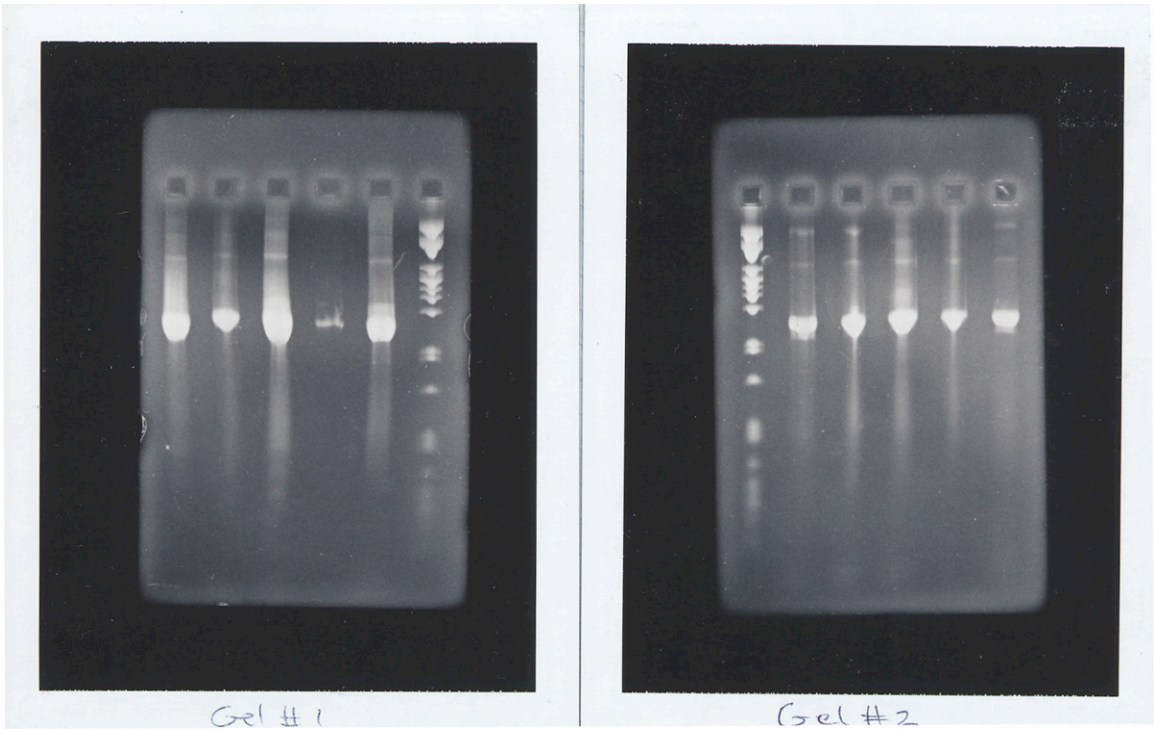

Figure 1 – Two gels containing PCR product DNA in high concentration (6μL per well). Lane 6 of gel #1 and lane 1 of gel #2 contain bacteriophage lambda DNA cut with *Pst*I.

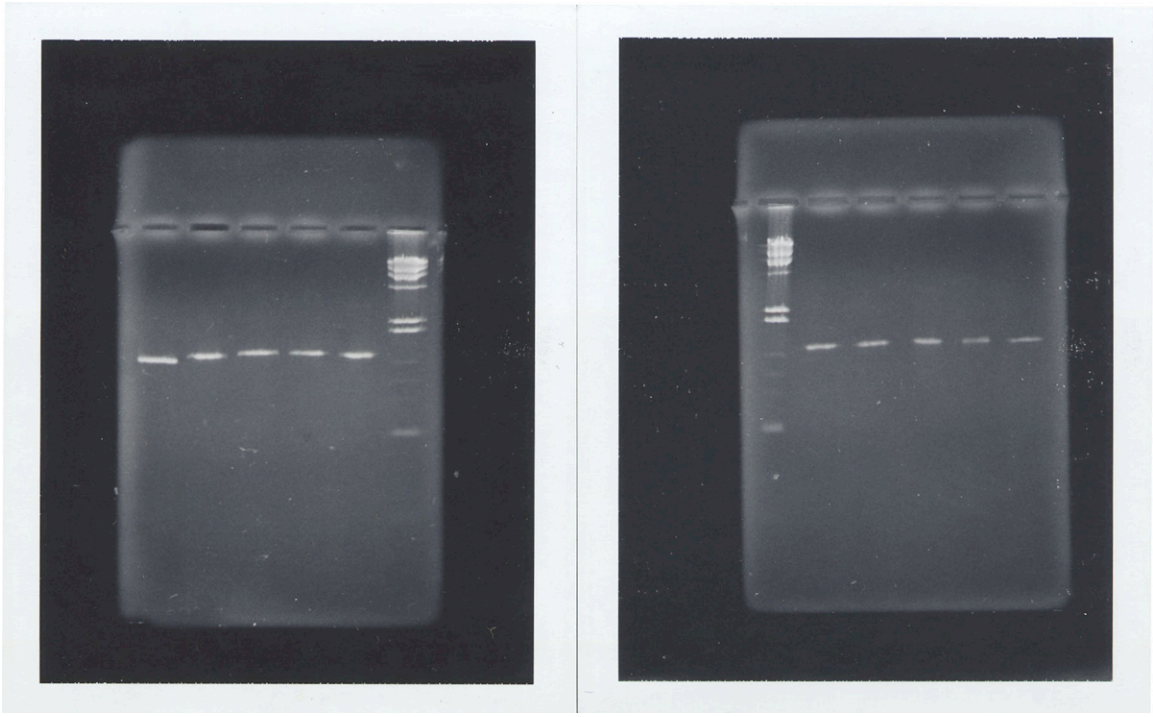

Figure 2 – Two gels containing PCR product DNA cleaned with QIAquick Gel Extraction kits (Qiagen). Bacteriophage lambda DNA is cut with HindIII.

PCR products subjected to gel electrophoresis appeared as clearly visible bands (bright patches) in gels, but often showed considerable smearing and were sometimes accompanied by accumulations of larger fragments presumed to be chromosomal DNA. In order to obtain clean DNA samples for nucleotide sequencing, PCR products from 10 different bacteria were cut from gels, weighed, and cleaned with QIAquick Gel Extraction Kits (Qiagen). The gel extraction procedure was successful, but significantly reduced the concentration of PCR product present in each sample, and did not yield DNA at a concentration suitable for nucleotide sequencing (30ng/μL). To alleviate this problem, three samples of PCR product were subjected to a concentrating procedure as follows. PCR product DNA from the same culture was run in four adjacent wells and then combined in a single spin column (QIAquick Gel Extraction Kit). The final elution volume used was 30μL. DNA samples from three different cultures were concentrated in this manner and then used to make dilutions (1:5 and 1:10) using Tris buffer. The concentrated DNA, 1:5 dilution and 1:10 dilution were then run in adjacent lanes of two gels in an attempt to determine the concentration of the original PCR product DNA.

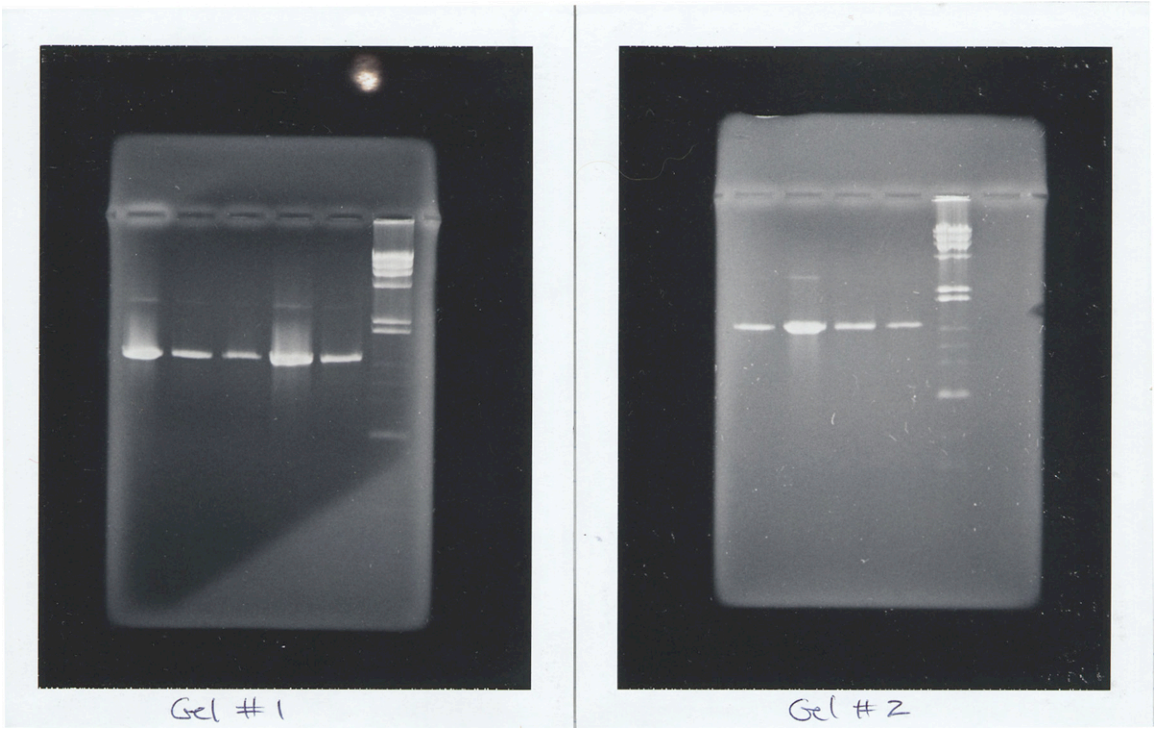

Figure 3 – Two gels containing PCR product DNA from three different cultures in concentrated form (lanes 1 and 4 of gel #1 and lane 2 of gel #2) followed by 1:5 and then 1:10 dilutions of the same samples. Bacteriophage lambda DNA is cut with *Hind*III.

The DNA concentration present in the doublet-pair bands of the bacteriophage lambda was calculated to be slightly more than 4ng/μL. Using this concentration as a comparison, it was estimated that the concentration of DNA present in the initial PCR product samples was approximately 40ng/μL, a concentration suitable for nucleotide sequencing.

Students working with unidentified Gram-negative bacterial cultures during the spring semester of 2004 sometimes used mini-spin extraction columns (Qiagen – DNeasy Tissue Kit) to extract genomic DNA from their cultures prior to amplification with the PCR and sometimes did not. Cell samples boiled for 10 minutes in Tris buffer often yielded PCR product comparable to that obtained with extracted DNA samples.

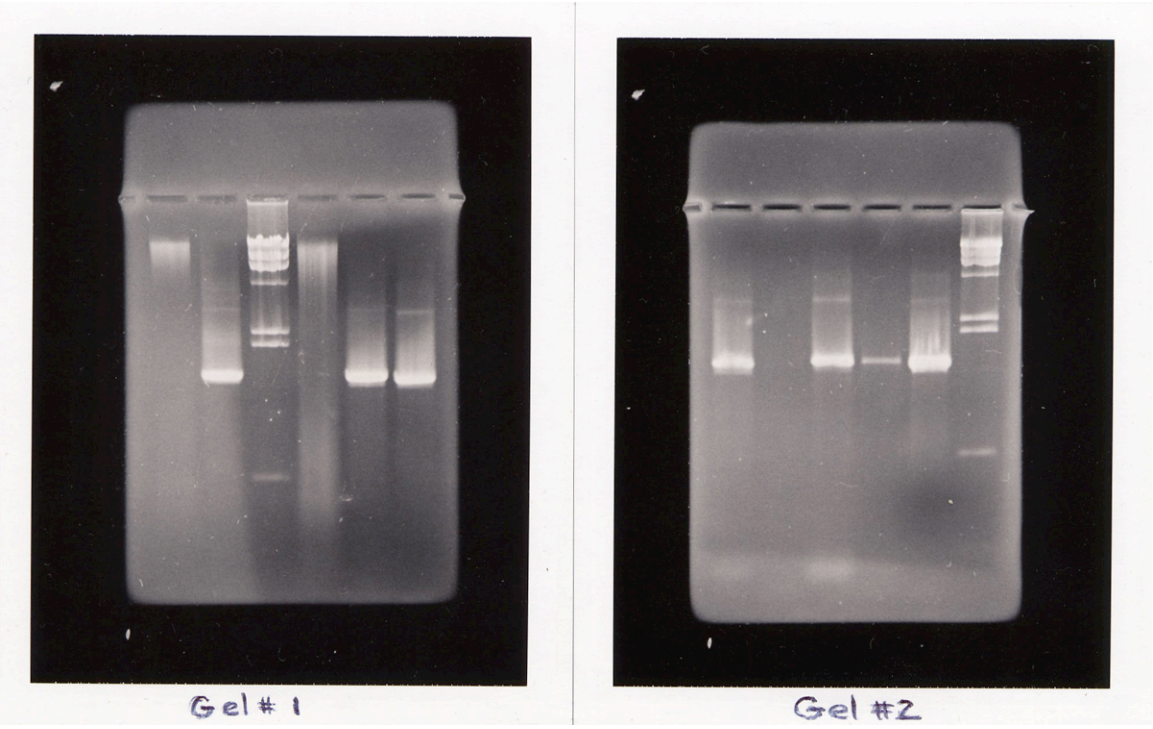

Figure 4 – Two gels containing PCR product DNA. Gel #1 contains extracted genomic DNA in lanes #1 and #4, PCR product in lanes #2, 5 and 6. Lane #6 contains PCR product from a boiled cell suspension. Gel #2 contains PCR product from four different bacterial cultures all boiled in Tris buffer. Bacteriophage lambda is cut with *Hind*III.

This data suggests that extracting DNA from Gram-negative cultures by boiling them for 10 minutes in Tris buffer (10mM, pH 8.5) is about as effective as extracting DNA using the more expensive spin-filter extraction kits. Although cell concentrations influenced outcomes with all DNA extraction methods, results were variable even when concentrations appeared to be the same, i.e., when cell suspensions were prepared in the same manner using cultures of the same age. PCR amplification was generally not successful if the cultures used as sources of genomic DNA were old (days or weeks depending on the culture). Freshly grown samples usually yielded the best results.

II. DNA Sequencing and Bioinformatics

DNA sequencing, the process of determining the sequence or arrangement of nucleotides (bases) in a sample of DNA molecules, is an essential step in the identification process, and in most cases involves the PCR. Sequencing facilities currently use highly sophisticated, automated systems that employ dideoxynucleotides and the chain termination method for nucleotide sequencing. A dideoxynucleoside triphosphate (ddNTP) is an analog of a dNTP that lacks a hydroxyl group on the third carbon of its sugar. When incorporated into DNA strands during replication, ddNTPs are unable to form phosphodiester bonds with incoming nucleotides, and thus efficiently terminate DNA synthesis. By adding small amounts of ddNTPs to a PCR reaction-mixture containing template DNA, dNTPs and a single primer, it is possible to generate populations of oligonucleotides that terminate at every position in the template strand. In Big Dye sequencing systems, the four different types of ddNTPs each carry a different colored fluorescent label, and the oligonucleotides generated are subjected to electrophoresis within capillary tubes. Lasers are used to excite the fluorescent labels, and a camera captures the color patterns generated. Data collection software and computer analysis is then used to generate an electropherogram, a visual record of the DNA sequence data.

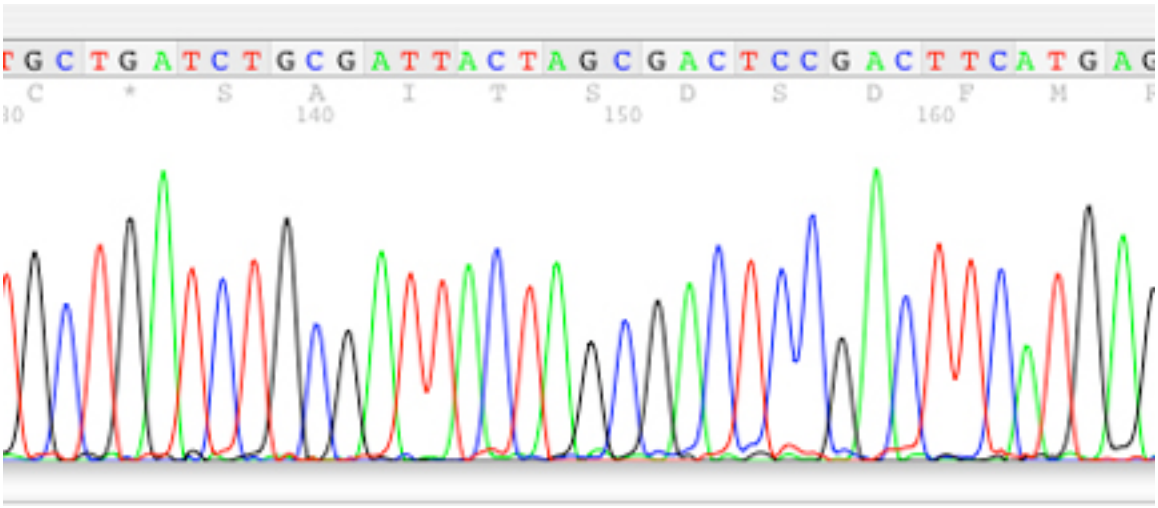

Figure 5 – A portion of an electropherogram showing four-colored peaks and the corresponding text file indicating the base sequence present in a sample of DNA.

The electropherograms generated by automated sequencers are four-color chromatograms displaying sequencing results as a series of peaks. Each peak represents an accumulation of oligonucleotides ending with a specific base as designated by color, and peak heights (intensity of signal) indicate the relative number of oligonucleotides present in each size category. In addition to the data represented by colored peaks, sequencing machines generate text files showing their interpretation of this data (typically 500bp with 98% accuracy). The machines cannot verify the validity of the text files generated, so human interpretation and editing is necessary. For assistance with sequence interpretation and editing, see Interpretation of

Sequencing Chromatograms presented by the DNA Sequencing Core at the University of Michigan (http://seqcore.brcf.med.umich.edu/).

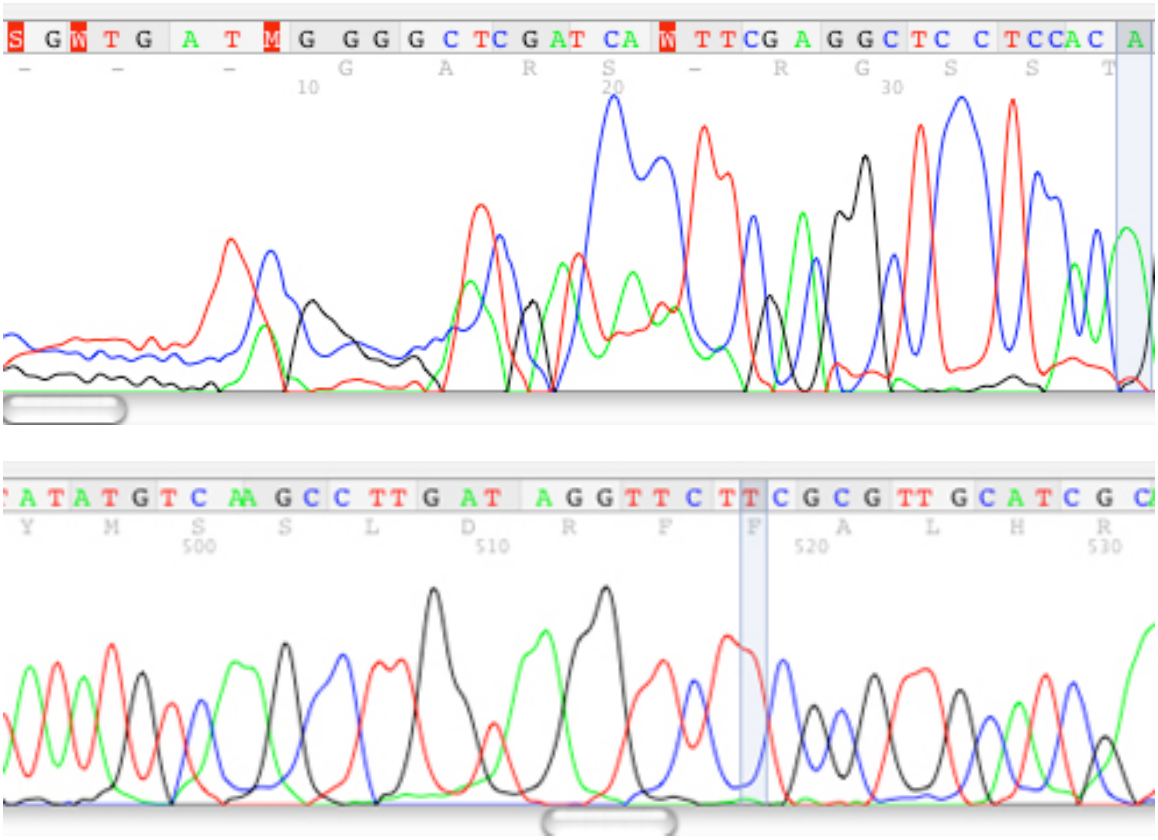

Figure 6 – Portions of an electropherogram showing beginning and later sections of a sequence where much of the data is unreliable. The researcher must edit this data to insure accuracy of the sequence recorded. Note overlapping peaks near the beginning, and miscalled bases around 510.

Computer technology is essential to bacterial identification using nucleotide sequencing because computers generate the data recorded in electropherograms, and are used to manipulate this data, store it and compare it to information available in public databases. Gene banks such as those maintained by the National Center for Biotechnology Information (NCBI) can be accessed via computers, and nucleotide sequence data obtained from bacterial isolates can be quickly and easily compared to data obtained by other researchers. The application of computer technology to the storage, comparison, analysis and interpretation of biological data has led to the development of a new field of study called Bioinformatics.

For the portion of this study completed during the spring semester of 2004, 16S ribosomal DNA from 14 different types of bacteria was amplified by means of the PCR, concentrated, purified and taken to a sequencing facility. Sequencing results were compared to the gene banks maintained by NCBI using the computer algorithm BLAST (Basic Local Alignment Search Tool). All of the isolates were found to have nucleotide sequences at least 97% homologous to previously identified bacterial species.

Materials and Methods

Bacterial strains – The bacteria used in this section of the study included both Gram-positive and Gram-negative forms obtained initially from air, soil and water. Some of these were tentatively identified using traditional laboratory methods and assigned putative names including *Escherichia coli*, *Janthinobacterium lividum*, *Azotobacter nigricans*, *Xenorhabdus poinarii,* and *Micrococcus roseus*, while others were only partially identified. None of the cultures investigated were clinical isolates.

PCR and Sequencing Preparation – All of the DNA samples prepared for sequencing were PCR amplicons obtained by amplifying PCR products previously diluted in Tris buffer (10mM, pH 8.5). Dilutions were usually 1:1000, but sometimes 1:500 or 1:100 depending on the initial concentration of the amplicons used as determined by gel electrophoresis (see above). In order to obtain the DNA concentration required for sequencing (30ng/μL), the contents of two tubes of PCR product from each isolate (50μL each) were purified and concentrated with QIAquick PCR purification kits (Qiagen) to obtain a total volume of 30μL. Each sequencing reaction required 8μL DNA. The primer initially used in all sequencing reactions was Universal 1492 reverse (5'ACGGCTACCTTGTTACGAC3') at a concentration of 3μM.

Sequencing Facility and Software – The sequencing facility used for this study was the Division of Biological Sciences DNA Sequencing Facility at UC Davis (http://dnaseq.ucdavis.edu/). Sequences were run on an ABI 3730 Capillary Electrophoresis Genetic Analyzer with ABI BigDye Terminator v3.1 sequencing chemistry. Sequence results were received by e-mail (Eudora) and analyzed with Macintosh OSX using the viewing software 4Peaks (http://www.mekentosj.com/4peaks/index.html).

Results

Samples of 16S ribosomal DNA (approximately 1500 base pairs in length) obtained from 14 different bacterial isolates were amplified by means of the PCR, concentrated, purified and partially sequenced. Sequence data containing 500-800bp was edited to eliminate miscalled bases and regions containing poorly resolved sequence. Text versions of edited and unedited sequences were then compared to the public database of NCBI using BLAST. An example of text file data from one isolate is shown below.

Sequence sample HSC-1 – Unedited version *Exiguobacterium gaetbuli* (811/825 = 98%)
TMGTTCGGCCTCWTCTGACACATCGGCGGCTGGCTCTTACGGTTACCTCACCGACTT
CGGGTGTTGCAAACTCTCGTGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGT
ATTCACCGCAGTATGCTGACCTGCGATTACTAGCGATTCCGACTTCATGCAGGCGAG
TTGCAGCCTGCAATCCGAACTGAGAACGGCTTTCTGGGATTGGCTCCACCTCGCGGC
TTCGCTGCCCTTTGTACCGTCCATTGTAGCACGTGTGTAGCCCAACTCATAAGGGGC
ATGATGATTTGACGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGA
GTGCCCAACTTAATGCTGGCAACTAAGGACAAGGGTTGCGCTCGTTGCGGGACTTAA
CCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTCACCCCTGCCC
CCGAAGGGGAAGATACATCTCTGTACCGGTCAGGGGGATGTCAAGAGTTGGTAAGG

TTCTTCGCGTTGCTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCA
ATTCCTTTGAGTTTCAGCCTTGCGACCGTACTCCCCAGGCGGAGTGCTTAATGCGTTA
GCTTCAGCACTGAAGGGCGGAAACCCTCCAACACCTAGCACTCATCGTTTACGGCGT
GCACTACCAGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTT
ATAGGCAAGAGTCGCTTCGCCACTGGTGTTCTCACATCTCTACGCATTCACCGCTAC
ACGKGGAATGCCACTCTTCTCTCCTATACTCAAGCCTCC

Sequence sample HSC-1 Edited version – *Exiguobacterium gaetbuli* (604/606 = 99%)
TMGTTCGGCCTCWTCTGACACATCGGCGGCTGGCTC
Reliable sequence begins here – TTACGGTTACCTCACCGACTTCGGGTGTTGCAA
ACTCTCGTGGTGTGACGGGCGGTGTGTACAAGACCCGGGAACGTATTCACCGCAGT
ATGCTGACCTGCGATTACTAGCGATTCCGACTTCATGCAGGCGAGTTGCAGCCTGCA
ATCCGAACTGAGAACGGCTTTCTGGGATTGGCTCCACCTCGCGGCTTCGCTGCCCTT
TGTACCGTCCATTGTAGCACGTGTGTAGCCCAACTCATAAGGGGCATGATGATTTGA
CGTCATCCCCACCTTCCTCCGGTTTGTCACCGGCAGTCTCCTTAGAGTGCCCAACTTA
ATGCTGGCAACTAAGGACAAGGGTTGCGCTCGTTGCGGGACTTAACCCAACATCTCA
CGACACGAGCTGACGACAACCATGCACCACCTGTCACCCCTGCCCCCGAAGGGGAA
GATACATCTCTGTACCGGTCAGGGGGATGTCAAGAGTTGGTAAGGTTCTTCGCGTTG
CTTCGAATTAAACCACATGCTCCACCGCTTGTGCGGGTCCCCGTCAATTCCTTTGAGT
TTCAGCCTTGCGACCGTACTCCCCAGGCGGAGTGCTTAATGCGTTAGCTTCAGCACT
GAA – and ends here.
GGGCGGAAACCCTCCAACACCTAGCACTCATCGTTTACGGCGTGCACTACCAGGTAT
CTAATCCTGTTTGCTCCCCACGCTTTCGCGCCTCAGCGTCAGTTATAGGCAAGAGTC
GCTTCGCCACTGGTGTTCTCACATCTCTACGCATTCACCGCTACACGKGGAATGCCA
CTCTTCTCTCCTATACTCAAGCCTCC

Analysis of all the sequence data obtained during the spring semester generated the following results:

1. Query sequence #1 showed greatest homology with *Exiguobacterium gaetbuli* (811/825 = 98% unedited, and 604/606 = 99% when edited).
2. Query sequence #2 showed greatest homology with *Photorhabdus luminescens*, *Escherichia fergusonii*, *Escherichia coli* K-12 and *Shigella flexneri* (828/838 = 98% unedited and 559/559 = 100% when edited).
3. Query sequence #3 showed greatest homology with *Janthinobacterium lividum* (507/518 = 97% unedited and 480/481 = 99% when edited).
4. Query sequence #4 showed greatest homology with *Rhodococcus luteus* or *R. fascines* (561/564 = 99% unedited and 523/528 = 99% when edited). Colony descriptions (Bergey's Manual) were most similar to *Rhodococcus luteus*.
5. Query sequence #5 showed greatest homology with *Bacillus simplex* and *B. macroides* (576/577 = 99% unedited). The edited data showed greater homology with *Bacillus simplex* (577/577 = 100%).
6. Query sequence #6 showed greatest homology with *Microbacterium schleiferi* (532/534 = 99% unedited and 590/591 = 99% when edited).
7. Query sequence #7 showed greatest homology with *Staphylococcus warneri* and *S. pasteuri* (532/536 = 99% unedited and 536/537 = 99% when edited).

8. Query sequence #8 showed greatest homology with *Rhodococcus erythropolis* (617/620 = 99% unedited and 609/609 = 100% when edited). Colonies formed by the live culture did not match the description for this species (Bergey's manual).
9. Query sequence #9 showed greatest homology with *Nocardia corynebacteroides* (853/865 = 98% unedited and 659/662 = 99% when edited). The edited sequence also showed homology with *Rhodococcus equi* (658/662 =99%).
10. Query sequence #10 showed greatest homology with *Micrococcus luteus* (828/847 = 97% unedited) and *Variovorax* sp. (829/847 = 97% unedited). The edited sequence showed homology with Antarctic bacterium (560/560 =100%), *Variovorax sp*. (559/560 = 99%) and *Micrococcus luteus* (558/560 = 99%).
11. Query sequence #11 showed greatest homology with *Erwinia cypripedii* and *E. amylovora* (495/511 =96% unedited and 402/407 = 98% when edited). The sequence data was weak and indicated that the culture was not pure.
12. Query sequence #12 showed greatest homology with *Micrococcus sp*. (461/477 = 96%) and *M. luteus* (461/478 = 96%). The edited version showed greater homology with *Micrococcus luteus* (254/254 = 100%). Data suggested this culture was not entirely pure as there was much evidence of peak overlap in the electropherogram.
13. Query sequence #13 showed greatest homology with *Vibrio alginolyticus* (483/513 = 94% unedited and 367/367 = 100%). Weak signal on all of this sequence indicated DNA concentration was low.
14. Query sequence #14 showed greatest homology with *Porphyrobacter donghaensis* (337/356 = 94%) and *P. tepidarius* (336/356 = 94%) when unedited. The edited version showed homology with *Porphyrobacter donghaensis* (306/315 = 97%) and *Erythromicrobium ramosum* (306/315 = 97%). The electropherogram showed weak signal and multiple overlapping peaks suggesting the presence two different amplicons.

Of the 14 samples sequenced, five were found to contain regions showing 100% homology with previously identified strains including *Escherichia coli*, *Bacillus simplex*, *Rhodococcus erythropolis*, *Micrococcus luteus* and *Vibrio alginolyticus*. One of the samples was found to be 100% homologous with an isolate obtained from frozen lake samples (Antarctic bacterium). Six of the samples were found to contain regions showing 99% homology with previously identified strains including *Exiguobacterium gaetbuli, Janthinobacterium lividum, Rhodoccus luteus* and *R. fascines, Microbacterium schleiferi, Staphylococcus warneri* and *S. pasteuri,* and *Nocardia corynebacteroides*. One of the samples contained a region of DNA 98% homologous to *Erwinia cypripedii* and *E. amylovora*, and one contained a region 97% homologous to *Porphyrobacter donghaensis* and *Erythromicrobium ramosum*. Both of the samples showing low sequence homology were taken from cultures of questionable purity, i.e., cultures possibly containing more than one type of bacteria.

Discussion

Findings associated with this study suggest that application of the polymerase chain reaction and bioinformatics provides and interesting addition to traditional methods of bacterial identification in undergraduate microbiology laboratories. Although kits involving spin-column technology increase the purity of genomic DNA samples, their application is not always necessary. DNA samples suitable for PCR amplification can be obtained from many commonly encountered

Gram-negative bacteria by means of a simple boiling procedure. Gram-positive bacteria present more of a challenge, and bead-beater technology or extraction kits involving lysozyme appear to be necessary for success.

Multiple sources provide detailed formulations for Taq polymerase, $MgCl_2$ and dNTP reaction mixtures, but for beginning researchers and students, a premade master mix is most convenient. The Taq polymerase master mix used in this study was relatively inexpensive and demonstrated remarkable stability. Storage and stability specifications included storage at –20ºC in a constant temperature freezer, but indicated that kits could be stored at 2-8ºC for up to 2 months without reduction in performance. The master mix used in this study was stored in a frost-free freezer, thawed and refrozen numerous times, and continued to function for over 3 months.

General recommendations for avoiding contamination during the preparation of PCR mixtures include the used of a laminar-flow hood equipped with ultra-violet light, and designated pipettes with cotton-plugged tips. Although important in research settings, these requirements effectively prohibit efficient application of the PCR in undergraduate laboratory classrooms with 24-30 students. During this study, students in four class sections applied the PCR while working on open lab benches cleaned with disinfectant solution. They used digital pipettes equipped with standard tips, and not restricted to PCR preparation. During one protocol, students attempted to amplify DNA extracted from bacterial cultures boiled in TSB, centrifuged, and suspended in Tris buffer. Less than half of these samples contained PCR product at visible concentrations following electrophoresis and UV illumination. Although significant quantities of the desired template DNA were present in all samples, two of the cultures failed to yield results for any class section. Students working independently and preparing materials for the PCR using DNA extraction kits were much more successful, i.e., all of their samples yielded results. During all portions of this study, achieving successful PCR amplification of desired template presented a greater challenge than restricting contaminants.

Identification involving 16S ribosomal DNA nucleotide sequencing often required augmentation or verification supplied by analysis of data obtained by more traditional methods. During this study, a 559bp segment of DNA (sample #2) found to have 100% sequence homology with *Escherichia coli* K-12, showed equal homology with *Photorhabdus luminescens*, *Escherichia fergusonii*, and *Shigella flexneri*. Results obtained with enzymatic testing were necessary to confirm the culture identity as *Escherichia coli*. Several 16S ribosomal DNA samples showed equal homology with two different species e.g., #4 (99% homology with both *Rhodococcus luteus* and *R. fascines*), #7 (99% homology with both *Staphylococcus warneri* and *S. pasteuri)*, #11 (98% homology with both *Erwinia cypripedii* and *E. amylovora*), and #14 (97% homology with both *Porphyrobacter donghaensis* and *Erythromicrobium ramosum*). In the case of *Rhodococcus luteus*, colony description and enzymatic testing confirmed the culture identity. Identity of the *Staphylococcus* culture could also be confirmed through enzymatic testing. Cultures #11 and 14 were not pure, and no confirmed identification was possible. In one case, the amplified DNA sample showed greatest sequence homology with an unidentified species, i.e., sample #10 (a 560bp sequence showing 100% homology with a sample identified as Antarctic bacterium). Identification of the culture was then based on the two genera showing 99% homology, i.e., *Micrococcus luteus* and *Variovorax sp*. Observation of yellow-pigmented colonies and application of enzymetic tests confirmed the culture identity as *Micrococcus luteus*.

Several of the DNA samples investigated showed strong sequence homology with unexpected genera or in some cases with genera not described in the Bergey's manuals. Sample #6 (showing 99% homology with *Microbacterium schleiferi*) was taken from a culture originally identified as *Xenorhabdus poinarrii*, in the family Enterobacteriaceae. This culture has been maintained for several years, and used in the laboratory as an unknown in various staining exercises. Students have consistently identified it as being Gram-negative, while *Microbacterium* cultures are Gram-positive. Validity of this surprising Gram-stain finding was reinforced by DNA extraction requirements. Attempts at PCR amplification were unsuccessful when the culture was subjected to methods effective for Gram-negative forms, but was successful when it was treated with a bead-beater kit. Sample #1 (showing 99% homology with *Exiguobacterium gaetbuli*) was taken from a culture previously identified as *Azotobacter nigricans*. Although the phenotypic characteristics of the culture were not entirely consistent with any *Azotobacter* species described in the Bergey's manual, the colonies were dark brown, the cells Gram-negative, large (more than 2μ in diameter) and the culture was obtained from soil using a medium formulated for the isolation of *Azotobacter* species (i.e., was nitrogen-free). PCR amplification of DNA from this culture was successful following extraction involving a kit and protocol designed for Gram-negative forms. Since a description of the genus *Exiguobacterium* is not included in currently available editions of the Bergey's manual, the identity of this culture remains questionable. The genera *Porphyrobacter* and *Erythromicrobium* along with several of the species indicated for our isolates are not included in the Bergey's manuals available. Additional research will be required to verify the identify of these cultures.

An interesting feature of this study was the diversity of the organisms investigated relative to the number. Only 14 isolates were used for the sequencing portion of the study, but they represented 12 different genera in 9 different families, 6 different orders, 5 different classes, and 3 different phyla. Most of the species indicated by sequence homology were new, i.e., not formerly encountered in our laboratory, and the descriptions available for some did not match the cultures being investigated. This may indicate that one or more of the cultures included represent bacterial species not yet described. Since the nucleotide sequences used in this study were incomplete, i.e., included only about 1/3 of the 16S ribosomal DNA amplified, definitive conclusions cannot yet be drawn. Additional research is necessary.

Taxonomic information:

Bacteria in the genera *Erythromicrobium* and *Porphyrobacter* belong to the family Sphingomonodaceae, Order Sphingomonodales, in the Class Alpha-Proteobacteria, phylum Proteobacteria
Bacteria in the genus *Janthinobacterium* belong to the family Oxalobacteraceae, Order Burkholderiales, Class Beta-Proteobacteria, phylum Proteobacteria
Bacteria in the genus *Vibrio* belong to the family Vibrionaceae, order Vibrionales, class Gamma-Proteobacteria, phylum Proteobacteria.
Bacteria in the genera *Escherichia* and *Erwinia* belong to the family Enterobacteriaceae, order Enterobacteriales, class Gamma-proteobacteria, phylum Proteobacteria.
Bacteria in the genus *Bacillus* belong to the family Bacillaceae, order Bacillales, Class Bacilli, Phylum Firmicutes.

Bacteria in the genus *Exiguobacterium* belong to the family Bacillaceae, order Bacillales, Class Bacilli, Phylum Firmacutes.
Bacteria in the genus *Staphylococcus* belong to the family Staphylococcaceae, order Bacillales, Class Bacilli, Phylum Firmacutes.
Bacteria in the genus *Micrococcus* belong to the family Micrococcaceae, suborder Micrococcinae, order Actinomycetales, class Actinobacteria, phylum Actinobacteria.
Bacteria in the genus *Microbacterium* belong to the family Microbacteriaceae, suborder Micrococcinae, order Actinomycetales, class Actinobacteria, phylum Actinobacteria.
Bacteria in the genera *Nocardia* and *Rhodococcus* belong to the family Nocardiaceae, suborder Corynebacterineae, order Actinomycetales, class Actinobacteria, phylum Actinobacteria.

Acknowledgments

I would like to thank Professor Diana E. Northup, Ph.D., University of New Mexico, Albuquerque for her inspiration, encouragement and technical assistance, Judith A. Kjelstrom, Ph.D., Associate Director, Biotechnology Program, University of California, Davis for her educational support and networking assistance, Hazel A. Barton, Ph.D., for UCD facilities tour and hallway tutorial, Jeffery O'Neal, Ph.D., Director, North Valley and Mountain Biotechnology Center, for educational opportunities and financial support, Debbie Birnby, Ph.D., Senior Technical Service Scientist (Qiagen), Sheryl Bernauer and Kerry Cloud (DBS DNA sequencing facility) for much needed technical assistance, Elaine Atnip and Benjamin Lewis for hours in the laboratory running and analyzing PCRs, and James L. Wilson for his patience, laboratory support and computer expertise.

References

Barton, H.A., Taylor, M. R., and Pace, N. R. 2002. Molecular Phylogenetic Analysis of a Bacterial Community in an Oligotrophic Cave Environment. Manuscript GJ (C) 02-21.

Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, JP, and Raoult, D. 2000. 16S Ribosomal DNA Sequence Analysis of a Large Collection of Environmental and Clinical Unidentifiable Bacterial Isolates. Journal of Clinical Microbiology. 38 (10): 3623-3630.

Frame, Peter. (2000-2002). Ten Things That Can Kill Your PCR. The Nucleus. biowire.com. Retrieved from http://www.biowire.com/nucleus/nucleus_1_1.jsp

Interpretation of Sequencing Chromatograms. (n.d.) Retrieved May 19, 2004 from the University of Michigan DNA Sequencing Core Web Site: http://seqcore.brcf.med.umich .edu/doc/dnaseq/interp.html

Northup, D. E., Barns, S. M., Yu, L. E., Spilde, M. N., Schelble, R. T., Dano, K. E., Crossey, L. J., Connolly, C. A., Boston, P. J., Natvig, D. O., and Dahm, C. N. 2003. Diverse Microbial Communities Inhabiting Ferromanganese Deposits in Lechuguilla and Spider Caves. Environmental Microbiology. 5(11):1071-86.

Vierstraete, A. *Principle of the PCR*. Retrived March 3, 2004, from Department of Biology, University of Ghent. http://allserv.rug.ac.be/~avierstr/principles/pcr.html

Sambrook, J. and Russell, D. W. 2001. *In Vitro Amplification of DNA by the Polymerase Chain Reaction*, Chapter 8. *In* Molecular Cloning a Laboratory Manual, Third Edition, vol. 2. Cold Springs Harbor, New York.

Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D., and Doré, J. 1999. Direct Analysis of Genes Encoding 16S rRNA from Complex Communities Reveals Many Novel Molecular Species within the Human Gut. Appl. Environ. Microbio. 65 (11): 4799–4807.

Tchan, YT. 1984. *Azotobacteraceae*, p. 219-234. *In* N. R. Krieg, and G. J. Holt (ed.) Bergey's Manual of Systematic Bacteriology, vol. 1. Williams & Wilkins Co., Baltimore, Md.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. 1991. 16S ribosomal DNA Amplification for Phylogenetic Study. Journal of Bacteriology. 173 (2): 697-703.